Doctoral Dissertation Academic Year 2017

# Context-Dependent Corporate Analysis Methods Using Integrated Multi-Domain Semantic Space

A dissertation for the degree of
Ph.D. in Media and Governance


Graduate School of Media and Governance
Keio University

**Shin Ito**

# Abstract

This thesis proposes a new evaluation method and system implementation for companies and exchange-traded financial assets.

Companies are the source of economic value generation in the modern societies and stakeholders (investors, suppliers, employees, customers, regional residents, analysts, etc.) play important roles in economic activities. Stakeholders attempt to avoid risks and maximize profits and to that end evaluate companies at various stages of their engagements with the companies. Today, the useful information they have access to is limited thus an elaborate methodology and evaluation system is highly demanded that compensates stakeholders' insufficient capabilities to evaluate companies in a proper manner.

This thesis presents a computational method that evaluates variety of companies according to the stakeholders' individual evaluation criteria. The computational method is realized by constructing a multidimensional semantic metric space that characterizes the entities, mapping each entity and a stakeholder's contextual needs in the space, and calculating the relevant semantic distances. It utilizes the Mathematical Model of Meaning as a semantic association mechanism in a multidimensional semantic metric space, define a stakeholder's evaluation criteria as a "context", and calculate the special distance within the subspace selected by the context. The methodology enables customized evaluation of entities by translating the stakeholder's context to defining a proper semantic subspace, thus allowing individualized evaluation unlike the currently available special purpose systems.

This thesis's evaluation methodologies apply to companies as well as exchange-traded financial assets and an investment return prediction model was demonstrated. States of an Exchange Traded Fund in time series was characterized to construct a multidimensional semantic metric space, the state of each instance in times series was mapped in the semantic space, current instance and historical instances were compared for similarities by semantic special distance calculations, and the instance of the highest similarities (shortest distance) is identified, and the very next instance of the discovered instance is taken as the predicted state for the immediate future from the current state. Manipulations of subspace selection caused different precision of the predictive capability of the model represented by the correlation between the actual return and the predicted return leading to improving the model. The use of historical data as actual "predicted" return values led to a discovery of a methodology for solving inverse problems of the Mathematical Model of Meaning. This thesis indicates that by specifying resulting value, the associated context can be found.



Keywords: Multidimensional Semantic Matric Space, Context-dependent, Corporate Evaluation, Stakeholders, The Mathematical Model of Meaning, Inverse Problem

# 論文要旨

本研究は、企業及び市場取引資産を対象とし、多様な視点から分析・評価する計量方式とそのシステムの実現を提案している。企業は経済価値創造の源泉であり、投資家、原材料供給者、従業員、分析者等多くのステークホルダーが企業の内外で活動している。ステークホルダーは目的に応じ、様々な段階において対象企業の評価を行うが、入手可能な情報や分析能力は限定的である。ステークホルダーが、夫々の個別要求に基づき、柔軟かつ包括的に企業評価を実行できる計量方式やシステムの実現が期待されている。

本研究は、様々なステークホルダーの個別要求基準を設定し、それらによる多様な企業評価を実現する計量方式とその実現システムを構築した。その計量方式は、企業についての多種多様な特徴量を表す多次元意味計量空間を構成し、評価対象企業群及びステークホルダーの個別評価基準をこの多次元意味計量空間に写像し、多次元意味計量空間における距離として計量する方式である。具体的には、多次元意味計量空間における意味的連想機構として提案されている意味の数学モデルを応用し、ステークホルダーの個別評価基準を“コンテキスト”という概念を導入することによって表現し、コンテキストに応じ多次元意味計量空間の部分空間を選択し、この部分空間における距離計算により、企業評価を行なう方式である。本方式は、特定の分析目的を対象に個別のシステムにより実現されている既存の企業評価とは異なり、個々のステークホルダーが要求する評価基準をコンテキストとして設定することで、ステークホルダーの多様な個別要求に対応する。

さらに、本研究では、多次元意味計量空間を用いた解析の応用事例として、市場取引される金融資産の投資リターン予測モデルを構築した。具体的には、投資信託資産の市場価格変動の時系列事象を各時点における状態として空間写像し、直近状態と過去における各時点の状態との距離計算を行うことにより類似時点を求め、当該過去の次時点状態を将来予測値とするものである。多次元意味計量空間から適切な部分空間を選択する実証実験により、本金融資産投資リターン予測モデルの有効性を明らかにした。さらに、本研究は、既知の予測評価値の部分空間による分布を解析することにより、市場取引される資産を評価する際の主要な部分空間、すなわち、評価基準に対応するコンテキストを発見する逆問題分析方式を示した。

キーワード：多次元意味計量空間、コンテキスト依存、企業評価、ステークホルダー、意味の数学モデル、逆問題

## Acknowledgements

I would like to express my deepest gratitude to my main research supervisor and advisor Professor Yasushi Kiyoki of Keio University for giving me invaluable and extremely insightful advice. Dr. Kiyoki's advise was sometimes very philosophical other times very technical. I always enjoyed hearing him, exchanging ideas with him, and learning from him. Dr. Kiyoki's enthusiasm and undivided attention toward scientific research greatly inspired me in pursuit of my research. I am totally indebted to Dr. Kiyoki for his kind help and contribution for my research.

I would also like to thank my research advisors. Professor Yoshiyasu Takefuji gave me very interesting and stimulating insights from many of his research activities. Professor Takefuji always had something new to share. Professor Emeritus Dr. Mitsuyo Hanada gave me very deep knowledge about human capital and business enterprises. Dr. Hanada was always very kind and supportive for my research.
I would also like to thank Professor Takeo Kuwahara for his comments and very pointing advise in my research. Dr. Kuwahara's questions motivated me to explore deeper in my thought process.

I would like to express my sincere gratitude to all members and associates of the MDBL and Kiyoki research family. Special thanks to Dr. Shiori Sasaki, Dr. Asako Uraki, Dr. Yasuhiro Hayashi, Dr. Shuichi Kurabayashi, and Professor Naofumi Yoshida of Komazawa University who always gave me kind support for my research activities and great questions and advice on my academic work. I am very thankful for the friendship with all members of MDBL.

I would also like to extend my thanks to associates at academic conferences, in particular, the organizers and participants at the European Japanese Conference on Information Modelling and Knowledge Bases (EJC) and Knowledge Creation and Intelligent Computing (KCIC) who presented very interesting ideas and gave me insightful and stimulating comments to my work.

Finally, I would like to thank my wife Mika Ito for giving me warm and dedicated support for my research work. Without her understanding and help, I could not have accomplished it.

## Table of Contents

# List of Tables

## List of Figures

# CHAPTER 1. INTRODUCTION

## ISSUES WITH CURRENT CORPORATE EVALUATION

Companies are the driving forces of the economic value creation in the modern industrialized societies and are a very important integral part of today's societies. Because companies interact with great many diverse constituencies, there are many stakeholders, such as shareholders, employees, lenders, suppliers, customers, governments, and analysts. Each of these stakeholders has different interests in relation to the companies they interact with and wishes to avoid risks and maximize the benefit in association with the companies.

The types of interest greatly vary depending on the stakeholders and their contexts in which they happen to be when interacting with companies. For instance, a university student who is interested in getting a job in certain technology area may wish to know how strong companies are in such technological field now and how committed they are in the forthcoming future. At the same time the student may wish to know the company's financial stability for securing a long-term employment while also interested in getting decent salary and benefits. Another example may be a potential component supplier to a company who is contemplating whether to start a business relationship. In such instance, the supplier is interested in knowing how secure the company is in making the payments for the sale of the supplier products in a timely manner. The supplier may also be interested in the volume growth by business success in gaining market share. As in these examples, stakeholders are interested in knowing about the company as a whole, and to a higher degree, certain areas of particular interest. However, despite their interest, it is extremely difficult to adequately understand the companies of interest as an outsider.

Professional stakeholders, such as institutional investors, tend to have better access to relevant information and are able to interpret and use such information to make critical business decisions. These professional stakeholders tend to have appropriate relationships with corporate personnel to get access to such relevant information in a timely manner and utilize the information to take full advantage for themselves or for their clients. But even these resourceful professionals sometimes exhibit conflicting views among themselves, indicating difficulty in the assessment.

Stakeholders with limited resources and less sophisticated skills are disadvantaged in making optimal decisions. These stakeholders tend to have less access to relevant information about the company and have inferior capabilities of making use of such information. Some have access to direct and relevant information through interactions with the company but because of the narrow scope of interaction, they lack the holistic view leading to partial and local decision making. Correct interpretation of publicly available information tends to be difficult for these stakeholders to make reasonably correct assessment of the company.

There exist a number of reasons that make it difficult for anyone to assess the state of companies based on public data. One of the major reasons of the problem is the fact that the publicly available data does not adequately describe the state of companies. Publicly traded companies in stock exchanges are required by respective government agencies to disclose financial information and any facts and potential risks that may materially affect companies' state. Such financial information is disclosed in certain format and rules that are defined by financial accounting authorities such as Financial Accounting Standard Board (FASB) that defines US Generally Accepted Accounting Principle (US GAAP) and International Accounting Standards Board (IASB) that sets up International Financial Reporting Standards (IFRS). Each jurisdiction usually institutes and adopts a financial accounting standard. There are some differences among such accounting standards but the vast majority is a common set of rules. Overall, the current accounting system has not caught up with the reality of corporate advancement in developing intangible assets[1], particularly in the area of intellectual properties. With ever rapidly evolving corporate entities in the global environment where intellectual properties provide differentiations from others and act as the core competitive edge among all players, financial accounting system has not captured such important assets in the balance sheet.

Companies are valued much more by the stock market than what is recorded in the balance sheet in financial reporting, manifesting as a great gap between the market value and the book value of companies. Hall (2000) estimates the total value of intangible capital as ranging between one half to two-thirds of the total market value of publicly traded corporations, indicating consistent departure of Tobin's Q[2] in excess of 1 among publicly traded companies in the US. Miyanaga et. al. (2014) also showed that the revised Tobin's Q incorporating intangibles such as software development, research and development, brand development, firm-specific human capital development, and organizational changes significantly approach 1 as compared with only tangible assets using publicly traded Japanese firms. Assuming certain market efficiency where firm values face constant scrutiny by market participants, each stock

---

[1] A framework developed by Lev (2001) for intangible capital classifies intangible assets into the following four groups:
(A). *Discovery/learning intangibles*—technology, know-how, patents and other assets emanating from the discovery (R&D) and learning (e.g., reverse engineering) processes of business enterprises, universities and national laboratories.
(B) *Customer-related intangibles*—brands, trademarks and unique distribution channels (e.g., internet-based sales), which create abnormal (above cost of capital) earnings.
(C) *Human-resource intangibles*—specific human resource practices such as training and compensation systems, which enhance employee productivity and reduce turnover.
(D) *Organization capital*—unique structural and organizational designs and business processes generating sustainable competitive advantages.

[2] Tobin's Q = (Firm equity market value + debt book value)/ (Replacement costs of all existing assets of the firm) The Q ratio is expected to be 1 in equilibrium.

price should equal the present value of the firm's future cash flow. As such cash flow is produced as a result of corporate operational activities based on the net asset, it follows that there exists some form of intangible assets that should be capitalized in the balance sheet.

Recognizing the gap between the reality and the accounting system, IFRS has instituted a new rule to recognize, measure, and record certain intangible assets in the balance sheet.

### MOTIVATION FOR THIS RESEARCH

In order to provide a means for multiple stakeholders to assess state of companies and thus allow them to make intelligent predictions of their interactions with the companies, I propose context-dependent integrated multi-domain corporate evaluation method. This method will allow the user to evaluate a group of companies and score them in accordance to the user's specific context. Such context may include the nature of the user's proposed interactions with companies such as investment and employment, associated conditions under which such proposed interactions are to be conducted, and other restrictions such as time. With this system, the user will be able to significantly better understand the companies of their interest and decide on the course of their actions intelligently.

### RELATED WORK

There are numerous academic papers and industry practices that have been published and known in public in corporate evaluation. However, most of them are intended to provide for narrow audience with specific purposes in mind. Most notably, financial evaluation of companies leads the effort. Financial terms are most developed as the measurement of companies' financial performance has been much better established and such measures have been standardized and adopted in many countries.

International Accounting Standards Board approves and updates such international standards.

Context-dependent semantic space calculation was pioneered by Kiyoki and Kitagawa in their monumental work in the Mathematical Model of Meaning (MMM)[3] in 1993. Based on this fundamental concept, there have been numerous applications of the MMM in the field of images, medicine, socio-politics, education, etc. but there have not been any applications in the field of business management.

**EXISTING CORPORATE EVALUATION SYSTEMS**

There exist corporate ranking publications on a global scale. Perhaps the most well-known is "America's Most Admired Companies 500" and "World's Most Admired Companies 500" published by Fortune Magazine– referred as "Fortune 500 companies." Other companies such as Forbes, Businessweek, Financial Times, and Interbrand provide their own rankings of companies. These rankings can be interpreted as generally accepted ratings of companies for wide audiences.

Nikkei provides a service called NICES as a successor to CASMA and PRISM which used to utilize multivariate analysis using certain variables with weight factors Nikkei generated. CASMA (Corporate Appraisal System by Multivariate statistical Analysis) was for large companies and used (i) corporate size, (ii) profitability, (iii) stability, and (iv) growth potential with fixed weight distribution among these factors for a given year. PRISM (PRIvate Sector Multi-angular evaluation system) generated rakings for smaller companies using (i) profitability and growth potential, (ii) flexibility and social factors, (iii) research and development capabilities, and (iv) youth – flexible human capital utilization, with constant weight distribution for a given year. The new raking system NICES (Nikkei Investor – Customer – Employee – Society) does not use

---

[3] T. Kitagawa and Y. Kiyoki, ``A mathematical model of meaning and its application to multidatabase systems," Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993.

multivariate analysis but evaluates companies from the view point of the named stakeholders. Each of these categories include 5 to 7 indexes obtained through financial data as well as survey and the weight was determined by Nikkei editors and Internet survey.[4]

NICES-based ranking is a step forward as it provides the evaluative view points of the stakeholders. However, all existing raking system utilizes constant weight factors thus provides very rigid ranking system. Even when NICES offers views for different stakeholders, they are stereotypical or representing instances, failing to provide flexibility for individual contextual needs.

## NEW PROPOSAL FOR CORPORATE EVALUATION

I propose a completely new corporate evaluation system with customizable user-context dependency, allowing users to individually set their evaluation criteria. And the significance of this system is that it uses a common set of databases and selects the subspace specified by the user context to evaluate the companies. I utilize the numeric space and Mathematical Model of Meaning (MMM) and expand the application of the model into the economic activities.

## SUMMARY OF CHAPTERS

Here I summarize each of the chapters in this doctoral dissertation.

### *Chapter 1 Introduction*

I identify issues with current company evaluation systems and state motivations for a flexible user context-dependent evaluation system. The issues are twofold. (1) Current evaluation system covers only limited aspects of company, and (2) that it does not cover multiple users with different purposes assuming their own contexts. This

[4] T. Furuyama, "Nikkei no Kigyou Hyouka System – NICES no hyouka wo chushin nisite" Keiei Bunseki Kenkyu Annual Report Vol 29.
.

dissertation addresses both of these issues utilizing semantic space and user-defined context.

***Chapter 2 Evaluating Companies – Current Practice***

Current company evaluation methods commonly used by practitioners are described, pointing out that most of the available methods are qualitative that require expert knowledge and skills to understand companies.

***Chapter 3 Semantic Space Analysis Methods***

The concept of semantic space is introduced in reference to the Mathematical Model of Meaning. The concepts of Characteristic Parameters and subspace selection by the user contexts are introduced as essential components for the analysis methods using a semantic space.

***Chapter 4 Constructing Sematic Spaces***

Construct of a semantic space is explained in detail. Procedure of creating semantic space in the domain of finance, technology, and multi-domain is described with examples of Characteristic Parameters.

***Chapter 5 Experiments on Context-Dependent Corporate Evaluation in Semantic Space***

Experiments were conducted for a single-domain and multi-domain semantic spaces with user-context dependency. US semiconductor companies as well as large multi-national companies were used for experimental illustration respectively. Different user context were tested and the model viability was examined.

***Chapter 6 Experiment on Time Series Analysis Using Financial Semantic Space***

As an application of the context-dependent semantic space evaluation methods, a market return predictive model is explored. The state at each historical instance is mapped in a financial semantic space and distance is calculated to identify the degree of similarity between instances. This method is used to build a predictive model for future market return.

***Chapter 7 An Improved Predictive Model and a Methodology to Solve Inverse Problem***

I introduce a market return predictive model which is simpler but is viable compared with the model introduced in Chapter 6. I also describe a methodology to solve inverse problem for the Mathematical Model of Meaning to further improve the effectiveness of the predictive model.

***Chapter 8 System Configuration***

The overall system architecture and implementation is explained. A demonstration of user interface in specifying the target companies as well as user context is described.

***Chapter 9 Summary and Future Research***

The thesis summary is given and new areas of research interests are explored.

# CHAPTER 2. EVALUATING COMPANIES – CURRENT PRACTICE

Companies are evaluated for a wide variety of purposes by great number of analysts and reporters. The purposes of such corporate evaluation include potential financial transactions such as investment and lending, commercial transactions both supply and purchasing, partnering with other companies, employment engagement, and contemplation to enter a new markets and industry, etc. In this chapter, I review how companies are currently evaluated to form a foundation of new approach that is presented in this dissertation.

## Qualitative Approach

Companies operate in the global socio-economic system, and therefore they are under the influence of many socio-economic environmental factors. There have been numerous research done on corporate performance at different layers of such socio-economic system. Below is a typical approach in assessing the state of a company in a qualitative way. This approach is a top-down system and analyses a company from at large environmental level, at industry level, and to the company level of interest.

### MACRO ENVIRONMENT

Many factors affect the current state and future directions of companies. Analysts consider these factors in analyzing companies. Typical framework used at this macro environmental level is PEST – Political, Economical, Social, and Technological environment that influence the company of interest.

- Politics

- Economy
- Society
- Technology

These are usually environmental factors that companies cannot change but must adapt themselves to survive and prevail. Analysts examine what industries these macro-level environmental factors affect, to what extent such influence is made, and how long it will take to affect the industries.

Political factors include regulations and deregulations, policy changes, new legislations, policy changes such as consumer protection, patent policies, human rights, natural environment, energy policies, etc. It also includes government-backed protection of certain industries such as agriculture, economic stimuli packages, and new bilateral and multi-lateral treaties. Economic factors include economic cycles, interest rates and money supply each central bank sets forth, savings rates, foreign exchanges, and price levels and changes. Social factors can be value standards, religions, public opinion formations, lifestyle changes, education levels, population structural changes, etc. Technology factors include new scientific findings, technology development, technology adoption, new infrastructure platform, efficient manufacturing, telecommunication methods, etc. These factors influence industries at various magnitude and in turn individual companies thus it is very important to analyze these factors and affect they have in great care. And at the company level, successful companies also try to predict how these environmental factors will shift and predict what new business opportunities they represent as well as what threats they introduce to their existing businesses and cope with such changes better than others.

As an illustration, an economic treaty promoting free trade greatly affect industries as country protection set forth in the form of tariff and quota is significantly altered. Thus, weaker domestic industries such as agriculture in case of Japan will be damaged while stronger industries such as automobile manufacturing will benefit from it over time. Another example is patent policy and judicial decisions on court cases as well as administrative attitude on certain patent-related issues. In 1980s, after significant manufacturing businesses particularly in automobile and electronics were taken away from the US to foreign countries, the US government decided to place strong emphasis on upstream in the supply chain – research and development, and made pro-patent policy giving strong rights to patent holders. The pro-patent policy worked quite effectively and American companies transcended themselves to new forms of businesses. However, another kind of new business emerged now called "non-practicing entity" that takes advantage of the patent policy and exploits companies for profit. Their business is perfectly legal but their litigation-based business style became harmful to many industries. The recent legislations along with supreme court decisions indicate moving away from full pro-patent policy and attempt to better discover the optimal point of the intellectual property policy. By these changes, particularly the electronics industry benefited while the software industry and e-commerce-related industries were disadvantaged to certain degree. Non-practicing entities now face higher costs of doing business overall.

**INDUSTRY ATTRACTIVENESS**

The attractiveness of industries can be assessed in many ways; however, the most prevalent methodology was proposed by Michael Porter at Harvard Business School[5] [6]. There have been a number of follow-on publications by Porter and others[7] [8] [9] on the subject. Porter states that the well being of a particular company is dictated

---

5 Porter, M. E. "How Competitive Forces Shape Strategy." Harvard Business Review 57, no. 2 (March–April 1979): 137–145.

6 Porter, M. E. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press, 1980.

7 Bensanko, David, David Dranove, and Mark Shanley. *The Economics of Strategy*. John Wiley & Sons, Inc. 1996

8 Day, George S., David J. Reibstein. *Wharton on Dynamic Competitive Strategy*. John Wiley & Sons, Inc. 1997.

9 Mintzberg, Henry, Bruce Ahlstrand, and Joseph Lampel. *Strategy Safari: The Complete Guide through the Wilds of Strategic Management, 02 Edition*. 2009

by the degree of attractiveness of the industry the company operates in and its unique positioning of the company within the industry. The attractiveness of an industry is defined as the potential of profit making of the industry. Porter claims that such industry attractiveness is shaped by five competitive forces exerted from within and around the industry. The five forces are listed below.

- Threat of New Entrants
- Threat of Substitute Products and Services
- Bargaining Power of Buyers
- Bargaining Power of Suppliers
- Rivalry Among Existing Firms

Threat of new entrants is great when barriers to entry into the industry are low. The barriers to entry can be large capital investment such as manufacturing plant or national brand, established distribution channels, technological advancement, strong relationships with existing customers and suppliers that causes high switching costs of buyers and suppliers, brands that buyers are attracted to, etc. The threat of new entrants represents the potential new players in the industry that increases competition potentially leading to lower average selling price and higher acquisition costs of good, thus resulting in decreasing the potential profit that can be made by the industry.

Threat of alternative products and services potentially lowers the profitability by reduced volume of products and services the industry should be able to sell to its customers. This results in lower revenue generated by the industry thus lower profits. Customers' buying activities may be directed towards such alternative goods and services replacing the inherent sales volume the industry should enjoy. Threat of alternatives may become significant when the switching costs of buyers are low, that alternative offerings meet the buyer needs, and that the price sensitivity of buyers is high.

Bargaining power of buyers influences the price at which the industry constituents sell their products and services to buyers. When the bargaining power of buyers is high, buyers can negotiate the prices down so hard that lowers the price points are lowered to make the trade. This inevitably reduces the sales revenue of the industry leading to reduced profits. Buyer tends to have higher bargaining power when they are consolidated as in the case of large general mass merchandisers where the players exhibit oligopolistic behavior. Also, when the switching costs of buyers are low, as in the case of commodity trade, it gives buyers higher bargaining power. Having alternative products and services also give buyers a bargaining advantage against the industry.

Bargaining power of suppliers controls the price points which the industry buys goods from suppliers. When there are only a small number of suppliers of goods and there is no alternatives then the bargaining power of suppliers tend to be high. Microprocessors and the operating systems for personal computers were typical examples of this situation. Also, when there is high switching cost imposed to the industry, the negotiation power is shifted to suppliers. Scarcity of goods such as certain natural resources or industrial goods that are hard to mass-manufacture gives bargaining power to suppliers.

Rivalry among existing firms within the industry is typically the competitive situation thought of by most industry practitioners. The rivalry works to voluntarily lower the selling prices to buyers due to competitive pressure. In order to get businesses of buyers, companies tend to compete on prices during the course of negotiation. Also, an industry that requires high fixed costs tends to drive price-based competition to cover the fixed costs and achieve break even. Such price-based competition tends to become keen when product differentiation is difficult; that it, the products are commodities or there is set industry standards product and service offerings need to conform to. Gasoline, commonly used metals such as copper, wheat, courier services, etc. all fall in this category. Companies voluntarily lowers price to get businesses, reducing the industry sales revenue and profits.

The degree of these five competitive forces is examined and the strengths of each force are assessed. Each one of the five forces is a potential threat and may reduce profitability of the industry. Estimating the magnitude of the current forces and predicting any changes to the forces enable analysts to estimate the profitability of the industry and its future state. In other words, assessment of five competitive forces provides qualitative estimate of average profit making capacity of the industry.

**COMPANY POSITIONING**

Porter also suggests that how a company is positioned within the industry provides information on the wellbeing of the company. When a company has established a well-defined position within an industry, it can perform better over time than those that hover without a definitive position. These companies that are "stuck in the middle"[10] situation will either need to establish a new positioning or stay in low profitability and may become extinct in the long run.

The possible positioning a company can take are (1) low-cost - all segments, (2) differentiation – all segments, (3) low-cost - focused segment, and (4) differentiation – focused segment. The "stuck in the middle" situation is not completely cost-leader; that is, there are other players in the industry that can produce products and services at lower costs. At the same time, the company is not differentiated enough with their products and services to a set of target customers. Under this situation, the company cannot command high enough prices and is vulnerable to competitors that better-fit products to the target segment. Thus, the companies in the middle end up as a low performer. These generic competitive strategies are essential in maintaining profitability of a company.

Low-cost leadership in the industry can be achieved by two key factors: volume production and experience curve. Volume production provides the economy of scale

---

[10] Porter, M. E. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press, 1980: 41-44

where fixed costs, such as production plant costs and human resources can be spread over the volume production leading to lower per-product fixed costs. Experience curve is a phenomenon that both fixed costs and variable costs are reduced by cumulative number of production over time. So by securing the number of volume production against competitors, one can lower overall costs per product. Initial investments are usually needed to realize such volume demand, by way of advertisement or other promotion activities, and supply side volume by investing in large production facility or more efficient plant and equipment. If a company is not the leading mass producer to reduce costs and there exists another company that exhibits even lower overall cost profile, price war can severely damage the company to the extent it can no longer stay in business. Therefore, a company must use caution to take cost leadership positioning.

Differentiation positioning on the other hand appeals to the customer in a way to command premium price over others. A company that decides to take a differentiation positioning must develop certain appeal to the customer by providing certain distinctive value that customers perceive. This distinctive value needs to be very unique and must be well understood and occupy the minds of the customers. Certain brands give customers proof of trust and reliability while others give consumers rich and luxury satisfaction along with status symbols. Analysts examine very hard whether a company's differentiation strategy is sustainable by evaluating the uniqueness of the attributes the company is offering. If the differentiation positioning is not sustainable, then the premium price will not be accepted or kept at the intended level over time, resulting in reduced revenue thus lowered profits. If the differentiation factors are strong enough and such differentiation is supported by the company's competitive advantage others cannot readily imitate, then the premium price is sustained with intended volume sale, resulting in superior revenue and profit performance.

Focus strategy is to target certain segment of the entire market comprised of certain set of customers. Companies that choose focus strategy decides to do so because they believe they could serve the chosen customers significantly better than the whole set of customers and, at the same time, construct sustainable competitive advantage. The size of the chosen set of customers depends on the chosen set of customers' needs the company decided to fulfill with unique value proposition. Focus strategy can work either for cost leadership or differentiation for the given segment of customers.

Whatever the strategic positioning the company decides to choose, the company must implement such strategy to achieve its strategic goals. An excellent strategy means nothing without effective implementation for company performance. Efficient operation in pursuit of profit maximization is extremely important and necessary for a company's bottom line performance. But that is not sufficient for company's long-term success. Proper design and implementation of strategic positioning is key in constituting sustainable competitive advantage thus the long-term financial success of the company. Analysts therefore look at company's performance from these directions by way of financial statement analyses, market survey, personal interviews with executives, and other means for discovery.

## Quantitative Approach

Here I consider a quantitative approach in evaluating companies from two distinctive areas, namely finance and technology. Both financial and technology aspects of a company can be evaluated in either qualitatively and quantitatively but only the quantitative approach is taken here. Financial reporting is readily available for all publicly traded companies in developed countries and exhibit similar construct on a global scale. Financial reporting describes the past record of financial performance as a result of actions that companies took. Similarly, technical information is also historical data that provides information on areas of interest, actual investment, and technical achievement. Because technology development is done by companies as investment for future business benefits, technical assessment is important in assessing future direction of companies' businesses.

### FINANCIAL STATEMENT ANALYSIS

Financial reporting of a company takes the form of financial statements that typically comprise of the balance sheet, the income statement, and the statement of cash flow. These financial statements are typically reported once a year as the annual report and also on a quarterly report indicating company's results in a 3-month period. Publicly traded companies are required to create and submit audited financial statements to each of the respective government agencies such as Securities and Exchange Commission in the US for fair trade of financial instruments. The representation of a company by these financial statements does not exactly reflect the economic realities of the company but rather is an approximation at best[11]. However, it is the most widely accepted and available tool to assess the economic reality of a company. It is therefore important to recognize the usefulness as well as the limitation in the analysis of financial statements.

The balance sheet is a snapshot of company's assets and how such assets had been financed at the time of the report. The left-hand side of the balance sheet lists out assets in the order of liquidity. The right-hand side of the balance sheet lists out the sources of finance for the aforementioned assets that include debts and owners' equity. Since assets indicate future benefits of the company, the types of assets and the amount thereof can imply company's potential course of actions to a degree. Also, the short-term and long-term debt obligations can tell financial risks and interest burdens that may limit company's activities.

The income statement is a cumulative historical set of activity-based financial results that changes the contents of the balance sheet. Thus, income statement describes

---

[11] White Gerald I., Ashwinpaul C. Sondhi, and Dov Fried, *The Analysis and Use of Financial Statements*, John Wiley & Sons, Inc, 1994.

what factors affected the changes in the balance sheet. It records revenue and expenses at various stages of company activities under normal operations as well as income and expenses incurred from extraordinary events such as divestiture of major assets, changes in foreign exchange rates, and material rules changes in the accounting systems.

The statement of cash flow reports the cash movement made by the company's operation, investment activities, and financing activities. The statement of cash flow was added to the balance sheet and income statement because management of cash is extremely important for maintaining the company operation. It provides financial health of the company and suggests points of improvement on a cash basis, which is necessary to avoid bankruptcy.

All of these financial statements are interrelated and one must get a holistic view of the company utilizing all three as well as footnotes associated with them. Close examination and proper analyses require rigorous skills to achieve insightful understanding about the company[12] [13].

**CAPITAL MARKETS**

Equity and debt instruments as well as their derivatives of publicly listed companies are traded in capital markets where such trades determine the instantaneous prices. The act of trades is motivated by profit seeking investors who try to take advantage of any miss-priced opportunities to make money. Having numerous market participants looking for any arbitrage opportunities with intelligent economic analyses make the market efficient and a correct asset price is reached at any instance for the given information. At the same time, there exist certain abnormalities in the market trade. One is high-speed trading conducted by computers that run under certain

[12] Bodie, Zvi, Alex Kane, and Alan J. Marcus. *Investments, Second Edition*. Irwin, Inc. 1993.
[13] Richard A. Brealey, and Stewart C. Myers. *Principles of Corporate Finance, Fourth Edition*. McGraw-Hills, Inc. 1991.

algorithms set by human programmers. There have been incidents of over trading due to inappropriate algorithms run at extremely high speed. Another type of incident is caused by human traders who not only trade by intelligence-based judgment but emotional movements[14] [15]. Advanced research in behavioral science shed lights in suboptimal trading practices by human nature. These abnormalities can make the market inefficient and provide misprice at times. Nonetheless, market data provides quite meaningful data about the instruments being traded and the underlying economics of the company.

**FINANCIAL RATIOS**

Financial ratios utilize line items of financial statements as well as market exchange trade data to provide information for specific purposes. Each line item in financial statements has significant meanings that indicate certain aspects of company's financial performance. Two or more line items may be combined to produce ratios. Financial ratios are widely used to characterize company's state and its performance. Financial ratios are also used to compare multiple companies within and across industries.

For example, to assess company's financial stability, financial ratios that indicate solvency and liquidity may be used. Solvency-related indicators such as Debt to Equity Ratio, Debt to Asset Ratio, and Interest Coverage Ratio suggests company's long-term financial health indicating the degree of company's capability in long-term financial commitments. Liquidity-related indicators such as Current Ratio and Quick Ratio indicate the degree of company's ability to meet its short-term financial obligations.[16]

---

[14] Richard H. Thaler. The Winners Curse. The Free Press. 1992

[15] Richard H. Thaler, and Cass R. Sunstein. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press. 2008.

[16] Erich A. Helfert. *Techniques of Financial Analysis: A Practical Guide to Managing and Measuring Business Performance. Eighth Edition*. Irwin Professional Publishing. 1994

Certain financial ratios have skew depending on the type of business or the type of industry the company operates in. But overall, financial ratios are useful common measure of companies across industries.

## Evaluating Companies Using Technology Information

Companies in many industries thrive by innovation. Innovation leads to new products and services, and higher productivity that bring value to companies. Particularly in technology-based companies, technological innovation is the source of their competitive advantage and is the core foundation of their corporate value.

Recognizing the importance of superior technology basis, companies invest in research and development to garner technological advancement. Such technological advancement is manifested in the portfolio of intellectual properties. Some intellectual properties are kept as trade secret but significant portion of them is converted into intellectual property rights to be legally protected as exclusive rights to such new technologies and designs. Among various intellectual properties, patents are companies' selected inventions as a result of research and development activities and that had been carefully determined as worthwhile as important investment for their businesses.

As other intellectual properties, despite their significant potential contribution to the future success of companies, patents are completely underestimated as important assets in the balance sheet. Patents are intangible assets and widely deployed accounting standards such as Generally Accepted Accounting Principle (GAAP) and International Financial Reporting Standard (IFRS) do not capture the value of patents well due to their conservatism. Patents that are acquired as asset purchase or a part of business acquisition are valued at costs and capitalized as asset which tend to give fair value treatment. On the other side, internally developed technology and associated patents are typically expensed thus no capitalized assets are recorded. Even when they are capitalized, the value of asset is limited to the cost of patent administrative processes such as application fees and patent attorney's fees. This unrealistic accounting representation of patent values sometimes mislead the potential of companies in today's information based economy.

Patent information is publicly available from respective patent offices worldwide. There are multiple research companies that provide user-friendly interface to manage such information[17].

Both qualitative and quantitative approaches are taken in analyzing patents. To fully understand the quality of invention and viability of such underlying technology, the contents presented in each claim and embodiments need to be diligently examined. The examination requires deep technology domain knowledge as well as legal knowledge about patents. Such examination may require research far beyond the actual patent at hand and it may involve exploration of prior art in earlier patents, scientific research papers, industry articles, etc. Therefore, full examination and appreciation of patented invention is extremely demanding in the level of knowledge and skills and only limited number of such close examination can be afforded even at large organizations. Extensive qualitative examination is practiced in the case of patent infringement based law suit cases where extreme high price is at stake.

Quantitative approach sometimes compensates for full qualitative examination. The number of granted patents under licensing negotiation of a portfolio licensing plays an important role in determining the value of the whole portfolio. For example, when two parties contemplate on potential cross-licensing of patents in the electronics industry, thousands of patents may be included in the portfolio. In such event, only champion patents from each side are carefully examined to see the significance of technology and the rest will be judged by the number of patents in the portfolio as a

[17] Questel was used as a data source for the experiments in this dissertation.

compromise to full qualitative examination. The practice is done to manage costs of examination by each side.

# CHAPTER 3. SEMANTIC SPACE ANALYSIS METHODS

This chapter presents the core algorithmic and procedural ideas of the semantic space analysis methods. I first introduce the Mathematical Model of Meaning (MMM)[18] [19] as a basis for the applied analytical methods. From the MMM, I apply the core concept of semantic space construct and its subspace selection as a result of user contextual predication.

I then describe two distinctive methodologies for different applications. The first method is relevance-based analysis method used for corporate evaluation with the user context specification. The second method is similarity-based analysis method used to build market return prediction model for a financial asset in time series. I also describe a general procedure in expanding the semantic space both in the horizontal and the vertical directions.

## The Mathematical Model of Meaning (MMM)

The central theme of the Mathematical Model of Meaning proposed by Kiyoki and Kitagawa is context-dependent semantic computing. The MMM provides the

---

[18] Kitagawa, T. and Kiyoki, Y., "The Mathematical Model of Meaning and its Application to Multi-database Systems," Proc. 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993.
[19] Kiyoki, Y. and Kitagawa, T., "A metadatabase system for supporting semantic interoperability in multidatabases," Information Modeling and Knowledge Bases, IOS Press, Vol. V, pp. 287-298, 1994

fundamental platform to conduct analysis[20 21 22] The MMM creates a semantic space, give a user context to select a subspace in which the analysis is conducted.

The MMM consists of the following components and processes:

- Designing metadata space and creating a metadata base space M
- Constructing metadata space
- Conducting semantic association

## DESIGNING METADATA SPACE

The first step is an overall designing of a metadata space based on the project objectives. Appropriate metadata and respective feature words or attributes are chosen.

### *1. Create metadata*

A set of metadata can be created by various means depending on the subject matter and availability of associated data. Keywords that characterize the subject may be produced by domain experts. Certain processes that extract characteristics of the subject matter may produce a set of metadata. Such process may involve measurement of physical dimensions such as image sensor recognition or may be keyword extraction from a document with an algorithm such as tf-idf (term frequency with inverse document frequency).

### *2. Creating feature words*

A set of feature words shall be selected that explain or indicate attributes of the entire metadata set. A set of feature words may be constructed with words listed in a dictionary or a textbook. Care needs to be taken to align the abstract levels between the metadata and the feature words.

---

[20] T. Kitagawa and Y. Kiyoki, ``A new information retrieval method with a dynamic context recognition mechanism," Proceedings of 47th Conference of International Federation for Information and Documentation, pp.210-215, Oct. 1994.

[21] Y. Kiyoki and T. Kitagawa, ``A semantic associative search method for knowledge acquisition," Information Modelling and Knowledge Bases (IOS Press), Vol. VI, pp.121-130, 1995.

[22] Y. Kiyoki, T. Kitagawa and T. Hayama, ``A metadatabase system for semantic image search by a mathematical model of meaning," Multimedia Data Management -- using metadata to integrate and apply digital media --," McGrawHill(book), A. Sheth and W. Klas(editors), Chapter 7, 1998.

### *3. Designing a base matrix M*

Once metadata is defined, each metadata $d_i$ is explained by n number of feature words ($f_1$, $f_2$, … $f_n$) forming a vector:

$$d_i \text{ (i=1 to m)} = (f_{i1}, f_{i2}, \ldots f_{in})$$

The metadata vector $d_i$ is stacked in the row direction to form a base space *M* (entry word set) while the feature words are thus extended in columns (feature word set). If the number of metadata is m and the number of attributes is n, the metadata space is expressed as an m by n matrix. The matrix M is normalized by 2-norm[23].

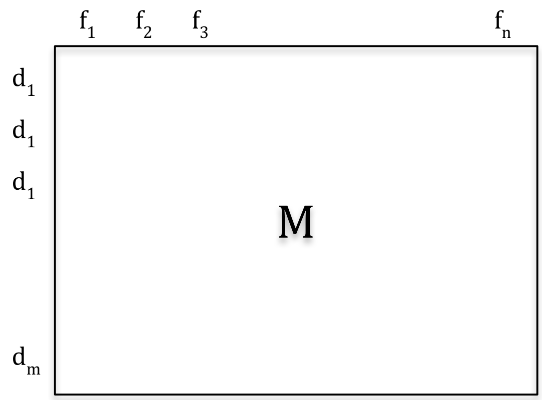



**Figure 1.** A Metadata Base Space M

### *4. Take the correlation matrix $M^TM$*

### *5. Conduct eigenvalue decomposition of the correlation matrix MTM such that:*

---

[23] Y. Kiyoki, T. Kitagawa and T. Hayama, ``A metadatabase system for semantic image search by a mathematical model of meaning, ACM SIGMOD Record, (refereed as the invited paper for special issue on metadata for digital media), Vol.23, No. 4, pp.34-41, Dec. 1994.

$$M^TM = Q \begin{pmatrix} \lambda_1 & & & & & & \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_v & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} Q^T$$

where $\lambda_i's$ are eigenvalues all in real numbers with $0 \leq v \leq n$. $Q$ is an orthogonal matrix and is defined as

$$Q = (q_1, q_2, q_3, \ldots, q_n)^T$$

Where $q_i$'s are the normalized eigenvectors of $M^TM$. Because $M^TM$ is symmetric, all of the eigenvectors are orthogonal to each other. I refer the eigenvectors $q_i$'s as "semantic elements" hereafter.

***6. Define the metadata space MDS***

I define the metadata space as below:

$$MDS := \text{span}(q_1, q_2, q_3, \ldots, q_v)$$

Thus MDS is defined as linear combinations of $(q_1, q_2, q_3, \ldots, q_v)$.

### DEFINING THE SEMANTIC PROJECTION SET $\Pi_v$

I define the projection function $P_{\lambda_i}$ as the projection of *MDS* to an eigenvector space that corresponds to a given eigenvalue $\lambda_i$.

$$P_{\lambda_i}: MDS \to span(q_i)$$

I define the semantic projection set $\Pi_v$ as follows:

$$\Pi_\upsilon := \left\{ \begin{array}{c} 0,\ P_{\lambda_1},\ P_{\lambda_2},\ \cdots,\ P_{\lambda_\upsilon}, \\ P_{\lambda_1}+P_{\lambda_2},\ P_{\lambda_1}+P_{\lambda_3},\ \cdots,\ P_{\lambda_{\upsilon-1}}+P_{\lambda_\upsilon},\ P_{\lambda_1}+P_{\lambda_2}+P_{\lambda_3},\cdots, \\ P_{\lambda_1}+P_{\lambda_2}+P_{\lambda_3}\cdots,\ P_{\lambda_{\upsilon-1}}+P_{\lambda_\upsilon} \end{array} \right\}$$

Because there are $\frac{\nu(\nu-1)\cdots(\nu-i+1)}{i!}$ semantic elements for the number of dimensions $i \; where \; i = 1, 2, \dots \nu$,, there exist $2^{\nu}$elements in the projection set $\Pi_\upsilon$. This implies that a maximum of $2^{\nu}$ different contexts can be expressed.

### DEFINING SEMANTIC OPERATOR $S_P$

I develop the semantic operator $S_P$. Let $s_k$ denote a set of k number of contextual search words:

$$s_k = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots \boldsymbol{u}_k)$$

where $\boldsymbol{u}_i$ represents a search word and is a subset of words defined in the base matrix *M*. $\boldsymbol{u}_i$ is thus presented in a vector form consisting of the feature word elements in the base matrix *M*. Let $T_k$ be the entire set of k number of search words thus $T_k \ni s_k$.

I defined the semantic projection set $\Pi_\upsilon$ in the previous section. $S_P$ is an operator that projects contextual search words onto the semantic element space. $S_P$ therefore maps each $s_k$ to $P_{\varepsilon_S}(s_k)$.

$$S_P: T_k \to \Pi_\nu \quad , \Pi_\nu \ni P_{\varepsilon_S}(s_k)$$

I introduce a threshold $\varepsilon_S$ for the projection where $0 < \varepsilon_S < 1$. $\varepsilon_S$ is used to select semantic element that is adequately relevant for the semantic operation $S_P$. I describe the procedure of the semantic operation below in steps.

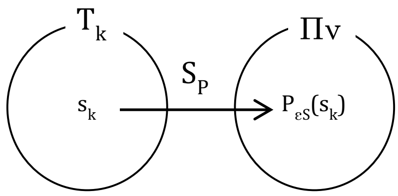



**Figure 2.** Projection by Semantic Operator

### CONDUCT FOURIER EXPANSION OF $\boldsymbol{u_i}$ (I = 1, 2, …, K)

To compute Fourier expansion, we take the inner product of $\boldsymbol{u}_i$ and $\boldsymbol{q}_j$ to produce $u_{ij}$.

$$u_{ij} \coloneqq (\boldsymbol{u}_i \ , \ \boldsymbol{q}_j)$$

where $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, \nu$.

I define a vector $\widehat{u_i}$ as the projection of contextual search word onto the semantic space *MDS*. Thus $\widehat{\boldsymbol{u_i}} \in MDS$ while $\boldsymbol{u}_i \in M$.

$$\widehat{u_i} \coloneqq (u_{i1,} u_{i2,} \ u_{i3,} \ \ldots, \ u_{iv,})$$

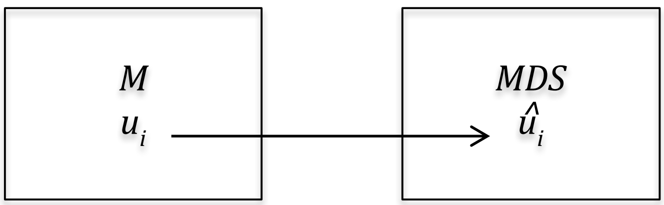



**Figure 3.** Projecting Context Words onto MDS

### FIND THE SEMANTIC CENTER OF GRAVITY

I find the semantic center of gravity $\mathbf{G}^+(s_k)$ from the sequence of contextual search words $s_k$ as follows:

$$\mathbf{G}^{+}(s_k) := \frac{\sum_{i=1}^{k} u_{i1}, \sum_{i=1}^{k} u_{i2}, \dots, \sum_{i=1}^{k} u_{i\nu}}{\|\sum_{i=1}^{k} u_{i1}, \sum_{i=1}^{k} u_{i2}, \dots, \sum_{i=1}^{k} u_{i\nu}\|\infty}$$

where $\|*\|\infty$ denotes the infinity norm.

The numerator of $\mathbf{G}^{+}(s_k)$ is a $\nu$ dimensional vector whose each element is a sum of attributes for all *i* from 1 to k. The denominator of $\mathbf{G}^{+}(s_k)$ is the infinity norm of the numerator and is the selection of the largest value. Therefore, $\mathbf{G}^{+}(s_k)$ is a normalized $\nu$-dimensional vector that has the semantic center represented by semantic element weighted by the sum of all contextual words projected onto the semantic elements.

**THE SEMANTIC PROJECTION WITH THRESHOLD ON SEMANTIC ELEMENT ADOPTION**

I introduce a threshold parameter $\varepsilon_S$ $(0 < \varepsilon_S < 1)$. $\varepsilon_S$ is used to select relevant semantic element space with respect to the given contextual search words. For a given $\varepsilon_S$, only the semantic elements with the respective projected absolute value of $\mathbf{G}^{+}(s_k)_i$ greater than $\varepsilon_S$ are selected and adopted for the final projection.

$$P_{\varepsilon_S}(s_k) := \sum_{i \in \Lambda_{\varepsilon_S}} P_{\lambda_i} \quad \in \Pi_\nu$$

where $\Lambda_{\varepsilon_S} := \{i \,|\, |\mathbf{G}^{+}(s_k) i| > \varepsilon_S\}$.

**CONDUCTING SEMANTIC ASSOCIATION**

In a given semantic subspace generated by the user-provided contextual search words, I take the norm of each projected object to be the representative value of such object. The size of norm is a manifestation of the degree of relevant association of an object with respect to the given set of contextual search words. Special care is required in the calculation of the norm to correctly incorporate the direction of projected vectors.

# Semantic Space Analysis

## RELEVANCE-BASED ENTITY EVALUATION

I describe the methodology in examining a set of entities according to the user context. In this analytic method, I create a semantic space by defining a set of "Characteristic Parameters," each of which comprises a dimension in a multi-dimensional semantic space and is the basis of semantic spatial analysis. I map each entity on to the semantic space as well as the user context both characterized by the Characteristic Parameters. The user context defines a semantic subspace in which intended analysis and evaluation is conducted for each of the entities under examination.

There are multiple options in selecting a subspace based on the user context, representing each entity in the selected subspace, and analyzing and evaluating the entity depending on the user objective. A careful analysis should be conducted in choosing such options to achieve the intended purpose. I describe below a typical methodology in steps for the relevance-based entity evaluation which I used for user context-dependent corporate analysis and evaluation system I implemented.

### *1. Creating a Semantic Space by Characteristic Parameters*

I construct a semantic space by defining a set of Characteristic Parameters that well describe the characteristics of the entities I want to evaluate. There is no limitation in the number of Characteristic Parameters but should be sufficient in describing the entities and user context.

Characteristic Parameters are fundamental elements for the construct of a semantic space. Because entities exhibit certain characteristics, Characteristic Parameters are chosen both in quality and quantity to sufficiently describe such entities with the intent for characterization. At the same time, the number of parameters must meet economic viability as a constraint for a practical application. A Characteristic Parameter consists a dimension of the semantic space.

Characteristic Parameters may be categorized into domains such as economy, business, technology, etc. Each domain needs to be described by a concrete set of Characteristic variables that are numerically measurable in the data type of ratio.

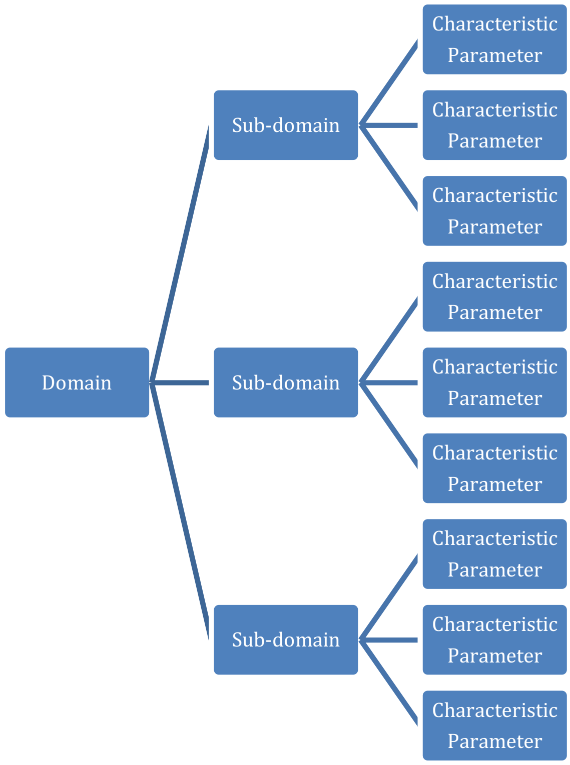



**Figure 4.** Multi-layer construct of Semantic Space

***2. Describing Each Entity by Characteristic Parameters***

Now that a semantic space has been constructed by the set of Characteristic Parameters, each entity is described as a vector whose elements correspond to the values of Characteristic Parameters. Suppose the semantic space comprises of n dimensions with n number of Characteristic Parameters, such vector that describe an entity has n elements. An Entity Vector (**EV**) is then consists of n numbers of Characteristic Parameters.

$$\mathbf{EV} \in R^n$$

$$\mathbf{EV} = ( cp_1 \, cp_2 \; \dots \; cp_i \; \dots \; cp_n )$$

where $cp_j$ is a Characteristic Parameter of $j^{\text{th}}$ dimension.

Multiple entities can be plotted into the semantic space using the same Characteristic Parameters. Suppose there are $x$ number of entities to be evaluated, there will be $x$ numbers of entity vectors.

$$\mathbf{EV}_1 = ( cp_{1,1} \;\; cp_{2,1} \;\; cp_{3,1} \;\; \dots \;\; cp_{1,n} )$$
$$\mathbf{EV}_2 = ( cp_{1,2} \;\; cp_{2,2} \;\; cp_{3,2} \;\; \dots \;\; cp_{2,n} )$$
$$\dots$$
$$\mathbf{EV}_i = ( cp_{i,1} \;\; cp_{i,2} \;\; cp_{i,3} \;\; \dots \;\; cp_{i,n} )$$
$$\dots$$
$$\mathbf{EV}_x = ( cp_{x,1} \;\; cp_{x,2} \;\; cp_{x,3} \;\; \dots \;\; cp_{x,n} )$$

where $cp_{ij}$ is a Characteristic Parameter element of the $i^{\text{th}}$ entity in the $j^{\text{th}}$ dimension.

I now define from the set of entity vectors $\mathbf{EV}_i$ an x by n matrix which I call the Entity Matrix (**EM**) in the following manner. The row represents the entity while the column represents the dimension of Characteristic Parameters, or features. In order for Characteristic Parameters to be independent, orthogonality needs to be secured by means of eigenvalue decomposition or a comparable process.

$$\mathbf{EM} = \begin{bmatrix} cp_{1\,1} & cp_{1\,2} & \dots & cp_{1\,n} \\ cp_{2\,1} & cp_{2\,2} & \dots & cp_{2\,n} \\ \vdots & \vdots & \vdots & \vdots \\ cp_{x\,1} & cp_{x\,2} & \dots & cp_{x\,n} \end{bmatrix}$$

***3. Normalizing Characteristic Parameters***

Characteristic Parameters take different values with a variety of ranges of absolute values. In order to compare values in different units and to avoid skew toward larger absolute values, for the analysis and evaluation of entities, normalization of data in each dimension is needed. Typical normalization methods are described below:

(1) Rescaling (Characteristic Vector Dimension)

Take the minimum value as 0 and maximum value as 1 in each dimension thus setting the range between 0 and 1.

$$x' = \frac{x - min(x)}{\max(x) - \min(x)}$$

(2) Standardization (Characteristic Vector Dimension)

Use standard deviation as a unit in each dimension and measure the deviation from the mean. The resulting value 0 indicates the mean and 1 as the standard deviation.

$$x' = \frac{x - average(x)}{\sigma}$$

(3) Unit Length (Per Entity)

Divide the Entity Vector by its Euclidean norm such that the range is scaled between -1 and 1 for each entity.

$$x' = \frac{x}{\|x\|}$$

I now define a normalized Entity Matrix **EM**'

$$\mathbf{EM}' = \begin{bmatrix} cp'_{1\,1} & cp'_{1\,2} & \dots & cp'_{1\,n} \\ cp'_{2\,1} & cp'_{2\,2} & \dots & cp'_{2\,n} \\ \vdots & \vdots & \vdots & \vdots \\ cp'_{x\,1} & cp'_{x\,2} & \dots & cp'_{x\,n} \end{bmatrix}$$

***4. Translating User Context to Construct Context Vector***

User context must be represented in terms of Characteristic Parameters. To prepare for the case where the user is unable to translate the user-defined context into the representation by Characteristic Parameters, I propose a system that translates the user context to Characteristic Parameters. The system incorporates necessary expertise to accomplish such translation task and alleviates the burden on the user. The user can specify her context such as objectives and conditions for the entity evaluation by layman's terms.

Expressing the user context in terms of Characteristic Parameters mean assigning weight on each dimension of Characteristic Parameters. This assignment leads to constructing a Context Vector, **CV**. The Context Vector is used to define a semantic subspace in which analysis and evaluation will be conducted on each entity.

### *5. Defining a Semantic Subspace*

The user Context Vector leads to defining a semantic subspace that properly reflects the interest of the user. A strict subspace selection is important because it significantly eliminates noisy components in Entity Vectors that are irrelevant with respect to the user context. Certain degree of freedom in the subspace selection process exists depending on the strictness to be represented by the user context as described below:

1. Select the Characteristic Parameters that has the highest absolute value in Context Vector.
2. Select certain number of Characteristic Parameters that exhibit the highest to the nth largest absolute values in the Context Vector.
3. Set forth a threshold level and select all dimensions of Characteristic Parameters whose components are above such threshold in the Context Vector.

### *6. Projecting Each Entity onto the Semantic Subspace*

For the selected k-dimensional subspace $\mathbf{U} = \mathrm{span}(u_1, u_2, \ldots u_k)$, then an Entity Vector $\boldsymbol{v}$ can be projected onto the subspace U as follows:

$$\boldsymbol{v}||\boldsymbol{U} = \sum_{i=1}^{k} \frac{(\boldsymbol{v}, \boldsymbol{u}_i)}{(\boldsymbol{u}_i, \boldsymbol{u}_i)} \boldsymbol{u}_i$$

where $(\boldsymbol{x}, \boldsymbol{y})$ indicates the inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

***7. Calculating the Norm for Each Projected Entity Vector*** $\boldsymbol{v}_i$

The norm of each of the projected entity vectors on the semantic subspace indicates the relevance of the entity for the given user context. I use the norm as a score for each entity and rank them accordingly. The larger number of such relevance score indicates higher relevance of the corresponding entity with respect to the specified user context.

## Time Series Analysis

I now turn to an application of semantic space analyses to the future instance prediction based on historical instances in time series.

### APPLICATION OF SEMANTIC SPACE ANALYSES TO PREDICTIVE MODEL

I present that the semantic space analysis method can be applied to create a predictive model of a future instance based on historical instances in time series. A semantic space is constructed with sufficient dimensions of Characteristic Parameters. The quantity and the quality of Characteristic Parameters are defined by the purpose of the analyses. Each instance is described by the defined Characteristic Parameters and mapped onto the semantic space. When the entire instances in the time series are mapped, the semantic space contains all relevant historical data.

The user context may be introduced to define user objectives and conditions for the analyses and evaluation of the instance entities. The context selects corresponding semantic subspace. Two arbitrary instances in the time series may be picked up and the degree of similarity may be examined by the distance calculation. The higher the similarity, the lower the distance value.

One of such instance to be chosen as a current instance and the distance between the current instance and all the historical instances may be calculated to find the least distance value for the most similar historical instance. The next instance in time series of the identified similar instance indicates the prediction of the future from the current instance.

### GENERAL PROCEDURE

***1. Constructing a Sematic Space using Characteristic Parameters***

A semantic space may be constructed with a range of Characteristic Parameters that are associated with the instances in time series. Characteristic Parameters may include domains such as macroeconomics, industry-specific indicators, market indicators, price indexes, etc.

***2. Map Each Instance onto the Semantic Space***

Each entity is a historical state that is periodically interspersed and is described as a set of numerical values of Characteristic Parameters. The current state, which is specially identified as the user context, is also described in the same way.

***3. Specify User Context to Select Semantic Subspace***

A user context may be introduced to select a subspace, which reduces the semantic space to limit the analyses only to the relevant aspects of the instance entity which results in savings of computational requirements. Subspace selection may be accomplished by multiple options as described earlier in this chapter.

***4. Calculate Distance for the Most Similar Instance***

The Euclidean distance (D) between two arbitrary n-dimensional vectors **p** = $(p_1, p_2, p_3, \ldots p_n)$ and **q** = $(q_1, q_2, q_3, \ldots q_n)$ is calculated as below:

$$D(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

The Euclidean distances between the current instance and all historical instances are calculated and the minimum distance values are sought. The instance time index with the minimum distance is identified.

***5. Predict the Future Instance***

The state of the next instance of the identified most similar instance is taken as the predicted state of the future next to the current instance.

**A NOTE ON PROJECTION AND VECTOR DISTANCE**

A projection of a vector $\boldsymbol{v}$ onto a subspace extracts the components of $\boldsymbol{v}$ in the span of the subspace, discounting the orthogonal components of $\boldsymbol{v}$. Therefore, the relative angle between the vector $\boldsymbol{v}$ and the subspace is significant. The norm of the projected vector onto the subspace becomes zero when $\boldsymbol{v}$ is orthogonal to the subspace. The norm of the projected vector onto the subspace becomes the same as that of $\boldsymbol{v}$ when $\boldsymbol{v}$ is a span of the subspace.

The distance D between the two vectors indicates the measure of similarities between the two n-dimensional vectors. The distance is affected by the angle between the vectors but is not significant as in the case of projection. The distance becomes zero when the two vectors are identical.

## Semantic Space Expansion

### MULTI-DIMENSIONAL SPACE CONSTRUCTION

In order to conduct semantic space analyses, such semantic space must first be constructed. Such semantic space needs to be flexible such that a new domain of dimensions may be added (horizontal expansion) and a dimension may be broken down to more detailed dimensions to explain an entity more properly (vertical expansion).

### HORIZONTAL EXTENSION OF SEMATIC SPACE

I consider a case to combine two metadata base space matrixes that do not have any mutual links between them. Referring to the left picture in Figure 9, combining the matrices M-A and M-B fall in this category. I apply the basic integration methodology of heterogeneous matrixes introduced by Kiyoki and Ishihara[24]. I concatenate each of the entry word set vertically and corresponding feature word set horizontally taking into account any duplication or synonymy issues in the set of feature words as well as the entry words. We then fill in the blank entries in the upper right section of the new matrix for feature words that correspond to the entry words of the M-A matrix. Likewise, we fill in the blank entries in the lower left section of the new matrix for feature words that correspond to the M-B' matrix, completing the new combined metadata base space matrix.

---

[24] Y. Kiyoki and S. Ishihara: "A Semantic Search Space Integration Method for Meta-level Knowledge Acquisition from Heterogeneous Databases," Information Modeling and Knowledge Bases (IOS Press), Vol. 14, pp.86-103, May 2002.

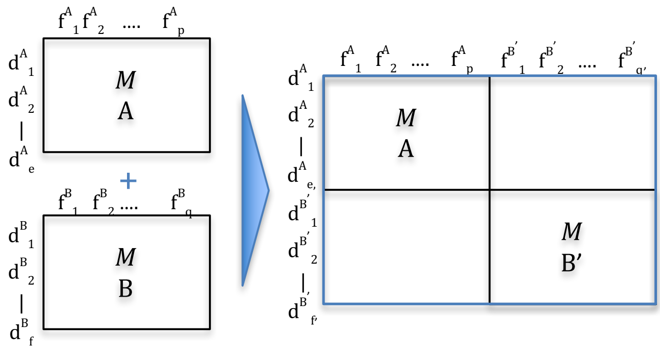



**Figure 5.** Combining two Metadata Base Space Matrices with no Mutual Links

### Vertical Extension of Sematic Space

I examine possible structures of multi-layered hierarchical metadata base spaces by looking at topological relationships between metadata base spaces that are directly connected. We define a parent metadata base space and a child metadata base space. A parent metadata base space exists higher in semantic hierarchy relative to a child metadata base space and the parent and the child metadata base spaces are directly connected. Direct connection means that at least one of the feature words of the parent metadata base space is an entry word of the child metadata base space. We allow an entry word to exist in multiple metadata base spaces.

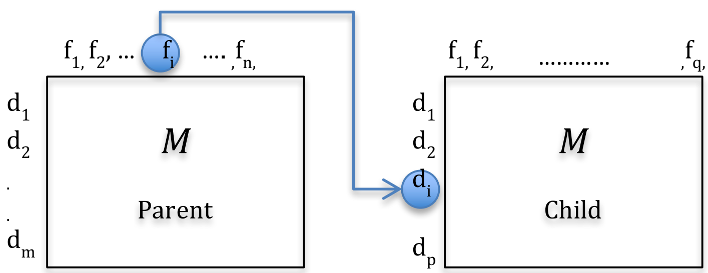



**Figure 6.** The Relationship between a Patent and a Child Metadata Base Space

### PATTERNS OF MULTI-LAYERED HIERARCHICAL METADATA SPACES

I recognize that there are three basic patterns of inter-relationships between metadata base spaces. As defined above, a parent-child relationship is established by a connection link between a feature word of a parent and an entry word of a child.

Pattern-A is a stand-alone case where there is no child metadata base space. This implies that this metadata base space is the terminal space located at the lowest end of technology chain in the semantic hierarchy. Therefore none of the feature words of this metadata base space matches with any entry words of all other metadata base spaces.

Pattern-B is a case where a parent metadata base space has one or more child metadata base spaces. The link or links between the patent and the child or children may be originated by the patents single feature word or multiple feature words.

Pattern-C is a case where more than two patent metadata base spaces exist for a given child metadata base space.

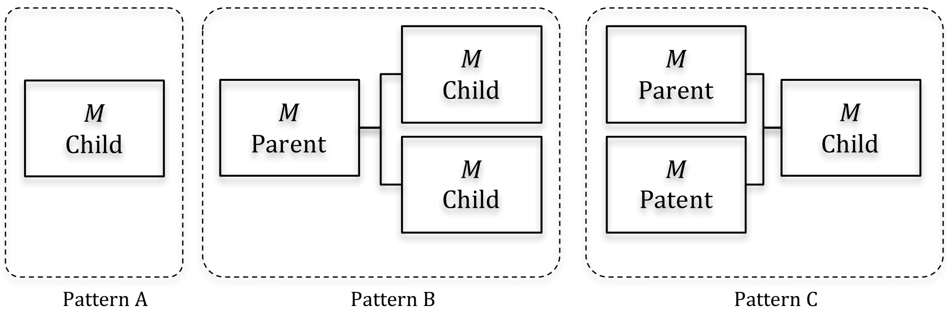



**Figure 7.** Patterns of Parent-Child Relationships between Metadata Base Spaces

### REPRESENTATION OF METADATA BASE SPACE SYSTEMS IN MATRICES

In this section, we consider representing all patterns of the parent-child metadata base space matrices that were presented in 2.1.

***Combining a Parent Metadata Base Space Matrix with a Child Metadata Base Space Matrix – a Simple Case***

We first consider the simplest case – combining a parent metadata base space matrix with a child metadata base matrix as depicted in Figure 6. To proceed, we expand the parent matrix with the subspace expansion matrix as follows.

Referring to Figure 8, we recognize that the feature word column vector $f_i$ of the parent metadata base space matrix that originates the link to the child metadata base space matrix is an m x 1 matrix. We also recognize that the entry word row vector in the child space matrix that terminates the link is a 1 x q matrix. We define the Subspace Expansion Matrix (SEM) by multiplying these two matrices as follows:

$$SEM \quad := \quad \begin{pmatrix} d_1 f_i \\ d_2 f_i \\ d_3 f_i \\ \vdots \\ d_m f_i \end{pmatrix} \begin{pmatrix} f_i f'_1 & f_i f'_2 & \cdots & f_i f'_q \end{pmatrix}$$

Where $\begin{pmatrix} d_1 f_i \\ d_2 f_i \\ d_3 f_i \\ \vdots \\ d_m f_i \end{pmatrix}$ is the column vector of the feature word of a parent metadata base space matrix that originates the link to a child metadata base space matrix, and $\begin{pmatrix} f_i f'_1 & f_i f'_2 & \cdots & f_i f'_q \end{pmatrix}$ is the row vector of the entry word in the child metadata base space matrix that terminates the link from the parent space matrix.

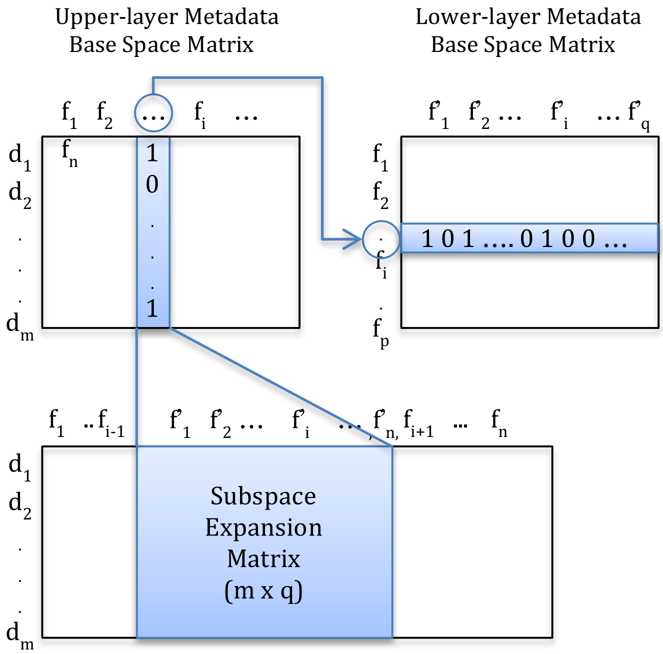



**Figure 8.** Subspace Expansion Matrix

### *Combining two Metadata Base Space Matrices without any Mutual Links*

We consider a case to combine two metadata base space matrixes that do not have any mutual links between them. Referring to the left picture in Figure 9, combining the matrices M-A and M-B fall in this category. We apply the basic integration methodology of heterogeneous matrixes introduced by Kiyoki and Ishihara. We concatenate each of the entry word set vertically and corresponding feature word set horizontally taking into account any duplication or synonymy issues in the set of feature words as well as the entry words. We then fill in the blank entries in the upper right section of the new matrix for feature words that correspond to the entry words of the M-A matrix. Likewise, we fill in the blank entries in the lower left section of the new matrix for feature words that correspond to the M-B’ matrix, completing the new combined metadata base space matrix.

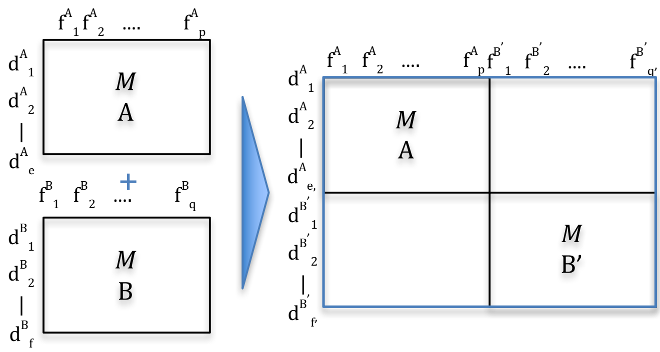



**Figure 9.** Combining two Metadata Base Space Matrices with no Mutual Links

### *Combining a Parent Metadata Base Space Matrix with a Child Metadata Base Space Matrix – a General Case*

The patterns described in Section 2.1 exhaust all parent-child relationships of metadata base spaces. We now present in Figure 9 a structure of multiple metadata base spaces that incorporates all the aforementioned patterns. The left diagram of Figure 9 shows a topological representation of parent-child relationships. The right diagram of Figure 9 indicates exemplifying links from feature words of parent matrices to entry words of child matrices forming the parent-child relationships.

Referencing the left diagram in Figure 9, M-C and M-D exemplify Pattern A of Figure 7. The relationships M-A to M-C, M-B to M-D, and M-B to M-D represent Pattern B while the relationships between M-A to M-C and M-B to M-C represent Pattern C.

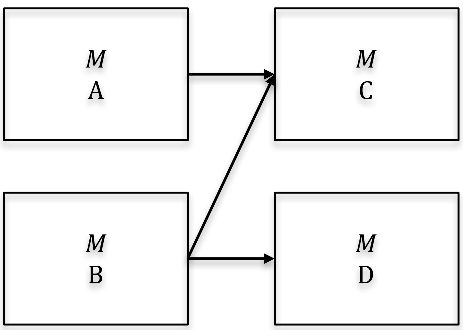



**Figure 10.** A General Topological Configuration of Parent-Child Metadata Base Space Combination

We now represent the configuration above in a matrix form. We assign the entry words and the feature words for each of the matrices M-A, M-B, M-C, and M-D as follows.

| | <Entry Words> | <Feature Words> |
|---|---|---|
| M-A: | $d_1^A, d_2^A, \cdots, d_e^A$ | $f_1^A, f_2^A, \cdots, f_p^A$ |
| M-B: | $d_1^B, d_2^B, \cdots, d_f^B$ | $f_1^B, f_2^B, \cdots, f_q^B$ |
| M-C: | $d_1^C, d_2^C, \cdots, d_g^C$ | $f_1^C, f_2^C, \cdots, f_r^C$ |
| M-D: | $d_1^D, d_2^D, \cdots, d_h^D$ | $f_1^D, f_2^D, \cdots, f_s^D$ |

Figure 10 indicates the basic structure with the link information between the upper-layer matrices and the lower-layer matrices. We combine these four matrices into a single matrix so as to treat the resulting matrix as the metadata base space for the MMM. We will take the following procedures:

- Incorporate the lower-layer metadata base space matrix into the upper-layer metadata base space matrix by introducing subspace expansion matrices (SEMs)
- Integrate upper-layer metadata base matrices with no mutual links

***Incorporate the lower-layer metadata base space matrix into the upper-layer metadata base space matrix by introducing subspace expansion matrices (SEMs)***

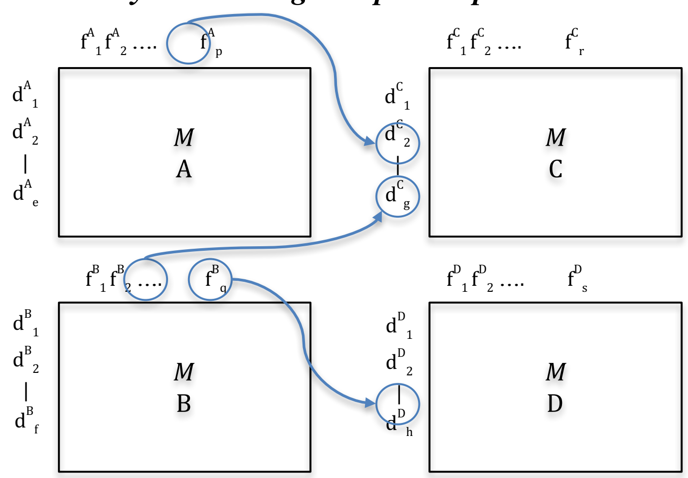



**Figure 11.** A General Configuration of Multi-layered Metadata Base Space Matrix

In Figure 11, there are three distinct links that connect upper-layer metadata base space matrices and lower-layer metadata base space matrices. Namely, these links are the following:

| | |
|---|---|
| From M-A to M-C: | The feature word of M-A $f^A{}_i$ is linked to the entry word of M-C $d^C{}_l$. |
| From M-B to M-C: | The feature word of M-B $f^B{}_j$ is linked to the entry word of M-C $d^C{}_m$. |
| From M-B to M-D: | The feature word of M-B $f^B{}_k$ is linked to the entry word of M-D $d^D{}_n$. |

For each of the links, we create Subspace Expansion Matrix (SEMs) and incorporate it into the upper-layered metadata base space matrix.

**The link from M-A($\boldsymbol{f_i^A}$) to M-C($\boldsymbol{d_i^C}$)**

$SEM_{AtoC}$ := $\begin{pmatrix} d_1^A f_i^A \\ d_2^A f_i^A \\ d_3^A f_i^A \\ \vdots \\ d_e^A f_i^A \end{pmatrix} \begin{pmatrix} d_i^C f_1^C & d_i^C f_2^C & \cdots & d_i^C f_r^C \end{pmatrix}$ is created where $f_i^A = d_i^C$, and the feature vector $\begin{pmatrix} d_1^A f_i^A \\ d_2^A f_i^A \\ d_3^A f_i^A \\ \vdots \\ d_e^A f_i^A \end{pmatrix}$ in the M-A matrix is replaced by the $SEM_{AtoC}$.

**The link from M-B($\boldsymbol{f_i^B}$) to M-C($\boldsymbol{d_i^C}$)**

$SEM_{BtoC}$ := $\begin{pmatrix} d_1^B f_i^B \\ d_2^B f_i^B \\ d_3^B f_i^B \\ \vdots \\ d_f^B f_i^B \end{pmatrix} \begin{pmatrix} d_i^C f_1^C & d_i^C f_2^C & \cdots & d_i^C f_r^C \end{pmatrix}$ is created where $f_i^B = d_i^C$, and the feature vector $\begin{pmatrix} d_1^B f_i^B \\ d_2^B f_i^B \\ d_3^B f_i^B \\ \vdots \\ d_f^B f_i^B \end{pmatrix}$ in the M-B matrix is replaced by the $SEM_{BtoC}$.

**The link from M-B($f_i^B$) to M-D($d_i^D$)**

$$SEM_{BtoD} := \begin{pmatrix} d_1^B f_i^B \\ d_2^B f_i^B \\ d_3^B f_i^B \\ \vdots \\ d_f^B f_i^B \end{pmatrix} \begin{pmatrix} d_i^D f_1^D & d_i^D f_2^D & \cdots & d_i^D f_s^D \end{pmatrix}$$ is created where $f_i^B = d_i^D$, and

the feature vector $\begin{pmatrix} d_1^B f_i^B \\ d_2^B f_i^B \\ d_3^B f_i^B \\ \vdots \\ d_f^B f_i^B \end{pmatrix}$ in the M-B matrix is replaced by the $SEM_{BtoD}$.

***Integrate upper-layer metadata base matrices with no mutual links***

We take the expanded matrices generated above and integrate them according to the process described in Section 2.2.2 resulting in the following integrated metadata base space matrix.

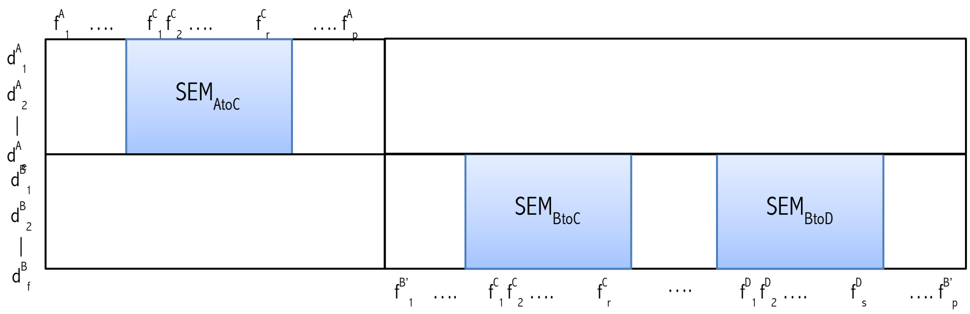



**Figure 12.** A General Integrated Metadata Base Space Matrix

### Chapter Summary

I introduced the Mathematical Model of Meaning as the core base of the context-dependent semantic space analysis methodologies. I described its application in entity evaluation as well as predictive modeling in time series. I also presented matrix expansion mechanism in the horizontal direction and vertical direction.

# CHAPTER 4. CONSTRUCTING SEMANTIC SPACES

In this chapter, I illustrate construction of semantic spaces in the domain of finance, technology, and brand for the purpose of evaluating companies. The essence of semantic space construction is the selection of Characteristic Parameters that sufficiently describe companies.

## CREATING FINANCIAL SEMANTIC SPACE

Company's state and their past performance can be described by financial statements. Financial statements are produced in compliance with the accounting principle and standardized rules, and all publicly traded companies are required to report to the authorities on an annual and quarterly basis. The line items in financial statements can be appropriate dimensions to describe companies.

Financial indexes, such as financial ratios can also serve the purpose of concise description of company states.[25] [26] [27] Financial ratios are typically derived from multiple line items of company's financial statements – the balance sheet, the income statement, and the statement of cash flow – as well as market data such as trading prices. Financial ratios can tell company's short-tern and long-term financial health, management efficiency, profitability, growth rates, relative valuation, etc. Typical financial ratios are classified and listed in the Table 1. Financial ratios are particularly useful for comparing companies within the same industry as some financial ratios exhibit different average values by industry.

---

[25] Bodie, Z., Kane, A., Marcus, A., "Investments Second Edition," Ch 19 Financial Statement Analysis Richard D. McGraw-Hill, Inc. 2009

[26] Brealey, R., Myers, S., "Principles of Corporate Finance Fourth Edition", Chapter 29 Financial Analysis and Planning, McGraw-Hill, Inc. 2003

[27] White, G., Sondhi, A., Fried, D., "The Analysis and Use of Financial Statements," 1994 John Wiley and Sons, Inc.

**Table 1.** Typical Financial Ratios

| | | |
|---|---|---|
| **Liquidity** | | |
| Current Ratio | Total Current Assets/ Total Current Liabilibies | Company's ability to meet its current liability obligations by its current assets |
| Quick Ratio | (Total Current Assets - Inventory)/ Total Current Liabilities | Company's ability to meet its current liability obligations by its liquid assets |
| **Solvency** | | |
| Debt to Equity | Total Liabilities/ Total Equity | Companies capital structure indication |
| Interest Coverage | (Earnings Before Interest & Taxes)/ Interest Charges | Company's ability to pay interest charges from operating profits |
| Debt Coverage | Total Long-term Debt/ Cash from Operations | Indicates the payback period for coverage of long-tern debt |
| **Profitability** | | |
| Gross Profit Margin | Gross Profit/ Net Sales | Company's profitability at the gross margin level suggesting its management effectiveness on inventory, buying power, and market power with its product |
| Net Profit Margin | Net Income/ Net Sales | Company's share of net revenue after paying all expenses including financing charges and taxes |
| Return on Assets | Net Income/ Total Assets | Company's ability to turn its assets into profits |
| Return on Equity | Net Income/ Shareholders' Equity | Company's ability to create profts for $1 invested by Shareholders |
| **Efficiency** | | |
| Accounts Receivable Turnover | Net Sales/ Accounts Receivable | Company's ability to collect cash from its customers on credit sales |
| Inventory Turnover | Cost of Goods Sold/ Inventory | Company's ability to turn inventory to sales |
| Sales to Total Assets | Net Sales/ Total Assets | Company's ability to generate sales on each dollar of assets |
| **Market Related** | | |
| Price Earnings Ratio | (Market Price/ Share)/ (Net Income/ Share) | Indicates company's market share price level in relation to its earnings |
| Price Book Ratio | (Market Price/ Share)/ (Book Value/ Share) | Indicates company's market share price level in relation to its shareholders' equity (book value) |

## Sources of Corporate Financial Data

There are numerous sources of financial data of companies that are publicly traded. The most notable source is the public depository of financial statements that companies submit on a regular basis. There are many other data sources that utilize such public data and publish information with analyses to provide certain insight about the company states.

## Characteristic Parameters based on Financial Information

Characteristic Parameters are driven by each stakeholder's role and objectives and their priority order. I list typical Characteristic Parameters by category as below.

**Table 2.** Characteristic Parameter Examples for the Finance domain

| | |
|---|---|
| Basic | Rev, EBITDA, EBIT, NI, Diluted EPS, Market Cap, Cash & Short-term Inv, TEV |
| Profitability | ROA, ROE, RO Common Equity |
| Margin | GM%, EBITDA%, EBIT%, NI% |
| Turnover | Total Asset Turnover, Fixed Asset Turnover, AR Turnover, Inventory Turnover |
| Liquidity | Current Ratio, Quick Ratio, Cash from Ops/Current Liabilities |
| Solvency | D/E, LTD/E, Total Lib/Total Assets |
| Growth-1 1-year | CAGR on Revenue, EBITDA, EBIT, NI, Cash from Ops., CapEX |
| Growth-3 3-year | CAGR on Revenue, EBITDA, EBIT, NI, Cash from Ops., CapEX |
| Growth-5 5-year | CAGR on Revenue, EBITDA, EBIT, NI, Cash from Ops., CapEX |
| TEV | TEV/Rev, TEV/EBITDA, TEV/EBIT |
| Volatility | Beta |
| Stability | Variance (Rev)/Rev, Variance (NI)/NI |

Above lists exemplify Characteristic Parameters and many more factors may be introduced to sufficiently describe the company state for the user objective. Orthogonality is assumed among all the Characteristic Parameters in constructing the semantic space.

### *Describing Companies by the Characteristic Parameters*

Each of the target companies under evaluation is described in terms of the selected Characteristic Parameter. That means acquiring financial information for each company, identifying necessary items such as line items in financial statements, calculate values that represent a Characteristic Parameter. As an example, ROA (Return on Asset) value is calculated by finding the net income in the income statement, finding the total asset value from the balance sheet, and make the division to come up with the ROA value.

### *Normalization*

Each Characteristic Parameter is unique in its measurement unit. In order to unify the measurement units in each dimension, each Characteristic Parameter values are normalized as described in Chapter 3. The range of Characteristic Parameter values as well as their signs (positive or negative numbers) must be individually considered to correctly normalize each dimension. Once normalization is complete for all Characteristic Parameters, the construct of the semantic space is complete.

## CREATING TECHNOLOGY SEMANTIC SPACE

Technology covers great many areas each of which has deep layers of sub technologies. To fully grasp the technological position of a company, rigorous qualitative research is required. However, to create a semantic space in the technology domain, numerical representation is necessary. It is quite difficult to quantitatively represent technological state of a company. I use patents as a basis of describing company's technological state. Patents are results of extensive research and development or costly acquisition form others. Companies make conscious decisions in selecting technology areas and individual technology to invest. Therefore, company's patent portfolio is a result of their investment. Thus examining patents can tell company's capabilities of certain technology areas and company's intent to make business in the applied fields.

### Sources of Patent Related Data

The Patent Office manages all phases of intellectual property rights including patents from application, prosecution, and grant or denial. The Patent Office manages such information for each intellectual property applied in their secure repository and make the information publicly available either the earlier of 18 months after the application date or patent grant date. The information is maintained on a permanent basis for public consumption on a global scale. There are private companies that utilize such information and provide value-added services. These service providers offer easy to navigate user interface, analytical tools such as heat maps, and services that have business implications.

**Characteristic Parameters based on Technology Information**

Here is a list of examples of technology-based Characteristic Parameters.

**Table 3.** Characteristic Parameter Examples for the Technology domain

| | |
|---|---|
| # Patents | Cumulative, Cumulative Currently Effective |
| # Patents | By year, By time period |
| # Patents | By IPC (International Patent Classification) – Multi-layer |
| # Patents | By country coverage |
| # Patents | By Internal Inventors, By Assignment |
| # Patents | Under registered license |
| # Patents | Granted per R&D Expenses |
| # FC[28] | By year, By time period |
| # FC | By IPC (International Patent Classification) – Multi-layer |
| # FC/# BC[29] | By IPC (International Patent Classification) – Multi-layer |
| R&D Exp. | By year, By time period |
| # Researchers | By year, By time period, By areas of expertise |
| # Papers [30] | By year, By time period, By scientific field |

*Normalization*

As was the case in the Characteristic Parameters in finance, Technology-based Characteristic Parameter values need to be normalized such that the differences in units as well as ranges in values are unified for integration.

**CREATING MULTI-DOMAIN SEMANTIC SPACE**

Here I describe construct of a multi-domain semantic space – in this example, finance, technology and brand. All of these domains are important in describing

[28] FC: Forward Citation – Citation made by other inventors and patent examiners, indicating a higher quality invention.

[29] BC: Backward Citation – Citation made by the inventor to other patents, indicating certain dependency on other patents

[30] Academic research papers published in journals

companies from multiple angles. Other domain factors may include company's social responsibilities, employee satisfaction, environmental conservation, etc.

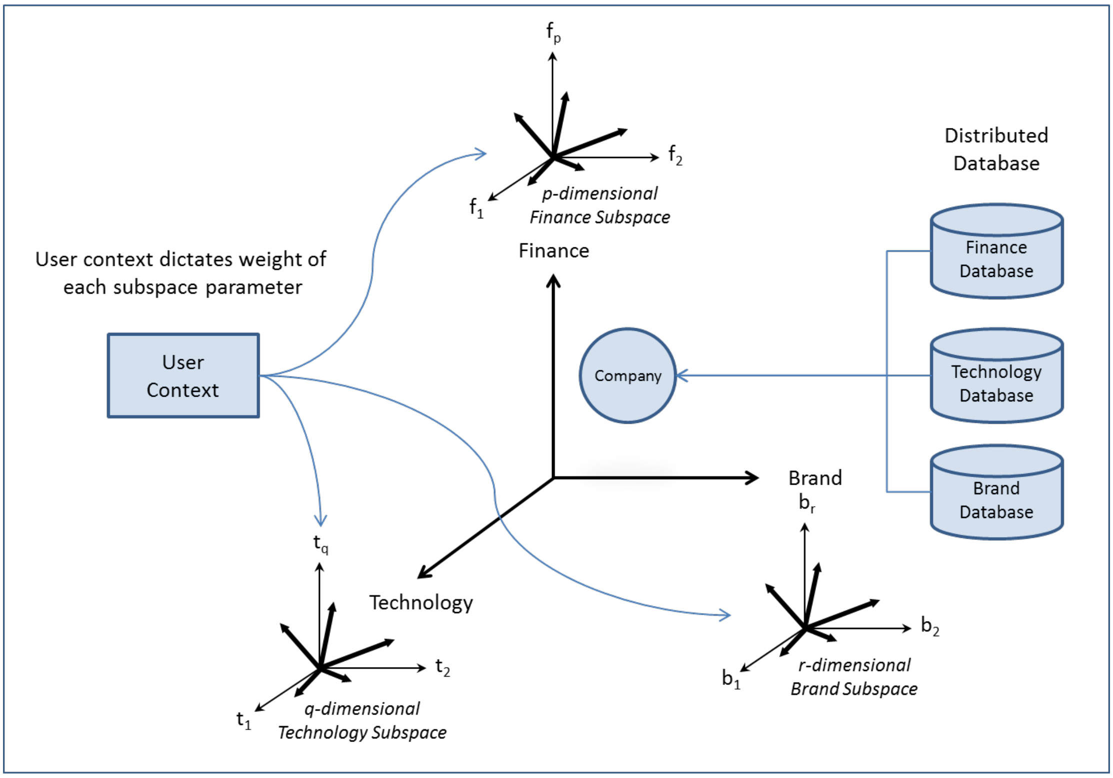



**Figure 13.** Conceptual Construct of a Multi-Domain Semantic Space

**Characteristic Parameters for Multi-Domain Semantic Space**

The Characteristic Parameter Matrix is one of the key elements of the Context-based Multidimensional Corporate Analysis Method. The Characteristic Parameter Matrix is an implementation of the distinct set of domain spaces (Finance, Technology, and Brand) that further consist of multi-dimensional parametric subspaces. The matrix contains i x j elements where i is the number of target companies determined by the Query Specifications and Context Specifications for conditions as described above, while j is the total number of parameters in all evaluation subspaces. Thus these elements in Characteristic Parameter Matrix serve as the basic representation of company characteristics.

As an illustration, I create Characteristic Parameter Matrix for three categories for m companies as below. The three categories may be finance, technology, and brand, for example.

For Domain-A: Finance

$$\left.\begin{array}{l}
CP_{comp\text{-}A} = (k_{domain\text{-}A\ 1\ 1}, k_{domain\text{-}A\ 1\ 2}, \dots\ k_{domain\text{-}A\ 1\ p}) \\
CP_{comp\text{-}B} = (k_{domain\text{-}A\ 2\ 1}, k_{domain\text{-}A\ 2\ 2}, \dots\ k_{domain\text{-}A\ 2\ p}) \\
CP_{comp\text{-}C} = (k_{domain\text{-}A\ 3\ 1}, k_{domain\text{-}A\ 3\ 2}, \dots\ k_{domain\text{-}A\ 3\ p}) \\
\quad \dots \\
CP_{comp\text{-}X} = (\underbrace{k_{domainA\text{-}m\ 1}, k_{domain\text{-}A\ m\ 2}, \dots\ k_{domain\text{-}A\ m\ p}}_{\textit{p elements}})
\end{array}\right\}\ \textit{m companies}$$

For Domain-B: Technology

$$\left.\begin{array}{l}
CP_{comp\text{-}A} = (k_{domain\text{-}B\ 1\ 1}, k_{domain\text{-}B\ 1\ 2}, \dots\ k_{domain\text{-}B\ 1\ q}) \\
CP_{comp\text{-}B} = (k_{domain\text{-}B\ 2\ 1}, k_{domain\text{-}B\ 2\ 2}, \dots\ k_{domain\text{-}B\ 2\ q}) \\
CP_{comp\text{-}C} = (k_{domain\text{-}B\ 3\ 1}, k_{domain\text{-}B\ 3\ 2}, \dots\ k_{domain\text{-}B\ 3\ q}) \\
\quad \dots \\
CP_{comp\text{-}X} = (\underbrace{k_{domain\text{-}B\ m\ 1}, k_{domain\text{-}B\ m\ 2}, \dots\ k_{domain\text{-}B\ m\ q}}_{\textit{q elements}})
\end{array}\right\}\ \textit{m companies}$$

For Domain-C: Brand

$$\left.\begin{array}{l}
CP_{comp\text{-}A} = (k_{domain\text{-}C\ 1\ 1}, k_{domain\text{-}C\ 1\ 2}, \dots\ k_{domain\text{-}C\ 1\ r}) \\
CP_{comp\text{-}B} = (k_{domain\text{-}C\ 2\ 1}, k_{domain\text{-}C\ 2\ 2}, \dots\ k_{domain\text{-}C\ 2\ r}) \\
CP_{comp\text{-}C} = (k_{domain\text{-}C\ 3\ 1}, k_{domain\text{-}C\ 3\ 2}, \dots\ k_{domain\text{-}C\ 3\ r}) \\
\quad \dots \\
CP_{comp\text{-}X} = (\underbrace{k_{domain\text{-}C\ m\ 1}, k_{domain\text{-}C\ m\ 2}, \dots\ k_{domain\text{-}C\ m\ r}}_{\textit{r elements}})
\end{array}\right\}\ \textit{m companies}$$

*Normalization*

Each element of the Characteristic Parameter Matrices is normalized according to its data characteristics. The normalized value has a range between -1 and +1. After the normalization, as the elements have become comparable, multiple matrices are combined to create an integrated Characteristic Parameter Matrix. Three matrices are horizontally combined to form an intermediate integrated data structure in the following example.

$$
\begin{array}{ll}
 & \begin{array}{ccc} \textit{Domain-A: p elements} & \textit{Domain-B: q elements} & \textit{Domain-C: r elements} \\ \textit{(Finance)} & \textit{(Technology)} & \textit{(Brand)} \end{array} \\
\begin{array}{l} CP_{comp\text{-}A} = \\ CP_{comp\text{-}B} = \\ CP_{comp\text{-}C} = \\ \\ CP_{comp\text{-}X} = \end{array} &
\left(\begin{array}{ccc}
(k_{domain\text{-}A\ 1\ 1}, \cdots\ k_{domain\text{-}A\ 1\ p}), & (k_{domain\text{-}B\ 1\ 1}, \cdots\ k_{domain\text{-}B\ 1\ q}), & (k_{domain\text{-}C\ 1\ 1}, \cdots\ k_{domain\text{-}C\ 1\ r}) \\
(k_{domain\text{-}A\ 2\ 1}, \cdots\ k_{domain\text{-}A2\ p}), & (k_{domain\text{-}B\ 2\ 1}, \cdots\ k_{domain\text{-}B\ 2\ q}), & (k_{domain\text{-}C\ 2\ 1}, \cdots\ k_{domain\text{-}C\ 2\ r}) \\
(k_{domain\text{-}A\ 3\ 1}, \cdots\ k_{domain\text{-}A\ 3\ p}), & (k_{domain\text{-}B\ 3\ 1}, \cdots\ k_{domain\text{-}B\ 3\ q}), & (k_{domain\text{-}C\ 3\ 1}, \cdots\ k_{domain\text{-}C\ 3\ r}) \\
\cdots & \cdots & \cdots \\
(k_{domain\text{-}A\ m\ 1}, \cdots\ k_{domain\text{-}A\ m\ p}), & (k_{domain\text{-}B\ m\ 1}, \cdots\ k_{domain\text{-}B\ m\ q}), & (k_{domain\text{-}C\ m\ 1}, \cdots\ k_{domain\text{-}C\ m\ r})
\end{array}\right)
\end{array}
$$

Depending on the selection of Characteristic Parameters, there may exist cross correlation between multiple parameters. When such correlation is suspected, eigenvalue decomposition is conducted to secure orthogonality between any pair of parameters for parametric independence.

A general description of an Integrated Normalized Characteristic Parameter Matrix can be given as below where the number of rows indicates the number of target elements to be evaluated and the number of columns indicates the number of Characteristic Parameters for all relevant domains.

$$
\mathrm{CP} = \begin{pmatrix}
k_{1,1} & k_{1,2} & k_{1,3} & \ldots k_{1,j} & \ldots k_{1,n} \\
k_{2,1} & k_{2,2} & k_{2,3} & \ldots k_{2,j} & \ldots k_{2,n} \\
 & & \ldots & & \\
k_{i,1} & k_{i,2} & k_{i,3} & \ldots \; k_{i,j} & \ldots k_{i,n} \\
 & & \ldots & & \\
k_{m,1} & k_{m,2} & k_{m,3} & \ldots k_{m,j} & \ldots k_{m,n}
\end{pmatrix}
$$

**Other Choices for Characteristic Parameters**

There exist numerous factors that are important in characterizing company competitiveness and longevity. One obvious factor is company's selling capabilities of their goods and services – activities that turn their assets into cash. There are other factors that exist within companies but very difficult to precisely identify and define as characteristic parameters. Further research in this area may lead to indicators that truly represent company strengths that are hidden from the surface.

## CHAPTER SUMMARY

I described how to construct semantic spaces in different domains. Proper choice of Characteristic Parameters is essential in each case for the purpose of analyses and evaluation of entities. Sufficient Characteristic Parameters need to be adopted to describe the entities of interest as well as the user context. Normalization of Characteristic Parameters is typically necessary to eliminate different units and value ranges of the Characteristic Parameters.

# CHAPTER 5. EXPERIMENTS ON CONTEXT-DEPENDENT CORPORATE EVALUATION IN SEMANTIC SPACE

In this chapter, I present two cases of context-dependent corporate evaluation experiments and demonstrate the viability of the methodology. The first case I present is in a financial semantic space with US-based semiconductor companies. Contexts are set as different purposes of stakeholders of the companies. I also present a case for large multi-national companies using multi-domain semantic space.

## Case 1: Context-dependent Corporate Evaluation in Financial Semantic Space

The first experiment is conducted to demonstrate the viability of a context-dependent corporate evaluation method in a single-domain multidimensional semantic space.

### OBJECTIVE OF THE EXPERIMENT

The objectives of this experiment are twofold. The first objective is to see how companies are evaluated and ranked differently with respect to the different user context set forth. The second objective is to test the viability of the methodology.

### EXPERIMENTAL PROCEDURE

The experiment was conducted in the following steps.

#### *Step 1: Design Financial Semantic Space with Characteristic Parameters*

I constructed a financial semantic space with financial ratios that are widely used to evaluate companies. Financial ratios represent corporate financial conditions[31] such as liquidity, solvency, growth, profitability, management efficiency, stock price levels, etc.[32] [33] in standardized formula making fair comparison among companies. The financial ratios are broken down to components of the formula and each component is translated to a line item of various financial statements. I considered over a hundred financial ratios and chose thirty-two as Characteristic Parameters.

***Step 2: Collect Raw Data and Create Characteristic Parameters***

I chose fourteen companies and collected data from financial statements during the years from 2003 to 2008 cited.[34] [35] Financial ratios as a representation of Characteristic Parameter was broken down to line items in financial statements – the balance sheet, the income statement, and the statement of cash flow – as well as line items in financial market data. After the raw data were acquired, they were combined to create financial ratios. Once the financial ratio data become available for all companies listed, they are normalized to fit between the values of (0, 1) by taking linear translation between the minimum and maximum values. Each value indicates company's relative position in financial criteria represented by the financial ratio.

---

[31] Damodaran, A., "Investment Valuation," John Wiley and Sons, Inc. 1996.
[32] Bodie, Ziv, Alex Kane, and Alan J. Marcus. Investments, Second Edition. Richard D. Irwin,Inc. 1993.

[33] White, G., Sondhi, A., Fried, D., "The Analysis and Use of Financial Statements," John Wiley and Sons, Inc. 1994.
[34]Capital IQ online information services, a Standard & Poor's Business: https://www.capitaliq.com/home.aspx
[35] Mergent Online, Mergent Inc., 2008: http://www.mergentonline.com/

***Step 3: Set User Context***

The user context selects a set of financial ratios and places appropriate weights on the ratios according to the contextual relevance. Such selection and weighing factors are expressed in the form of relevance vector. I set the following six distinctive sets of user context as in Table 4.

**Table 4.** List of User Contexts

| | |
|---|---|
| *Context-1* | *Investor, Risk-taking, High-return, Short-term* |
| *Context-2* | *Investor, Risk-averse, Moderate-return, Long-term* |
| *Context-3* | *Job Seeker, Risk-averse, Stable growth, Long-term* |
| *Context-4* | *Supplier, Risk-averse, Secure payment, Steady business, Long-term* |
| *Context-5* | *Lender, Risk-averse, Moderate-return, One-time transaction* |
| *Context-6* | *Lender, Risk-averse, Low-return, Long-term* |

***Step 4: Select a Subspace***

In the subspace selection process, I chose to have five Characteristic Parameters with the highest values of weighting factors set as the user context.

***Step 5: Project and Evaluate Companies***

Each company was projected onto the selected subspace and the Euclidean norm for each company Entity Vector was calculated as the relevance score. Companies were ranked by the magnitude of the Euclidean norm. The same process was repeated for each context setting.

## RESULTS AND OBSERVATIONS

I experimented a context-dependent multidimensional corporate analysis method in a financial domain semantic space. The results indicate diverse ranking among different contexts as shown in the figure below. The length of each horizontal bar signifies the degree of each company's fit in the given context. To observe the effectiveness of the model, I compared the analysis results for Context 1 and Context 2 with the factual data of the historical stock market prices[36]. I set the stock prices at the time of analysis, close of December 2008, as the base price level (100%) and observed how the stock prices behaved for a short period (Context 1: 12 months) and a longer period (Context 2: 27 months). Disregarding the over-fluctuated outliers, the high ranked Company F and Company N performed well in relation to other companies on a steady basis.

[36] Yahoo! Finance: http://biz.yahoo.com/r/

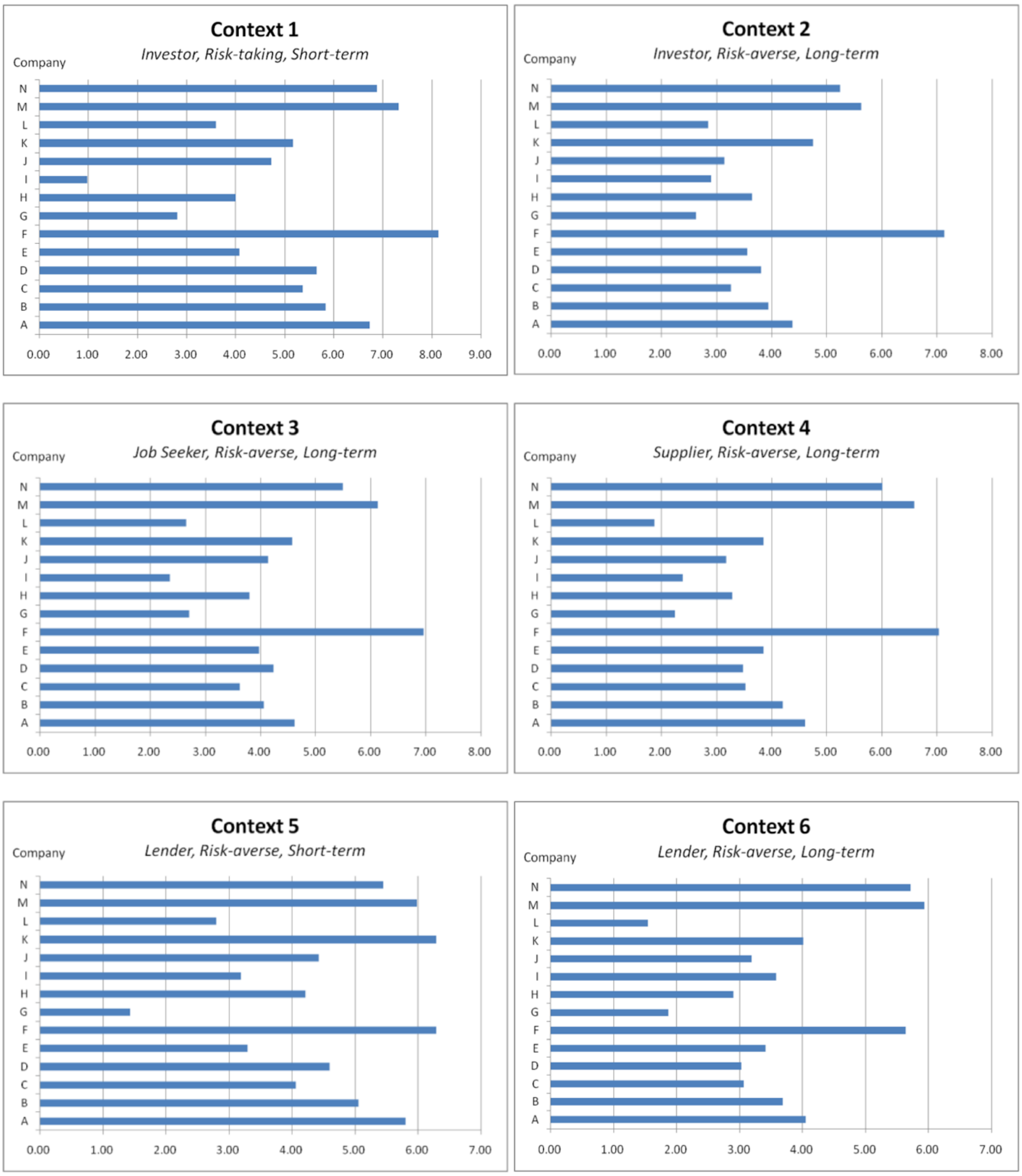



**Figure 14.** Company Ranking by different Stakeholder Context settings

*Each horizontal bar indicates the value of inner product, signifying the company's fit in accordance with the criteria set forth by each context. Longer bar means a better fit and thus a higher rank.*

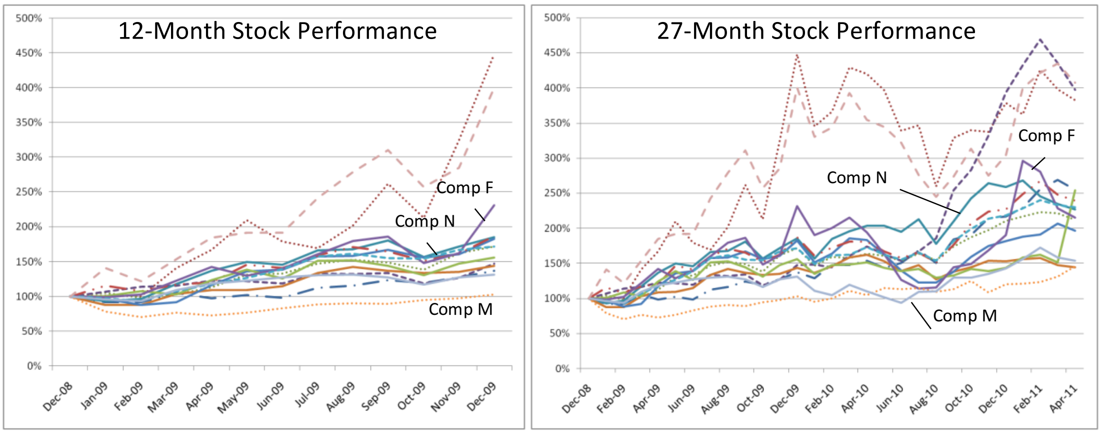



**Figure 15.** Context-based Corporate Ranking v.s. Actual Stock Performance

*Corporate rankings for short-term investment (Context-1) and long-term investment (Context-2) were compared against the actual stock price movements. The analysis was made as of December of 2008 and the stock price of each company at that time serves the basis (100%). The graphs show how each stock price moved over time relative to the basis.*

## Case 2: Context-dependent Corporate Evaluation in Integrated Multi-Domain Semantic Space

I demonstrate construction and use of a context-dependent company evaluation system using integrated semantic space with domains in finance, technology and brand. I experiment the validity of Data Analysis Module with fifty-eight global companies of a number of sectors with their headquarters located in the USA, Germany, France, Finland, Canada, Japan, Netherland, Taiwan, and Korea. The sectors cover business services, software, financial services, electronics, internet services, automobile, FMCG (Fast Moving Consumer Goods), restaurants, luxury items, etc.

### Objective of the Experiment

I integrate three distinctive semantic spaces in the domains of finance, technology, and brand to create a unified multi-domain semantic space. I map each of the target companies onto the integrated semantic space and evaluate them based on the context vectors set forth for each user context. I examine the operation and resulting ranks.

## EXPERIMENTAL PROCEDURE

The experiment was conducted in the following steps.

### *Step 1: Design an Integrated Multi-Domain Semantic Space with Characteristic Parameters*

I used the following Characteristic Parameters to construct an integrated multi-domain semantic space. The parameters in finance are financial ratios and key indicators commonly used in corporate finance. The parameters in technology consist of patent related data. The parameters in brand were published indexes.

**Table 4.** Characteristic Parameters for the Multi-Domain Semantic Space

| | |
|---|---|
| Finance | |
| Basic: | Market Capitalization, Revenue, Gross profit, Net income, R&D Expense |
| Liquidity: | Cash and equivalent, Current ratio, Quick ratio |
| Solvency: | Total debt-equity ratio, Long-term debt-equity ratio |
| Volatility: | Beta (1 year), Beta (2-year average), Beta (5-year average) |
| Profitability: | Gross margin, EBITDA margin, Net income margin |
| Efficiency: | Return on asset, Return on common equity, Total asset turnover |
| Growth: | Revenue growth (1 year, 3 years, 5 years), EBITDA growth (1 year, 3 years, 5 years) |
| Income: | Dividend yield |
| Technology | |
| Total active patents | |
| Total top-10% active patens, Top-10% ratio | |
| Top-10% patent generation with priority date in 2001 – 2002 | |
| Top-10% patent generation with priority date in 2003 – 2006 | |
| Top-10% patent generation with priority date in 2007 – 2010 | |
| Brand | |
| Brand economic value in 2011 | |
| Brand value growth – 1 year | |
| Brand value growth – 5 year | |
| Brand value growth – 9 year | |
| Global ranking 2010, Global ranking 2011 | |
| Green score 2011 | |

### *Step 2: Collect Raw Data and Create Characteristic Parameters*

I collect relevant data for the companies from the following sources:

a. Financial - Financial statements of publicly traded companies for 10 years[37] [38]
b. Technology – Patent data for 10 years[39]
c. Brand – Brand indexes as value estimate for 10 years and publicly available "Green Score" [40] [41]

***Step 3: Set User Context***

I set the following user context for this experiment. I translate these user contexts and represent them as weights on appropriate Characteristic Parameters. I used a fixed weight points to distribute them among associated Characteristic Parameters.

**Table 5.** User Context Specification for Multi-Domain Semantic Space

| Role-Objective Model | |
|---|---|
| $RO_1$: Investor | Income gain, Eco-conscious, Long-term, Risk-averse, Non-tech, Socially accepted |
| $RO_2$: Investor | Capital gain, Short-term, Innovative, Risk-taking |
| $RO_3$: Customer | Pleasure of ownership, Stable, Growing, Innovative |
| $RO_4$: Job Seeker | Well-known, Socially accepted, Growing, Somewhat stable, Large, Well-managed |
| $RO_5$: Supplier | Cash-rich, Stable, Growing, Innovative, Profitable, Efficient |
| $RO_6$: Business Acquirer | Innovative, High-quality patent-rich, Not heavily leveraged, Efficient |

The contextual settings heavily depend on the role of the user, company stakeholders, and their objectives in association with the users' risk-benefit profiles and other concerns.

---

[37]Capital IQ online information services, a Standard & Poor's Business: https://www.capitaliq.com/home.aspx

[38] Yahoo! Finance: http://biz.yahoo.com/r/

[39] Innography: http://www.innography.com/

[40]Interbrand Best Global Brands 2006, 2007, 2008, 2009, 2010, 2011: http://www.interbrand.com/en/best-global-brands/BGB-Interactive-Charts.aspx

[41] Interbrand Best Global Green Brands 2011: http://www.interbrand.com/en/best-global-brands/Best-Global-Green-Brands/2011-Report/BestGlobalGreenBrandsTable-2011.aspx

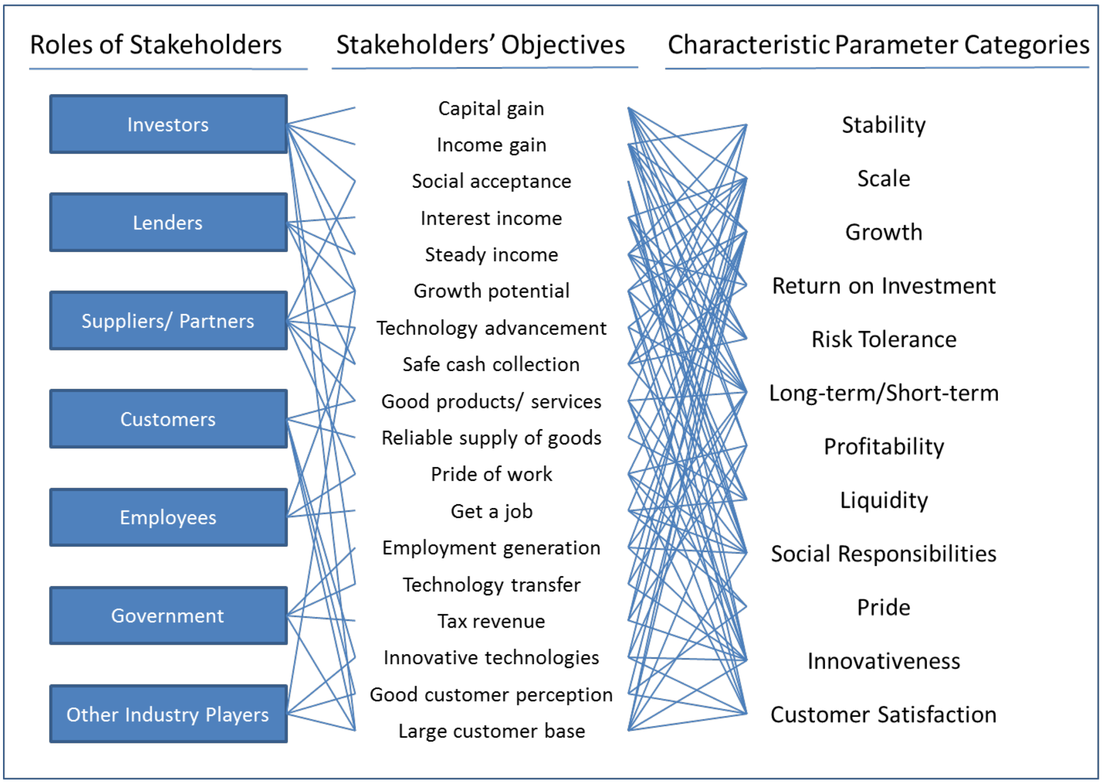



**Figure 16.** Translation of Context Specifications into Characteristic Parameter Categories

The figure above indicates how company stakeholders' context may be broken down to Characteristic Parameters. Weight assignment to each of the Characteristic Parameters are subjective and thus differ greatly by individuals. In this experiment, I assumed typical representations of stakeholders as examples. This translation process requires expert knowledge at times and the system shall absorb this translation functionality to relieve the burden from the end user. I took the system's role to accomplish such translation in this experiment.

Each Characteristic Parameters were normalized by rescaling where the maximum absolute value is set to either 1 or -1 and the minimum value is set to 0. Thus the resulting Characteristic Parameters have ranges between -1 and 1 inclusively. Characteristic Parameters were assumed to be orthogonal to each other.

***Step 4: Select a Subspace***

In the subspace selection process, I chose to have five Characteristic Parameters with the highest values as weighting factors set as the user context.

***Step 5: Evaluate Companies***

Each company represented by Characteristic Parameters were projected onto the subspace and Euclidean norm was calculated to rank companies according to the value of the norm. The higher the value of the norm indicates the closer companies match the stakeholders' criteria given as the user context.

## RESULTS AND OBSERVATIONS

The experiment results show that the corporate ranking significantly vary depending on the contexts. The scores and rankings are shown in the table below. The rankings were graphically represented in the figure below where companies are listed horizontally from Company 1 through Company 58 while the rankings of each company are shown on the vertical axis according to the user context specifications.

**Table 6.** Resulting Score and Associated Ranking

| | | | Score | | | | | | RANK | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sector | HQ Country | Company | OR-1 | OR-2 | OR-3 | OR-4 | OR-5 | OR-6 | OR-1 | OR-2 | OR-3 | OR-4 | OR-5 | OR-6 |
| Alcohol | Netherlands | 1 | 4.0 | 1.6 | 1.5 | 3.6 | 4.2 | 3.0 | 50 | 46 | 56 | 54 | 40 | 56 |
| Apparel | USA | 2 | 3.3 | 0.7 | 0.7 | 3.7 | 3.6 | 3.6 | 56 | 55 | 58 | 51 | 54 | 47 |
| Automotive | Germany | 3 | 4.4 | 2.4 | 2.4 | 4.6 | 4.6 | 3.7 | 46 | 33 | 51 | 42 | 32 | 46 |
| Automotive | Germany | 4 | 6.0 | 1.9 | 4.3 | 5.9 | 4.3 | 3.4 | 18 | 41 | 26 | 21 | 37 | 54 |
| Automotive | Germany | 5 | 6.0 | 1.7 | 4.7 | 5.7 | 4.1 | 3.8 | 19 | 45 | 19 | 23 | 44 | 44 |
| Automotive | Germany | 6 | 6.1 | 2.7 | 4.1 | 5.8 | 5.3 | 3.9 | 17 | 27 | 29 | 22 | 17 | 40 |
| Automotive | Japan | 7 | 5.4 | 0.9 | 4.5 | 5.4 | 4.2 | 4.2 | 32 | 53 | 23 | 31 | 39 | 33 |
| Automotive | Japan | 8 | 3.1 | 1.9 | 2.0 | 3.1 | 3.9 | 3.8 | 57 | 42 | 53 | 57 | 47 | 43 |
| Automotive | Japan | 9 | 5.8 | 0.9 | 5.2 | 6.2 | 4.9 | 4.5 | 27 | 52 | 14 | 17 | 25 | 30 |
| Automotive | USA | 10 | 4.1 | 3.4 | 3.9 | 3.1 | 1.9 | 3.3 | 48 | 18 | 31 | 56 | 58 | 55 |
| Automotive | USA | 11 | 1.6 | 2.2 | 1.4 | 1.8 | 2.6 | 3.9 | 58 | 35 | 57 | 58 | 57 | 41 |
| Beverages | USA | 12 | 8.3 | 2.6 | 5.4 | 7.7 | 5.7 | 4.0 | 3 | 29 | 13 | 6 | 12 | 38 |
| Beverages | USA | 13 | 7.0 | 2.5 | 4.6 | 6.5 | 5.2 | 4.1 | 10 | 32 | 21 | 11 | 20 | 34 |
| Bus Svc | Germany | 14 | 6.8 | 4.1 | 5.1 | 6.4 | 5.8 | 5.4 | 11 | 9 | 15 | 13 | 11 | 15 |
| Bus Svc | USA | 15 | 5.7 | 2.3 | 3.1 | 6.0 | 4.6 | 4.7 | 28 | 34 | 42 | 20 | 30 | 24 |
| Bus Svc | USA | 16 | 6.4 | 3.7 | 6.5 | 6.5 | 5.7 | 5.9 | 12 | 12 | 7 | 12 | 13 | 11 |
| Bus Svc | USA | 17 | 7.9 | 4.7 | 7.9 | 8.3 | 6.7 | 6.6 | 4 | 6 | 3 | 3 | 6 | 4 |
| Bus Svc | USA | 18 | 6.2 | 5.1 | 5.9 | 6.2 | 6.4 | 6.3 | 15 | 3 | 9 | 16 | 9 | 7 |
| Diversified | Germany | 19 | 5.9 | 2.5 | 4.6 | 5.6 | 5.2 | 4.8 | 23 | 31 | 20 | 27 | 22 | 22 |
| Diversified | USA | 20 | 5.2 | 3.7 | 4.4 | 5.1 | 5.2 | 6.1 | 36 | 11 | 24 | 36 | 21 | 10 |
| Diversified | USA | 21 | 4.2 | 3.2 | 3.6 | 4.4 | 3.7 | 4.0 | 47 | 20 | 34 | 46 | 52 | 39 |
| Diversified | USA | 22 | 7.5 | 3.5 | 5.4 | 7.8 | 6.7 | 6.4 | 8 | 15 | 12 | 4 | 7 | 6 |
| Electronics | Canada | 23 | 5.2 | 3.0 | 3.5 | 5.6 | 4.9 | 4.9 | 35 | 23 | 36 | 24 | 24 | 19 |
| Electronics | Finland | 24 | 5.5 | 0.5 | 4.3 | 5.4 | 4.1 | 4.6 | 30 | 58 | 25 | 29 | 43 | 27 |
| Electronics | Japan | 25 | 5.5 | 3.4 | 5.7 | 6.0 | 5.5 | 6.5 | 29 | 17 | 11 | 19 | 14 | 5 |
| Electronics | Japan | 26 | 4.0 | 1.9 | 4.2 | 4.8 | 5.3 | 5.6 | 51 | 43 | 27 | 41 | 19 | 13 |
| Electronics | Japan | 27 | 4.0 | 0.7 | 4.0 | 4.9 | 4.1 | 4.6 | 49 | 56 | 30 | 39 | 42 | 26 |
| Electronics | South Korea | 28 | 7.1 | 4.0 | 6.8 | 7.5 | 7.1 | 6.2 | 9 | 10 | 6 | 7 | 3 | 8 |
| Electronics | Taiwan | 29 | 6.4 | 4.8 | 2.9 | 5.6 | 6.0 | 4.7 | 13 | 5 | 44 | 26 | 10 | 23 |
| Electronics | USA | 30 | 9.8 | 7.4 | 9.0 | 10.4 | 8.5 | 6.7 | 1 | 1 | 1 | 1 | 1 | 3 |
| Electronics | USA | 31 | 4.6 | 2.1 | 3.3 | 5.2 | 3.7 | 4.1 | 42 | 37 | 38 | 35 | 51 | 36 |
| Electronics | USA | 32 | 6.2 | 3.3 | 6.3 | 6.8 | 4.8 | 5.5 | 14 | 19 | 8 | 10 | 27 | 14 |
| Electronics | USA | 33 | 7.8 | 4.8 | 7.0 | 7.8 | 7.1 | 7.0 | 5 | 4 | 5 | 5 | 4 | 2 |
| Electronics | USA | 34 | 4.5 | 3.6 | 4.9 | 4.6 | 3.6 | 5.3 | 45 | 13 | 18 | 43 | 53 | 16 |
| Energy | Netherlands | 35 | 6.2 | 4.2 | 4.6 | 6.1 | 5.3 | 4.7 | 16 | 8 | 22 | 18 | 18 | 25 |
| Fin Svc | France | 36 | 4.8 | 0.7 | 2.6 | 4.3 | 3.1 | 2.7 | 41 | 57 | 45 | 48 | 55 | 58 |
| Fin Svc | Germany | 37 | 4.6 | 0.9 | 2.5 | 3.9 | 3.0 | 2.8 | 43 | 54 | 49 | 49 | 56 | 57 |
| FMCG | France | 38 | 5.8 | 1.6 | 3.3 | 5.3 | 4.7 | 3.6 | 25 | 48 | 39 | 34 | 28 | 48 |
| FMCG | France | 39 | 6.0 | 1.6 | 3.7 | 5.6 | 5.0 | 4.1 | 22 | 47 | 32 | 25 | 23 | 37 |
| FMCG | Germany | 40 | 3.8 | 1.1 | 1.6 | 3.6 | 4.3 | 3.5 | 53 | 51 | 55 | 53 | 38 | 53 |
| FMCG | USA | 41 | 4.9 | 1.4 | 2.5 | 4.5 | 3.7 | 3.9 | 38 | 49 | 48 | 45 | 50 | 42 |
| FMCG | USA | 42 | 5.4 | 2.7 | 2.6 | 5.0 | 4.9 | 4.5 | 31 | 28 | 46 | 37 | 26 | 29 |
| FMCG | USA | 43 | 4.9 | 2.1 | 2.5 | 4.3 | 4.0 | 3.6 | 39 | 40 | 50 | 47 | 45 | 51 |
| FMCG | USA | 44 | 5.9 | 3.6 | 5.0 | 5.4 | 6.8 | 6.2 | 24 | 14 | 16 | 30 | 5 | 9 |
| FMCG | USA | 45 | 6.0 | 1.4 | 3.2 | 5.3 | 4.1 | 3.6 | 21 | 50 | 40 | 32 | 41 | 52 |
| Internet Svc | USA | 46 | 5.3 | 3.4 | 5.7 | 6.2 | 4.5 | 5.1 | 33 | 16 | 10 | 15 | 35 | 18 |
| Internet Svc | USA | 47 | 4.8 | 3.1 | 3.4 | 4.5 | 4.4 | 4.1 | 40 | 22 | 37 | 44 | 36 | 35 |
| Internet Svc | USA | 48 | 7.7 | 4.6 | 7.3 | 7.5 | 6.6 | 5.6 | 6 | 7 | 4 | 8 | 8 | 12 |
| Internet Svc | USA | 49 | 3.4 | 2.8 | 2.5 | 3.2 | 4.0 | 4.5 | 55 | 25 | 47 | 55 | 46 | 28 |
| Luxury | France | 50 | 5.0 | 2.6 | 2.3 | 5.0 | 5.5 | 4.4 | 37 | 30 | 52 | 38 | 15 | 31 |
| Luxury | France | 51 | 5.8 | 2.1 | 3.7 | 5.5 | 4.6 | 3.6 | 26 | 38 | 33 | 28 | 29 | 50 |
| Luxury | USA | 52 | 3.6 | 1.8 | 1.6 | 3.7 | 3.7 | 3.8 | 54 | 44 | 54 | 50 | 48 | 45 |
| Media | USA | 53 | 5.3 | 2.1 | 3.5 | 5.3 | 3.7 | 3.6 | 34 | 39 | 35 | 33 | 49 | 49 |
| Restaurants | USA | 54 | 7.5 | 2.2 | 4.2 | 7.2 | 5.3 | 4.3 | 7 | 36 | 28 | 9 | 16 | 32 |
| Restaurants | USA | 55 | 4.5 | 2.7 | 3.0 | 4.8 | 4.6 | 4.8 | 44 | 26 | 43 | 40 | 31 | 21 |
| Software | USA | 56 | 3.9 | 3.2 | 3.2 | 3.6 | 4.6 | 4.9 | 52 | 21 | 41 | 52 | 34 | 20 |
| Software | USA | 57 | 8.6 | 5.7 | 8.1 | 8.5 | 8.0 | 7.6 | 2 | 2 | 2 | 2 | 2 | 1 |
| Sport Goods | USA | 58 | 6.0 | 3.0 | 4.9 | 6.3 | 4.6 | 5.1 | 20 | 24 | 17 | 14 | 33 | 17 |

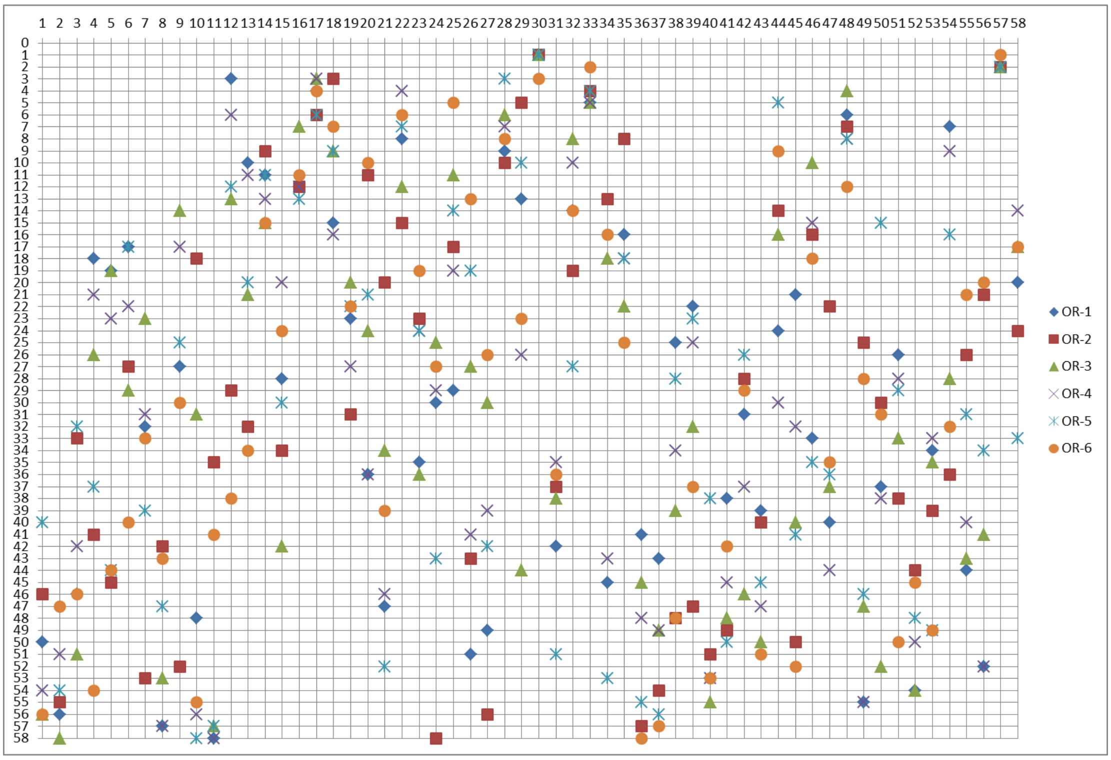



**Figure 17.** A Resulting Company Ranking According to the Contextual Specification

*OR-1 through OR-6 Indicating Diverse Preference of Companies (Vertical axis indicates the ranking; Horizontal Axis indicates the Company)*

I extract the best and worst rankings for all six context settings for each of the analyzed companies. I then make comparisons among companies by observing the degree of differences between the best ranking and the worst ranking. I note that there is a general positive correlation between the best ranking and the worst ranking. I interpret that company rankings tend to converge among all contextual settings – in general, high ranking companies tend to score high for all contextual settings while low ranking companies tend to score low for all contextual settings.

Typical examples are indicated by large circles in the figure below. On the contrary, there seem to exist certain groups of companies that exhibit significant differences between the highest ranking and the lowest ranking. Companies whose best rankings have higher rankings among all companies but their worst rankings indicate relatively low rankings as depicted by small circles. Technology-based companies such as electronics and automobile tend to dominate the second group of companies.

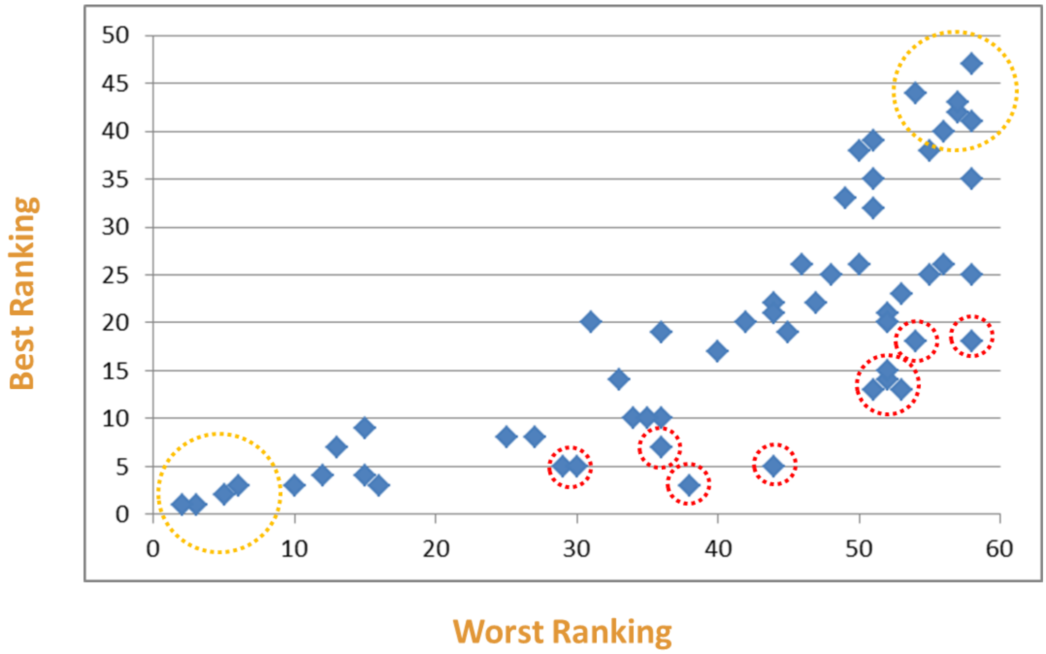



**Figure 18.** Best Ranking vs Worst Ranking among All Six Contextual Specifications (Smaller numbers indicate higher rankings)

### CHAPTER SUMMARY

I presented two cases of experiments of context-dependent multidimensional corporate analysis methods. I described the experimental procedures and the resulting rankings of companies according to each of the user context. The resulting ranking exhibited a wide variance with respect to the user contextual settings in both cases.

A special care was taken to integrate Characteristic Parameters in different domains because they have different measurement units and different ranges in data values. To construct a metric semantic space, each Characteristic Parameters were normalized.

I assumed that selected Characteristic Parameters that constituted the semantic space were orthogonal to each other to secure linear independence. To fully claim the orthogonality, a process such as eigenvalue decomposition may be introduced. This process was omitted because it was not an essential part of the experiment.

I note that there can be multiple options in selecting the subspace by the user context as well as evaluating company parameters depending on the objective and constraints for the analyses. For the subspace selection by the user context, I chose the top five parameters of the highest values in the user context represented by Characteristic Parameters. This process is essentially equivalent to taking the inner product between the context vector and the basis vectors of the Characteristic Parameters. In the company evaluation process, I chose to taking the Euclidean norm of companies' Entity Vectors that are projected onto the subspace as a measure of relevance. The normalization of Characteristic Parameters is an important preprocess for the company evaluation phase that allows Euclidean norm to be a valid relevance measure.

Proper and sufficient adoption of Characteristic Parameters in constructing a semantic space is essential to the viability of the methodology. Companies need to be adequately represented by the Characteristic Parameters for the analyses. Careful examination may be necessary to identify very important but unobvious parameters that are embedded in corporate realities. High context cultures tend to include such hidden parameters.

# CHAPTER 6. EXPERIMENTS ON TIME SERIES ANALYSIS USING FINANCIAL SEMANTIC SPACE

I introduce an application of the semantic computing methodology[42] to a time-series analysis in finance. In this application, I regard the state of exchange-traded assets at each time interval as an entity vector and compare it with the current state entity by the degree of similarity. The resemblance of the states and the changes that took place in the past suggests predictive capability of the model into the future[43].

## INTRODUCTION

Capital markets are a major component of modern economic system and its role is becoming increasingly important. More companies finance themselves from the equity and debt markets rather than going through the traditional banking system. Simultaneously, the proliferation of Internet and brokers' competition yielded significant reduction in trading costs of financial products. This new environment has made it possible for many individual investors to participate on a global scale.

Experienced investors make their investment decisions based on thorough analyses of companies with the investors' goals and constraints in mind. There are two major schools of analytic thoughts - fundamental analysis and technical analysis. Fundamental analysis takes a long-term view and takes into account relevant business and economic factors to assess future state of the company, estimating its value. Technical analysis takes a short-term view and considers the movement of company's

---

[42] Ito Shin, Kiyoki Yasushi, "A Context-based Multi-Dimensional Corporate Analysis Method" INFORMATION MODELING AND KNOWLEDGE BASES 2012, Vol. XXIV, IOS Press.

[43] Ito Shin, Kiyoki Yasushi, "A Multidimensional Market Analysis Method Using Level-Velocity-Momentum Time-Series Vector Space" INFORMATION MODELING AND KNOWLEDGE BASES 2014, Vol. XXV, IOS Press.

stock prices in the market. Stock prices are considered a reflection of all market's intelligence and participants' psychological states. Technical analysis only considers the trends and movements of these market factors. Investors tend to take both aspects of these analyses[44] but they are taken as two independent guides for investment decisions. This paper incorporates both aspects of fundamental and technical analyses simultaneously in the form of multi-spatial description.

The basic premise of this method is that given enough historical records of relevant market and socioeconomic patterns, the current pattern of interest may be identified with similarity. As the similarity-identified historical patterns already have factual records of the forward-looking markets and individual stocks, certain inferences may be made for the current situation.

## BACKGROUND

Numerous researches have been done both in academia and practice to understand and predict stock prices[45] [46] [47] [48]. A major controversy exists around random walk - whether the immediate future price of an individual stock price or the entire market follows the Brownish motion or not. The school of random walk denies predictability of future movement. Another school expresses that the probability distribution changes when there exists trend of price movement as observed in the significant rising or falling prices. I take the position that the stock prices movement is not completely random and thus prediction of stock prices is possible.

---

[44] Bettman, Jenni L. et. al., "Fundamental and technical analysis: substitutes or complements?", Accounting & Finance; Mar2009, Vol. 49 Issue 1, p21-36, 16p.

[45] Hickman, Kent, "A Comparison of Stock Price Predictions Using Court Accepted Formulas, Dividend Discount, and P/E Models", The Journal of the Financial Management Association; Summer1990, Vol. 19 Issue 2, p76-87, 12p.

[46] McCurtain, Robert, "Getting Technical with Economic Data", Futures: News, Analysis & Strategies for Futures, Options & Derivatives Traders; Nov2010, Vol. 39 Issue 11, p44-50, 4p

[47] Boni, Leslie. et. al., "Analysts, Industries, and Price Momentum", Journal of Financial & Quantitative Analysis; Mar2006, Vol. 41 Issue 1, p85-109, 25p.

[48] Bulkley, George., et. al., "Can the Cross-Sectional Variation in Expected Stock Returns Explain Momentum?", Journal of Financial & Quantitative Analysis; Aug2009, Vol. 44 Issue 4, p777-794, 18p.

Fundamental analysts look for an undervalued company with high growth potential. They observe a candidate company and its operating environments with great care to assess the company's economic value, from which they calculate the "right" stock price. They compare it with the price determined by the market. When fundamental analysts recognize that the market is significantly underpricing such company, they decide to invest. Fundamentalists tend to invest for a long-term return as they wait for the company growth and market's adjustments to the right price level.

Technical analysts have identified a number of stock price movement patterns that reflect investors' varying psychology between profit taking and risk tolerance. Technical traders take advantage of these price and volume movement patterns and make buy-hold-sell decisions. Technical analysis methods have been researched for different markets[49] [50] [51] [52] and widely practiced today. While these patterns may guide investors for their investment decisions, they present certain shortcomings. Because price movement occurs in real time, identifying the "correct" pattern may be difficult. It is easy to recognize the current price movement as a "wrong" pattern especially at an early stage of such pattern formation. In addition, even when the pattern is correctly identified, the shape of such pattern may take different forms or diverge completely, making it difficult for investors to take optimal actions in the market.

Advancement of artificial intelligence and its applications have proposed predictive models. Examples of notable models are based on foundations of Artificial

---

[49] Treynor, Jack L., et. al., "In Defense of Technical Analysis", Journal of Finance; Jul1985, Vol. 40 Issue 3, p757-773, 17p.

[50] Lo, Andrew W. et. at., "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation", Journal of Finance; Aug2000, Vol. 55 Issue 4, p1705-1765, 61p.

[51] Lai, Hung-Wei, et. al., "Technical Analysis, Investment Psychology, and Liquidity Provision: Evidence from the Taiwan Stock Market", Emerging Markets Finance & Trade; Sep/Oct2010, Vol. 46 Issue 5, p18-38, 21p.

[52] Varadharajan, P., "Effectiveness of technical analysis using candlestick chart for selection of equity stock in Indian capital market", Journal of Contemporary Management Research; March 2011, Vol. 5 Issue 1, p12-23, 12p.

Neural Networks (ANN) and Bayesian Networks (BN)[53] [54] [55] [56]. Even these robust models may not capture influential factors well for their predictive capabilities. Our model proposes a methodology that utilizes both factors from fundamental analysis and technical analysis and that allows very flexible inclusion of any number of new factors as analytic variables.

**AN OVERVIEW OF THE PREDICTION METHODOLOGY**

I provide an overview of the Semantic Space based analysis and prediction model. In this method, I define and construct a set of Characteristic Parameters to describe economic instances. Each instance in time series is represented as an entity in the semantic space.

To describe the economic reality, I take a set of Characteristic Parameters that are layered from macro to micro in scale as depicted in Figure 19. Factors exist in a number of subgroups of economic activities such as geographic regions and countries, and different sectors that are all related at various degrees in today's global economy. Examples of such descriptive factors are given in Table 7 for illustration purposes.

---

53 Huang Wei, et. al., Neural Networks in Finance and Economics Forecasting", International Journal of Information Technology & Decision Making; Mar2007, Vol. 6 Issue 1, p113-140, 28p.

54 Zuo Yi., et. al., "Up/Down Analysis of Stock Index by Using Bayesian Network", Engineering Management Research; Vol. 1, No. 2; 2012.

55 Tsai Chih-Ling, et. al., "Does a Bayesian approach generate robust forecasts? Evidence from applications in portfolio investment decisions", Published on line - The Institute of Statistical Mathematics, Tokyo 2009.

56 Otrok, Christopher and Whiteman, Charles H., "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa", The NBER/NSF Seminar on Forecasting and Empirical Methods in Macroeconomics, July 1996.

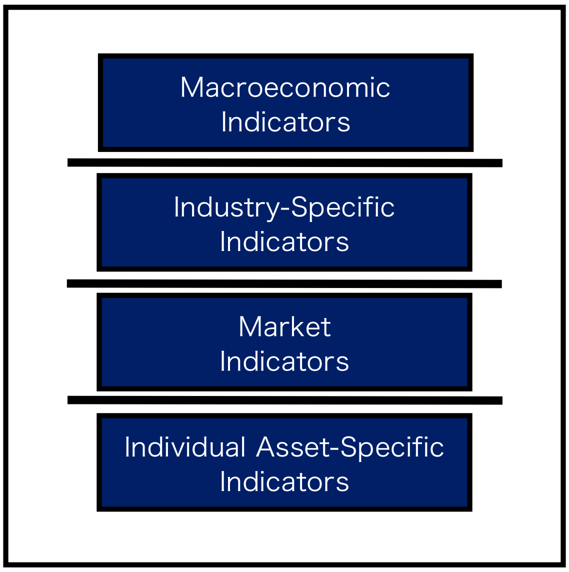



**Figure 19.** Layered Descriptive Factors

**Table 7.** Multi-layered Indicators

Macroeconomic Indicators
- GDP, GDP growth
- Inventory levels
- Trade balance
- Government's debts per GDP
- University of Michigan Consumer Sentiment Index
- Inflation rate, Consumer price index, Producer price index
- Unemployment rate, New unemployment insurance registration

Industry-specific Indicators
- Manufacturers' sentiment
- New construction permit
- Industry specific indicators

Market Indicators
- Interest rates
- Foreign exchange rates
- Stock market indexes - DJIA, S&P500, NIKKEI225, FTSE100, etc.

Individual Asset-specific Indicators
- Individual stock prices
- Financial results - financial statement elements, financial ratios, levels
- Technological advancement - patent related scores, # of patents
- Social acceptance - brand score

I collect appropriate data to create a matrix "Raw Data Matrix" (RDM) from which Characteristic Parameters are developed. Using Characteristic Parameters, I construct the Characteristic Matrix (CM). Each column of CM is an instance that describes socioeconomic state at that time. The rows of the CM represent instances in time series. The temporal resolution is to be determined by the investor's investment cycle. A current instance specified as the Current Vector is compared with each column of Characteristic Matrix by calculating the Euclidean distance between each set. The next instance of the most similar instance to the Current Vector shall be the predicted state.

### OBJECTIVE OF THE EXPERIMENT

The objective of the experiment is to implement a market predictive model based on semantic metric space representation of Characteristic Parameters and prove its model concept with predictive effectiveness.

### DETAILED OPERATIONAL PROCEDURE OF THE METHOD

I look at the operational procedure in more detail. Figure 20 shows the overall procedural sequence with intermediary matrixes and vectors.

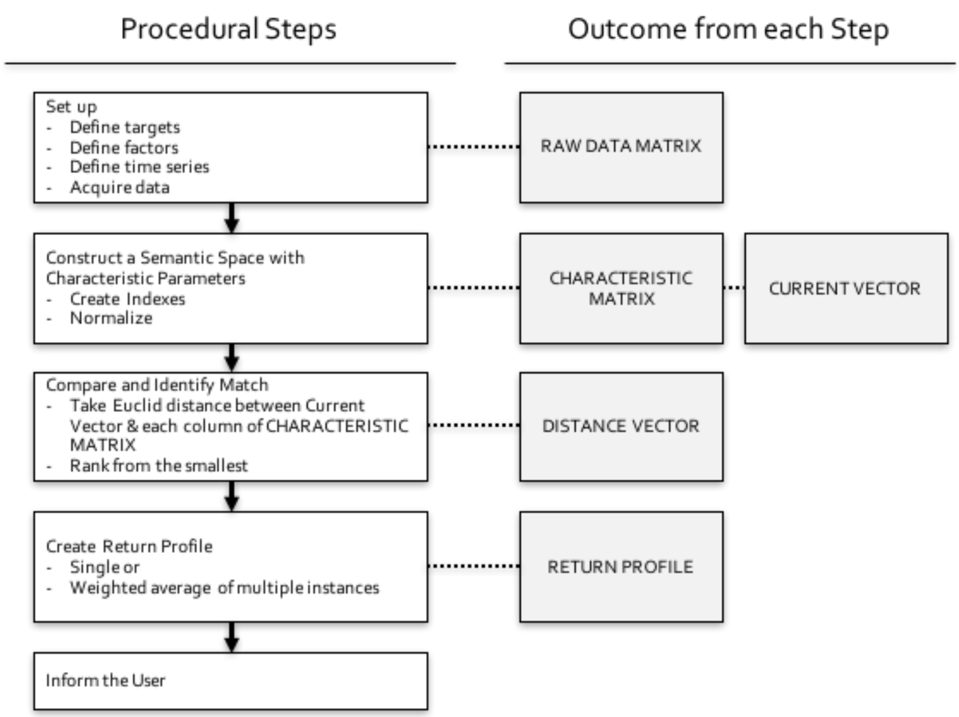



**Figure 20.** An overall procedural sequence

***Select investment target***

Many investors construct a portfolio of investing assets to maximize return while minimizing risks. Assets of interest may be central government's, municipal, or corporate bonds, stock market indexes, mutual funds of certain industries, or individual stocks. Investor's interest will determine the investment target and such target will determine the variables in the schema.

***Define Characteristic Parameters***

Characteristic Parameters are the construction elements of a semantic space. Characteristic Parameters should be defined so that the state of a financial asset at any given instance is well described for analyses. The target asset shall dictate relevant data types that may be highly correlated with or may have causal effect to the price movement of the target asset. In defining the Characteristic Parameters, relevant raw data shall be carefully considered.

I define several financial market specific Characteristic Parameters to describe market movements at a given instance. Characteristic Matrix contains three types of Characteristic Parameters, namely, level-type, velocity-type, and the momentum type. The level-type elements indicate singularly sampled quantity such as price of an asset at certain time or volume of traded asset during a day, or Gross Domestic Product during a three-month period. The velocity-type indicates the strengths of price movements between two time instances, describing long-term to short-term price moves. The momentum-type is defined as a multiplication of a level and velocity. The momentum data compares well with physical definition of momentum as defined by a multiplication of mass and velocity. In our case mass is the volume traded and the velocity is the rate of price change. Some typical Characteristic Parameters are described below.

A basic Velocity Vector Element (VVE) is given as a rate of price change between two time periods. Thus, VVE can be described as:

$$VVE_{i,i-x} = \frac{P_i - P_{i-x}}{P_{i-x}}$$

where x denotes the time interval toward the past.

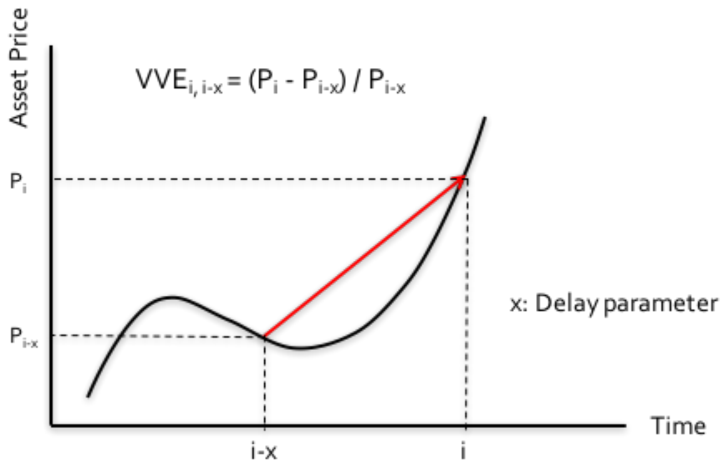



**Figure 21.** Constructing Velocity Vector Elements

In the example shown in Figure 21, the time i indicates the data point of interest, for example as the current instance, while the time i-x indicates a data point at different time periods, namely 1-day prior, 1-week prior, 1-month prior, and 3-months prior to the reference point. So the different VVEs indicate price changes over different time intervals all up to the reference point. The reference data will be taken at each of the time period with sweeping calculations over time.

Likewise, I define Moment Vector Element (MVE) as below:

$$MVE_{i,i-x} = \frac{(P_i - P_{i-x})(\sum_{i-x}^{i} V_k)}{P_{i-x} \; V_{i-x}}$$

where x specifies the time distance into the past from the current time i.

The MVE indicates the rate of price change in association with the traded volume.

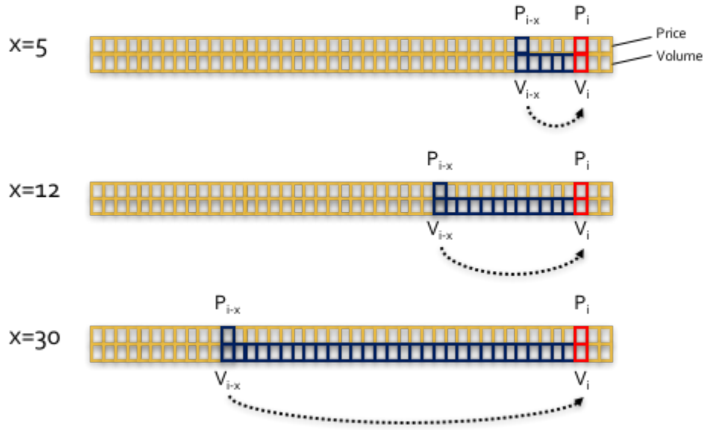



**Figure 22.** Momentum Vector Element

I also utilize a technical analysis indicator, the W%R parameter, with various time windows as below[57]:

$$W\%R_{i,n} = \frac{P_{n,max} - P_i}{P_{n,max} - P_{n,min}}$$

where n designates the number of time intervals toward the past starting at i.

The W%R indicates where the ending price level is in relation to the price moved during the time window i - n + 1 and i.

***Collect data and create Raw Data Matrix (RDM)***

Relevant data shall be acquired from the identified sources in a predetermined format and precision. The raw data will be assembled into a matrix called Raw Data Matrix (RDM).

---

[57] Wikipedia, http://en.wikipedia.org/wiki/Williams_%25R

$$\begin{bmatrix} \text{factor-1}_{i\text{-}k} & \ldots & \text{factor-1}_{i\text{-}2} & \text{factor-1}_{i\text{-}1} & \text{factor-1}_{i} \\ \text{factor-2}_{i\text{-}k} & \ldots & \text{factor-2}_{i\text{-}2} & \text{factor-2}_{i\text{-}1} & \text{factor-2}_{i} \\ \text{factor-3}_{i\text{-}k} & \ldots & \text{factor-3}_{i\text{-}2} & \text{factor-3}_{i\text{-}1} & \text{factor-3}_{i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \text{factor-n}_{i\text{-}k} & \ldots & \text{factor-n}_{i\text{-}2} & \text{factor-n}_{i\text{-}1} & \text{factor-n}_{i} \end{bmatrix}$$

** The vector with factors 1 through n indicate the multi-dimensional descriptors.*
** The subscript i, i-1, i-2, ..., i-k, ... indicate time-series with i being the current.*

**Figure 23.** Raw Data Matrix - time-series multi-dimensional vectors

***Construct the Characteristic Matrix (CM)***

The Characteristic Matrix is a collection of column vectors each of which represents an instance in time series. The description of an instance may be broken down to layers as exemplified in Figure 19. In this example, the layers consist of macroeconomic indicators, industry-specific indicators, market-specific indicators, and individual asset-specific indicators. I pre-process and manipulate data elements in the Raw Data Matrix to create Characteristic Parameters. Characteristic Parameters form the basis for constructing a semantic space. Each Characteristic Parameter is normalized such that the maximum absolute value is 1. Characteristic Matrix is created by filling instances for the entire time period of interest. This process is a projecting of each instance into the semantic space.

***Identify the Current Vector***

Similar to configuring the Characteristic Matrix, the Current Vector shall be constructed. The Current Vector is considered a special case of a column vector of the Characteristic Matrix -- it is the reference instance, usually the most recent instance in time series. All processes in constructing the Characteristic Matrix applies to making the Current Vector.

***Calculate Euclidean Distance between the Current Vector and Each Column of Characteristic Matrix***

I now have a means to compare for similarity between the current instance as represented by the current Characteristic Vector and past instances as represented by columns of the Characteristic Matrix. The similarities are measured by the Euclidian distance. The smaller the value of Euclidian distances, the higher the similarity. The reciprocal of the vector distance gives the score (dScore) indicating the degree of similarity.

Euclidean Distance: $\|d_i\| = \sqrt{\ (CM_i - CV)^2}$

where i denotes the $i^{th}$ column vector of CM

**DATASET USED IN THE EXPERIMENT**

***Investment Target***

A semiconductor sector-based ETF SPDR S&P Semiconductor (XSD) with 46 publicly traded semiconductor company stocks and has a market capitalization of over eight billion US dollars at the time of this writing.

***Time Series***

Data is acquired for the period from March $1^{st}$, 2006 to April $26^{th}$, 2013 for market open days. 1802 instances were created.

***Characteristic Parameters***

I select 36 Characteristic Parameters as described in Table 8 and Figure 24.

**Table 8.** Characteristic Parameters for Financial Assets

GDP Growth

- Quarterly GDP percent change from preceding period based on current dollars. The data is seasonally adjusted annual rates.
- Variable = Gdp-g $_t$
- Data type = level
- Frequency = Quarterly (3months)

Source = US Department of Commerce, Bureau of Economic Analysis

Interest Rate

- US Treasury Bill (4week) rate in the secondary market
- Variable = T-Bill $_t$
- Data type = level
- Frequency = Daily

Source = Board of Governors of the Federal Reserve System

Consumer Price Index (CPI)

- Variable = CPI $_t$
- Data type = level
- Frequency = Monthly

Source = US Bureau of Labor Statistics

Unemployment Rate

- Non-farm unemployment rate
- Variable = Unempl R $_t$
- Data type = level
- Frequency = Monthly

Source = US Bureau of Labor Statistics

Productivity Change Rate

- Non-farm labor productivity change in percentage
- Variable = Product R $_t$
- Data type = level
- Frequency = Quarterly

Source = US Bureau of Labor Statistics

Bill to Book Ratio

- The global billings and bookings of North American headquartered semiconductor equipment producers. All billings and bookings figures are based on a three-month moving average.
- Variable = BBR $_t$
- Data type = level
- Frequency = Monthly, Announcement date reflected

Source = David Powell, Incorporated

Semiconductor-sector ETF

- SPDR S&P Semiconductor (XSD)
- Original data = opening price, closing price, intraday high price, intraday low price, dividend split adjusted closing price, traded volume
- Variables:
    - XSD CP $_t$ : Closing price position as percentage of intraday trading range (level)
    - XSD Vel-1d $_t$ : Percent change in price at date t from date t-1 at closing (velocity)

- XSD Vel-5d $_t$ : Percent change in price at date t from date t-5 at closing (velocity)
- XSD Mom-1d $_t$ : Percent change in price at date t from date t-1 at closing multiplied by the traded volume during the date t (momentum)
- XSD Mom-5d $_t$ : Percent change in price at date t from date t-5 at closing multiplied by the cumulative traded volume during the 5-day period from date t-5 to date t , and divided by 5-times the heaviest intraday volume traded during the period of interest (momentum)
- Frequency: Daily

Source: Yahoo Finance

Telecommunication-sector ETF

- Vanguard Telecom Services ETF (VOX)
- Original data = opening price, closing price, intraday high price, intraday low price, dividend split adjusted closing price, traded volume
- Variables:
    - VOX CP $_t$ : Closing price position as percentage of intraday trading range (level)
    - VOX Vel-1d $_t$ : Percent change in price at date t from date t-1 at closing (velocity)
    - VOX Vel-5d $_t$ : Percent change in price at date t from date t-5 at closing (velocity)
    - VOX Mom-1d $_t$ : Percent change in price at date t from date t-1 at closing multiplied by the traded volume during the date t (momentum)
    - VOX Mom-5d $_t$ : Percent change in price at date t from date t-5 at closing multiplied by the cumulative traded volume during the 5-day period from date t-5 to date t , and divided by 5-times the heaviest intraday volume traded during the period of interest (momentum)
    - Frequency: Daily

Source: Yahoo Finance

Materials-sector ETF

- Vanguard Materials ETF (VAW)
- Original data = opening price, closing price, intraday high price, intraday low price, dividend split adjusted closing price, traded volume
- Variables:
    - VAW CP $_t$ : Closing price position as percentage of intraday trading range (level)
    - VAW Vel-1d $_t$ : Percent change in price at date t from date t-1 at closing (velocity)
    - VAW Vel-5d $_t$ : Percent change in price at date t from date t-5 at closing (velocity)
    - VAW Mom-1d $_t$ : Percent change in price at date t from date t-1 at closing multiplied by the traded volume during the date t (momentum)
    - VAW Mom-5d $_t$ : Percent change in price at date t from date t-5 at closing multiplied by the cumulative traded volume during the 5-day period from date t-5 to date t , and divided by 5-times the heaviest intraday volume traded during the period of interest (momentum)
    - Frequency: Daily

Source: Yahoo Finance

Financial Services-sector ETF

- Financial Select Sector SPDR (XLF)
- Original data = opening price, closing price, intraday high price, intraday low price, dividend split adjusted closing price, traded volume
- Variables:
    - XLF CP $_t$ : Closing price position as percentage of intraday trading range (level)
    - XLF Vel-1d $_t$ : Percent change in price at date t from date t-1 at closing (velocity)
    - XLF Vel-5d $_t$ : Percent change in price at date t from date t-5 at closing (velocity)
    - XLF Mom-1d $_t$ : Percent change in price at date t from date t-1 at closing

multiplied by the traded volume during the date t (momentum)

- XLF Mom-5d $_t$ : Percent change in price at date t from date t-5 at closing multiplied by the cumulative traded volume during the 5-day period from date t-5 to date t , and divided by 5-times the heaviest intraday volume traded during the period of interest (momentum)

Source: Yahoo Finance

NASDAQ index

- NASDAQ Composite
- Original data = opening price, closing price, intraday high price, intraday low price, dividend split adjusted closing price, traded volume
- Variables:
  - NDQ CP $_t$ : Closing price position as percentage of intraday trading range (level)
  - NDQ Vel-1d $_t$ : Percent change in price at date t from date t-1 at closing (velocity)
  - NDQ Vel-5d $_t$ : Percent change in price at date t from date t-5 at closing (velocity)
  - NDQ Mom-1d $_t$ : Percent change in price at date t from date t-1 at closing multiplied by the traded volume during the date t (momentum)
  - NDQ Mom-5d $_t$ : Percent change in price at date t from date t-5 at closing multiplied by the cumulative traded volume during the 5-day period from date t-5 to date t , and divided by 5-times the heaviest intraday volume traded during the period of interest (momentum)
  - Frequency: Daily

Source: Yahoo Finance

$$\begin{bmatrix}
\cdots & \text{Gdp-g}_{t-1} & \text{Gdp-g}_{t} & \text{Gdp-g}_{t+1} & \cdots \\
\cdots & \text{T-Bill}_{t-1} & \text{T-Bill}_{t} & \text{T-Bill}_{t+1} & \cdots \\
\cdots & \text{CPI}_{t-1} & \text{CPI}_{t} & \text{CPI}_{t+1} & \cdots \\
\cdots & \text{Unempl R}_{t-1} & \text{Unempl R}_{t} & \text{Unempl R}_{t+1} & \cdots \\
\cdots & \text{Product R}_{t-1} & \text{Product R}_{t} & \text{Product R}_{t+1} & \cdots \\
\cdots & \text{BBR}_{t-1} & \text{BBR}_{t} & \text{BBR}_{t+1} & \cdots \\
\cdots & \text{XSD CP}_{t-1} & \text{XSD CP}_{t} & \text{XSD CP}_{t+1} & \cdots \\
\cdots & \text{XSD Vel-1d}_{t-1} & \text{XSD Vel-1d}_{t} & \text{XSD Vel-1d}_{t+1} & \cdots \\
\cdots & \text{XSD Vel-5d}_{t-1} & \text{XSD Vel-5d}_{t} & \text{XSD Vel-5d}_{t+1} & \cdots \\
\cdots & \text{XSD Mom-1d}_{t-1} & \text{XSD Mom-1d}_{t} & \text{XSD Mom-1d}_{t+1} & \cdots \\
\cdots & \text{XSD Mom-5d}_{t-1} & \text{XSD Mom-5d}_{t} & \text{XSD Mom-5d}_{t+1} & \cdots \\
\cdots & \text{VOX CP}_{t-1} & \text{VOX CP}_{t} & \text{VOX CP}_{t+1} & \cdots \\
\cdots & \text{VOX Vel-1d}_{t-1} & \text{VOX Vel-1d}_{t} & \text{VOX Vel-1d}_{t+1} & \cdots \\
\cdots & \text{VOX Vel-5d}_{t-1} & \text{VOX Vel-5d}_{t} & \text{VOX Vel-5d}_{t+1} & \cdots \\
\cdots & \text{VOX Mom-1d}_{t-1} & \text{VOX Mom-1d}_{t} & \text{VOX Mom-1d}_{t+1} & \cdots \\
\cdots & \text{VOX Mom-5d}_{t-1} & \text{VOX Mom-5d}_{t} & \text{VOX Mom-5d}_{t+1} & \cdots \\
\cdots & \text{IGV CP}_{t-1} & \text{IGV CP}_{t} & \text{IGV CP}_{t+1} & \cdots \\
\cdots & \text{IGV Vel-1d}_{t-1} & \text{IGV Vel-1d}_{t} & \text{IGV Vel-1d}_{t+1} & \cdots \\
\cdots & \text{IGV Vel-5d}_{t-1} & \text{IGV Vel-5d}_{t} & \text{IGV Vel-5d}_{t+1} & \cdots \\
\cdots & \text{IGV Mom-1d}_{t-1} & \text{IGV Mom-1d}_{t} & \text{IGV Mom-1d}_{t+1} & \cdots \\
\cdots & \text{IGV Mom-5d}_{t-1} & \text{IGV Mom-5d}_{t} & \text{IGV Mom-5d}_{t+1} & \cdots \\
\cdots & \text{VAW CP}_{t-1} & \text{VAW CP}_{t} & \text{VAW CP}_{t+1} & \cdots \\
\cdots & \text{VAW Vel-1d}_{t-1} & \text{VAW Vel-1d}_{t} & \text{VAW Vel-1d}_{t+1} & \cdots \\
\cdots & \text{VAW Vel-5d}_{t-1} & \text{VAW Vel-5d}_{t} & \text{VAW Vel-5d}_{t+1} & \cdots \\
\cdots & \text{VAW Mom-1d}_{t-1} & \text{VAW Mom-1d}_{t} & \text{VAW Mom-1d}_{t+1} & \cdots \\
\cdots & \text{VAW Mom-5d}_{t-1} & \text{VAW Mom-5d}_{t} & \text{VAW Mom-5d}_{t+1} & \cdots \\
\cdots & \text{XLF CP}_{t-1} & \text{XLF CP}_{t} & \text{XLF CP}_{t+1} & \cdots \\
\cdots & \text{XLF Vel-1d}_{t-1} & \text{XLF Vel-1d}_{t} & \text{XLF Vel-1d}_{t+1} & \cdots \\
\cdots & \text{XLF Vel-5d}_{t-1} & \text{XLF Vel-5d}_{t} & \text{XLF Vel-5d}_{t+1} & \cdots \\
\cdots & \text{XLF Mom-1d}_{t-1} & \text{XLF Mom-1d}_{t} & \text{XLF Mom-1d}_{t+1} & \cdots \\
\cdots & \text{XLF Mom-5d}_{t-1} & \text{XLF Mom-5d}_{t} & \text{XLF Mom-5d}_{t+1} & \cdots \\
\cdots & \text{NDQ CP}_{t-1} & \text{NDQ CP}_{t} & \text{NDQ CP}_{t+1} & \cdots \\
\cdots & \text{NDQ Vel-1d}_{t-1} & \text{NDQ Vel-1d}_{t} & \text{NDQ Vel-1d}_{t+1} & \cdots \\
\cdots & \text{NDQ Vel-5d}_{t-1} & \text{NDQ Vel-5d}_{t} & \text{NDQ Vel-5d}_{t+1} & \cdots \\
\cdots & \text{NDQ Mom-1d}_{t-1} & \text{NDQ Mom-1d}_{t} & \text{NDQ Mom-1d}_{t+1} & \cdots \\
\cdots & \text{NDQ Mom-5d}_{t-1} & \text{NDQ Mom-5d}_{t} & \text{NDQ Mom-5d}_{t+1} & \cdots
\end{bmatrix}$$

**Figure 24.** Characteristic Matrix constructed with 36 parameters (column) and 1802 instances in time series (row)

### THE RESULTS AND OBSERVATIONS

The core mechanism of this predictive model is the inferences from the past to the future. To observe the effectiveness of the model, I rank each column vector of the Characteristic Matrix by the distance value. I identify ten instances of the lowest distance values as the best fit and ten instances of the highest distance values as the worst fit. I compare these two groups by the mean and variance of the return. I conduct this comparison for five sets of current Characteristic Vector and the Characteristic Matrix. For these operations I use the one-day returns and the five-day returns.

Figure 25 shows the effectiveness of the model in terms of mean returns for the four cases – the best fit for 1-day returns, the worst fit for 1-day returns, the best fit for 5-day returns, and the worst fit for 5-day returns. Five distinctive current Characteristic Vectors are demonstrated for each of the four cases. The actual values were calculated by subtracting the actual return values from the predicted return values to yield the model's predictive return error in percentage.

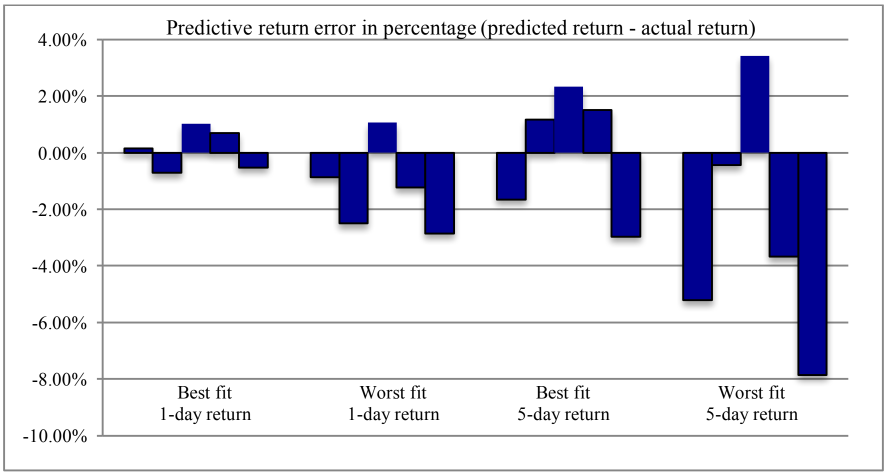



**Figure 25.** Model effectiveness comparison: Magnitude of predictive errors from the factual returns

I observe that the predictive errors are smaller in the best fits than the worst fits. I also note that the 5-day returns have higher errors compared with the 1-day returns.

Figure 25 shows the degree of variability in predicted returns for each of the four cases. The horizontal axis indicates the standard deviation of predicted returns by the model while the vertical axis show the predictive error in terms of corresponding standard deviation. The variability in the predicted return values is smaller in the best fit than the worst fit predictions. The 5-day returns generate higher variability than the 1-day return. I also observe that the predictive errors represented by multiples of standard deviation show comparable results.

I demonstrate that a good fit model of this methodology has certain predictive merit for a short-term prediction.

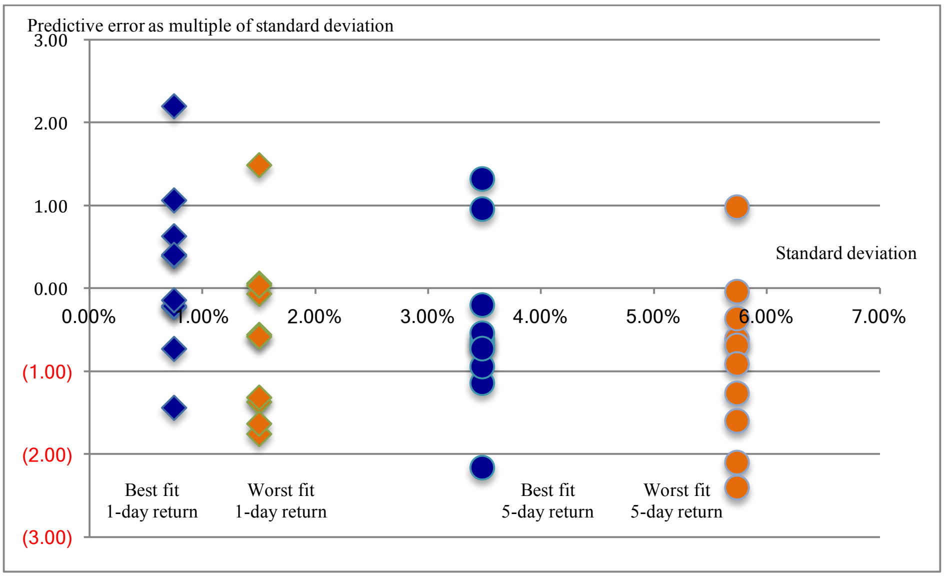



**Figure 26.** Errors in Predictive Capability

### Chapter Summary

I introduced a conceptual market return predictive model using a financial semantic space. The model offers a new methodology for market analysis by introducing schema in time series and making comparisons between the current and the historical instances. I demonstrated that this new model has certain predictive capabilities. The advantage of this model is its flexibility in adopting wide scope of data allowing adaptation of numerous applications. I note that this conceptual model serves as a good basis for a predictive model and its predictive accuracy may be significantly improved by optimizing the selection of characteristic parameters depending on the target financial assets under consideration.

# CHAPTER 7. AN IMPROVED PREDCTIVE MODEL AND A METHODOLOGY TO SOLVE INVERSE PROBLEM

I introduced a conceptual model that predicts future price movement of a financial asset of interest based on multidimensional economic and market factors. While I demonstrated possibilities of a wide range of adaptation of multidimensional parameters, the predictive effectiveness of the model needs further improvement. Such improvements may be made based on the nature of the asset under consideration as well as generic selection criteria of characteristic parameters. In the preceding chapter, I observed that a wide range of numerous dimensions of characteristic parameters may have caused diversification of instances making it more unlikely to find close match of similar vectors. In this experiment, I take a limited approach in selecting characteristic parameters that characterize the price movements of an asset. Not only in scope but the number of characteristic parameters is significantly reduced to only market oriented parameters disregarding other seemingly relevant socioeconomic factors.

I then introduce a very important methodology that can solve inverse problems for Mathematical Model of Meaning. I discovered this methodology when I was constructing the predictive market movement model.

## THOUGHT STANCE IN PRICE PREDICTION

There exist two major schools that attempt to explain and predict price movements in financial markets. One is chartists, or technical analysts, who recognize historical price movement patterns and apply them in the current situation. The other school is fundamentalists who examine the intrinsic value of the asset of interest. By comparing the assessed intrinsic value and the market-evaluated price, fundamentalists make buy-hold-sell decisions. In this experiment, I take a position similar to the chartists. However, my position is different from that of the chartists because I do not seek graphical patterns of price movements. I take instantaneous view of the asset price movements and trade volume with limited cumulative historical change to generate a conclusive predictive model.

### THE PRICE PREDICTION MODEL – A HYPOTHESIS

I hypothesize that an effective daily market return prediction model may be built using a semantic metric space with a relatively small set of characteristic parameters to constitute a Descriptive Vector for each instance. I take the latest instance of the current descriptive vector and compare it with the historical Descriptive Vectors at each instance in time series. By identifying the highest similarity between two Descriptive Vectors and examining how the price moved at the next instance of the similarity-identified historical instance, I should be able to predict what happens at the next instance in of the latest instance – predicting the one-day price movement.

### OBJECTIVES OF THE EXPERIMENT

The objective of this experiment is to build a price predicting model with varying number of characteristic parameters in the descriptive vectors to discover variance in the model effectiveness.

### EXPERIMENT OVERVIEW

I choose an extremely liquid highly traded asset in major exchanges thus excluding any liquidity-related price movement factors. I choose solely market-related factors to generate Characteristic Parameters to simplify the model - they are price changes, volume traded, and price deviation from moving averages. I start with a whole set of characteristic parameters – the entire span of the semantic space. I construct a descriptive vector for each instance in time series and compare each one with the current descriptive vector to find the most similar set of vectors between the historical and the current descriptive vectors. The similarity between vectors shall be calculated by the Euclidian distance between the respective vectors. I then shift the one instance back in defining the current descriptive vector and conduct the same process. This will allow exhaustive comparison for the given set of data points. I continue the backward time shift of the current descriptive vector until it reaches the very beginning of data minus one instance in the time series. Once this comparison process is complete, I take subspaces of the entirely spanned semantic space and repeat the process and see if the selected subspaces will generate different predictive model effectiveness. I choose different number of dimensions as the subspace construct to compare the outcome in the model effectiveness.

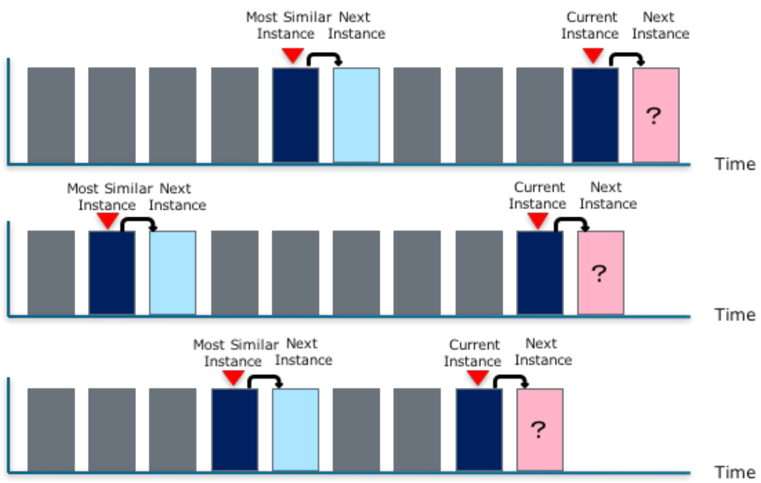



**Figure 27.** Mechanism of the Predictive Model using Financial Semantic Space

## DATASET

I use an Exchange Traded Fund, SPY, that mimics the S&P500 index. This ETF can be considered as "the market" is traded in very high volume and provides very high liquidity. A data set was obtained from a publicly available source[58]. The raw data set has the following characteristics:

***Raw Data***

- US ETF "SPY" closely tracking the S&P 500
- Daily data from January 29, 1993 to May 2, 2016 – 5858 valid samples
- Opening Price of the day in US$
- Highest Price during the day in US$
- Lowest Price during the day in US$
- Adjusted Closing Price of the day in US$ (split and dividend adjusted)
- Volume traded during the day

From the raw data above, I generate the following Characteristic Parameters.

***Characteristic Parameters***

I set three categories of Characteristic Parameters as ratios derived from the raw data:

**Change in Return** – Prices are compared and returns are calculated by taking the ratio of price differences (gain or loss) and the base price of interest. Various price categories and associated time durations characterize different return profiles. Higher positive returns are sought by long investors, call option holders, put option issuers while higher negative returns are sought by short investors, put option holders, and call option issuers.

**Volume Traded** – Volumes are the number of units traded for a given asset in the exchange. Higher volume indicates an active trading and lower trade indicates relatively inactive trading. Volume is an important factor that indicates market's interest in trading. High traded volume implies that there is high demand to buy and sell the asset. High volume trade may lead to price advancement, price decline or unchanged price depending on the balance of buy/sell demand.

[58] The data was obtained from Yahoo Finance: https://finance.yahoo.com/quote/SPY/history?p=SPY

**Deviation from Moving Average –** A moving average is an average of prices for a given window of time period. The duration of time window is chosen depending on the use purpose. Deviation from a moving average indicates the degree of price separation in percentage at the instance of interest from the moving average to that instance. A positive deviation indicates a higher current price with respect to the moving average while a negative deviation means a lower current price compare with the moving average.

## CONSTRUCTING A SEMANTIC SPACE

Construct a semantic metric space with a set of Characteristic Parameters for each of the 3 cases. Then fill with entries as instances in time series. This means creating an m x n matrix for each case where m is the number of instances in time series (3340) while n is the dimension represented by the number of Characteristic Parameters (19 for case-1, 4 for Case-2, and 3 for Case 3).

**Table 9.** Semantic Space Construct with Various Characteristic Parameter Sets

| **<Case 1>** | **<Case 2>** | **<Case 3>** |
|---|---|---|
| **Change in Return** | **Change in Return** | **Change in Return** |
| 1: Open/Close$_{-1d}$ | 4: Close /Close$_{-1d}$ | 4: Close /Close$_{-1d}$ |
| 2: High/Close$_{-1d}$ | **Volume** | **Volume** |
| 3: Low/Close$_{-1d}$ | 17: Volume Ratio | 17: Volume Ratio |
| 3: Low/Close$_{-1d}$ | **Deviation from MA** | **Deviation from MA** |
| 4: Close/Close$_{-1d}$ | 18: 25$_{d}$MA | 18: 25$_{d}$MA |
| 5: Close/Close$_{-2d}$ | 19: 300$_{d}$MA | |
| 6: Close/Close$_{-3d}$ | | |
| 7: Close/Close$_{-4d}$ | | |
| 8: Close/Close$_{-1w}$ | | |
| 9: Close/Close$_{-1m}$ | | |
| 10: Close/Close$_{-3m}$ | | |
| 11: Close/Close$_{-6m}$ | | |
| 12: Close/Close$_{-1y}$ | | |
| 13: Close/Close$_{-2y}$ | | |
| 14: Close/Close$_{-3y}$ | | |
| 15: Close/Close$_{-5y}$ | | |
| 16: Close/Close$_{-10y}$ | | |
| **Volume** | | |
| 17: Volume Ratio | | |
| **Deviation from MA** | | |
| 18: 25$_{d}$MA | | |
| 19: 300$_{d}$MA | | |

**Explanation of the Characteristic Parameters**

1. Change at Opening Price from the last Adjusted Closing Price (%)
2. Change of the Highest Price from the last Adjusted Closing Price (%)

3. Change of the Lowest Price from the last Adjusted Closing Price (%)
4. Change at the Adjusted Closing Price from the last Adjusted Closing Price (%)
5. Change at the Adjusted Closing Price from the Adjusted Closing Price at 2 trading days prior (%)
6. Change at the Adjusted Closing Price from the Adjusted Closing Price at 3 trading days prior (%)
7. Change at the Adjusted Closing Price from the Adjusted Closing Price at 4 trading days prior (%)
8. Change at the Adjusted Closing Price from the Adjusted Closing Price at 5 trading days prior (%)
9. Change at the Adjusted Closing Price from the Adjusted Closing Price at 1 month prior (%)
10. Change at the Adjusted Closing Price from the Adjusted Closing Price at 3 months prior (%)
11. Change at the Adjusted Closing Price from the Adjusted Closing Price at 6 months prior (%)
12. Change at the Adjusted Closing Price from the Adjusted Closing Price at 1 year prior (%)
13. Change at the Adjusted Closing Price from the Adjusted Closing Price at 2 years prior (%)
14. Change at the Adjusted Closing Price from the Adjusted Closing Price at 3 years prior (%)
15. Change at the Adjusted Closing Price from the Adjusted Closing Price at 5 years prior (%)
16. Change at the Adjusted Closing Price from the Adjusted Closing Price at 10 years prior (%)
17. Volume Traded as a percentage of the maximum volume during the observed period (%)
18. Price difference from the 25-day trailing moving average (%)
19. Price difference from the 300-day trailing moving average (%)

The valid parameter spans from January 28, 2003 to May 2, 2016 due to the 10-year comparison

## PROCEDURE

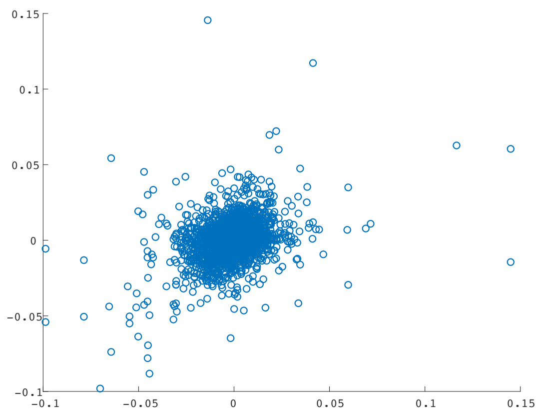
From the perspective of the Context-dependent Multidimensional space analysis, I define each row of the m x n matrix as the Instance Vector. I set Current Vector as a reference point whose next instance in time series is the one to be predicted. The Euclid distance is calculated between the Current Vector and each of the Instance Vectors. Then the Euclidean distance is compared to identify the Instance Vector that exhibits the minimum Euclidean distance. That is the Instance Vector that resemble the Current Vector the most. Only Instance Vectors that precede the Current Vector in time is compared. The procedure is repeated by moving the Current Vector into the past until it reaches m-1.

I use three subspace selection. Case 1 is the entire space whose Instance Vectors are constituted with 19 Characteristic Parameters. Case 1 serves the basis of comparison with dimension-reduced subspace selections as in Case 2 and Case 3. Case 2 represents a subspace whose dimension is reduced to 4. Case 3 represents a subspace with only 3 Characteristic Parameters. These three cases were chosen to represent subspace selection for comparison.

## THE RESULTS AND OBSERVATIONS

Experiments in all three cases were conducted. In all cases the model demonstrated return prediction capability. The predicted return and historical return of the most similar historical record showed positive correlation. In addition, this particular experiment exhibited that the smaller number of dimensions for the subspace selection performed more precise return prediction with higher correlation. I proved my hypothesis that an effective daily market return prediction model may be built using a semantic metric space with a relatively small set of Characteristic Parameters to constitute an Instance Vector for each instance.

Predicted Daily Return (%)

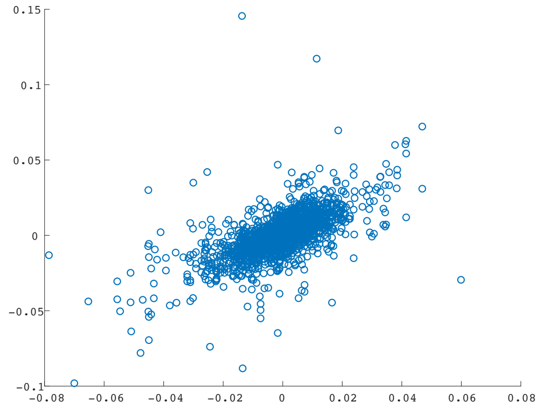



Actual Daily Return (%)

**Correlation = 0.3453**

**Figure 28.** Effectiveness of Predictive Model – Case 1

Predicted Daily Return (%)

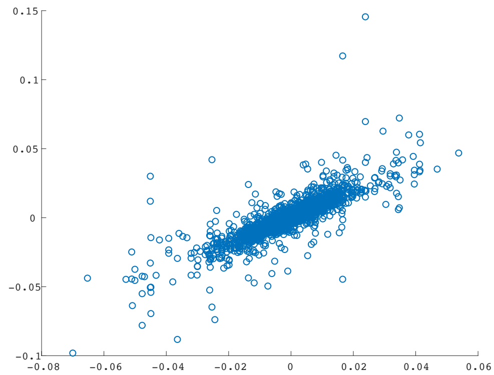



Actual Daily Return (%)

**Correlation = 0.6503**

**Figure 29.** Effectiveness of Predictive Model – Case 2

Predicted Daily Return (%)




Actual Daily Return (%)

**Correlation = 0.8362**

**Figure 30.** Effectiveness of Predictive Model – Case 3

Further improvement to the model may be made by employing other subspace selection. All combinations of Characteristic Parameters that constructs subspaces can be tested to find the optimal subspace selection that produces the most effective predictive capability for the model. This exhaustive trial shall produce the best fit model for the given set of Characteristic Parameters. Even further enhancement may be possible by introducing other Characteristic Parameters that may yield potentially even better prediction model. There are 524,287 subspace combinations excluding the case where the selected Characteristic Parameters is zero.

$$\text{Number of Subspace} = \sum_{k=0}^{19} \binom{19}{k} - 1$$

The important fact is that there exists a relationship between the choice of subspace and the resulting precision of the prediction model.

### SOLUTION TO THE INVERSE PROBLEM IN THE MMM

I built a return predictive model as an application of the Mathematical Model of Meaning (MMM). I constructed a semantic space with Characteristic Parameters and mapped each instance onto the space as Descriptive Vector. The "context" was the choice of Characteristic Parameters which translates to the selection of the subspace. The input parameters for the model are the instance entry and the subspace selection, and the output was the precision of the predictive model as the correlation between the predicted value and actual value. As the input was given, the output was produced as MMM was designed as a context-dependent forward problem solving methodology.

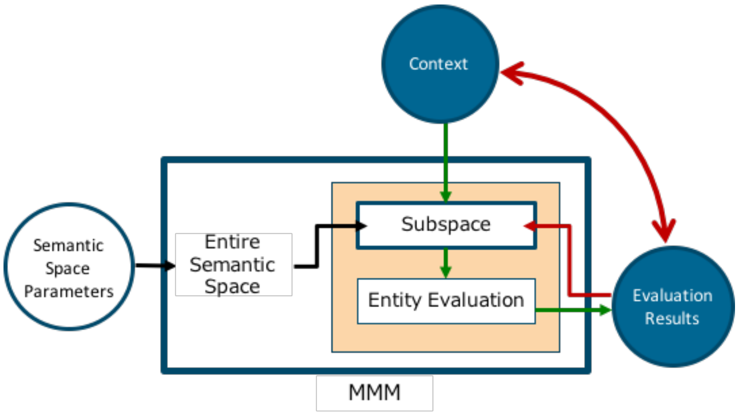



**Figure 31.** Forward and Inverse Problem with the MMM

***Changing the Input Parameter***

The model effectiveness is represented by the correlation between the actual value and the predicted value of market return for the financial asset. The correlation value is now used as an input.

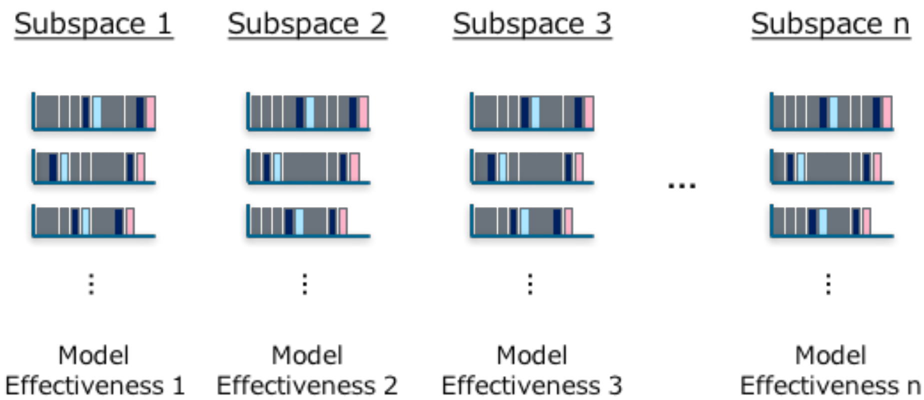



**Figure 32.** Subspace – Model Outcome Relationship

As in this experiment, when the output is known because even the “future” values to be predicted already exist as historical factual values, such output can be used as input to find the subspace selection, the context, as the output. It is this very mechanism that solves the inverse problem of the Mathematical Model of Meaning.

## Chapter Summary

I built a market return prediction model that utilizes Context-dependent Multidimensional Semantic Space Analysis method and demonstrated the effectiveness of the model. The level of effectiveness may be further improved by optimizing the Descriptive Vector as a set of Characteristic Parameters. Defining Descriptive Vector is equivalent to choosing the subspace of the semantic space.

The existence of a relationship between the output as the model precision and the input as subspace selection leads to an extremely important discovery of solving the inverse problem of the Mathematical Model of Meaning. The subspace selection can be found as an output when the model precision is given as an input. The MMM research has been done as solving forward problems in many applications. The new methodology to solve the inverse problem for the MMM shall open a magnitude of research topics.

## *Appendix*

Below is a sample Matlab code that compares the current and historical instances to produce predictive values.

```
function  C = Comp( P, I ,S, E)
% [Pactual, Ppredict] = Comp (P, I)
%   Create a two-column matrix of actual and predicted return
values
%    Input:
%         P = Price change from the previous day  Pclose (3340 x 1)
%         I = Index vector produced by Find_min_distance
%         S = Comparison start index in I (min = 2)
%         E = Comparison end index in I (max = 3333 > min)
%    Output :
%         C  =  [Pactual, Ppredict]
%         Pactual = P (index with the highest similarity I-1)
%         Ppredict = P (index for the comparison  I-1)

for i = S:E

C(i,1) = P(I(i-1));
C(i,2) = P(i-1);

% [Pactual, Ppredict] = (P(I(i-1), P(i-1));

end

function [ M,I ] = Find_min_dist( X )

 % Find_min_dist
% Calculate Euclid Distance for each row and find the minimum,
return the index for the minimum as the row number
%   X is an input matrix: the row is data entry and the column is
n-dimensional vector
%   M is the minimum value vector
%     M(1) is the minimum value between the 1st row and the later
rows of X
%     M(2) is the minimum value between the 2nd  row and the later
rows of X
%     M(n) the minimum value between the nth row and the later
rows of X
%   I is the index vector pointing to the minimum values
%     I(1) is the index to the minimum value between the 1st row
and the later rows of X
%     I(2) is the index to the minimum value between the 2nd row
and the later rows of
%     I(n) is the index to the minimum value between the nth row
and the later rows of X

dist = pdist(X);
D = squareform(dist);

%   Column 1 of D is the result of comparison between the 1st row
of X and other rows of X
%   Column 2 of D is the result of comparison between the 2nd row
of X and other rows of X
%   Column n of D is the result of comparison between the nth row
of X and other rows of X

L = tril(D);
L(~L)=inf;
[M,I] = min(L,[],1)

end
```

# CHAPTER 8. SYSTEM CONFIGURATION

In this chapter, I describe a typical configuration of a context-dependent integrated multi-domain corporate evaluation system. I assume a three-domain system as discussed in the previous chapter in the field of finance, technology, and brand as an illustration[59]. A system can accommodate other domains by adopting appropriate Characteristic Parameters. Employing the expertise in selecting appropriate Characteristic Parameters, translating the user context and company attributes into the Characteristic Parameters is still required as in the case of this example.

## SYSTEM ARCHITECTURE

A typical system architecture of a Context-dependent Integrated Multidimensional Corporate Analysis System is described below. The system consists of four major components: an external Application Software, Corporate Analyzer Subsystem, Context/Query Qualification Subsystem, and external Distributed Databases.

### Application Software

Application Software manages the user presentation layer of the system. It manages the user interface for context word input as well as query word input. For input and output, it should have graphical user interface for easier input and better presentation of the results. It should also allow real-time interactivities between the input and output to allow the user to reach optimal results.

[59] Shin ITO, Yasushi KIYOKI, A Context-based Multi-Dimensional Corporate Analysis Method, Information Modelling and Knowledge Bases XXIV, pp255-270, 2013

### Corporate Analyzer Subsystem

Corporate Analyzer Subsystem manages the interface with the Application Software, Context/Query Qualification Subsystem, and external distributed databases via the wide-area computer network. It consists of the following three components:

***User I/O Processing Module***

The User I/O Processing Module takes in the user context and query words from the application software, inquires the Context/Query Qualification Subsystem for a set of appropriate context and query specifications. The module can be designed to accept textual words that will be translated to actual context and target company data. A more explicit method may be deployed to avoid any ambiguous interpretation of textual word based input method by introducing a user-selectable preset options. Such preset options need to cover a very wide variety of choices with ease of reaching to the desired intent.

***Data Analysis Module***

The Data Analysis Module takes in context specifications, query specifications, and data set from external distributed databases to evaluate companies and generate the results that are passed to the User I/O Processing Module for user consumption.

***Data Fetch Module***

The Data Fetch Module receives the location information of each of the line item from the Corporate Analyzer Subsystem and issues inquiry to receive appropriate data from the distributed databases. Once proper data is received, it returns it to the Data Analysis Module.

### Context/Query Qualification Subsystem

Context/Query Qualification Subsystem accepts context and query words from the User I/O Processing Module and returns context specifications and query specifications to it. The subsystem may utilize the Mathematical Model of Meaning

(MMM)[60] to extract the meaning of user-specified context words and query words in reference to field-specific terms and generates context specifications and query specifications. Context specification contains two types of data - one is a set of objectives in association with companies and the other is a set of conditions and restrictions to be considered in evaluating companies. Query specification is a list of target company names to be evaluated. These specifications are passed to the Data Analysis Module.

### Distributed Databases

Distributed Databases are the sources of data that house data to be used for the corporate analyses. Databases include both public and private domains and span numerous categories.

---

[60] T. Kitagawa and Y. Kiyoki, ``A mathematical model of meaning and its application to multidatabase systems," Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993.

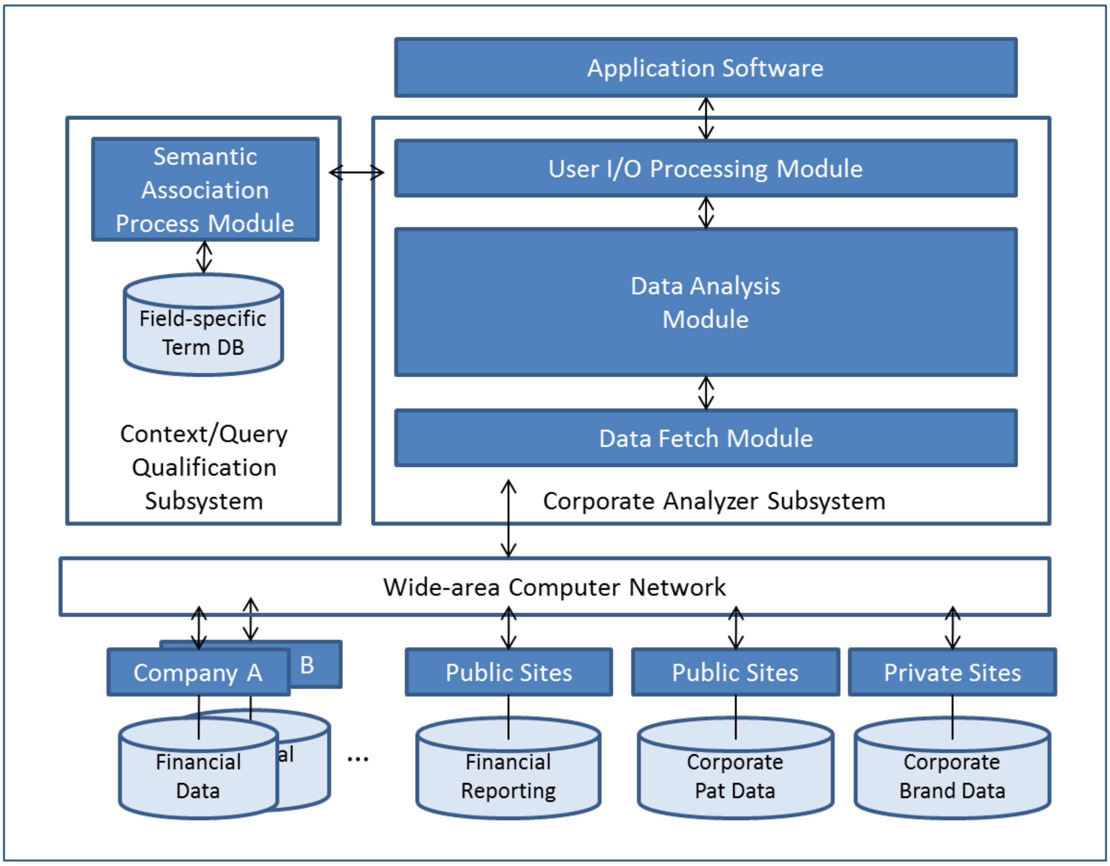



**Figure 33.** A typical Configuration of Context-dependent Integrated Multi-Domain Corporate Analysis System

### Data Analysis Module

The Data Analysis Module of the Corporate Analyzer Subsystem is the central module of the entire system and discussed here in detail. The Data Analysis Module comprises of five distinctive sub-modules as depicted in Figure 27. They are:

- Semantic Space Construction Module
- Characteristic Parameter Value Assign Module
- Normalization/Integration Module
- Context Translator Module
- Evaluation Process Module

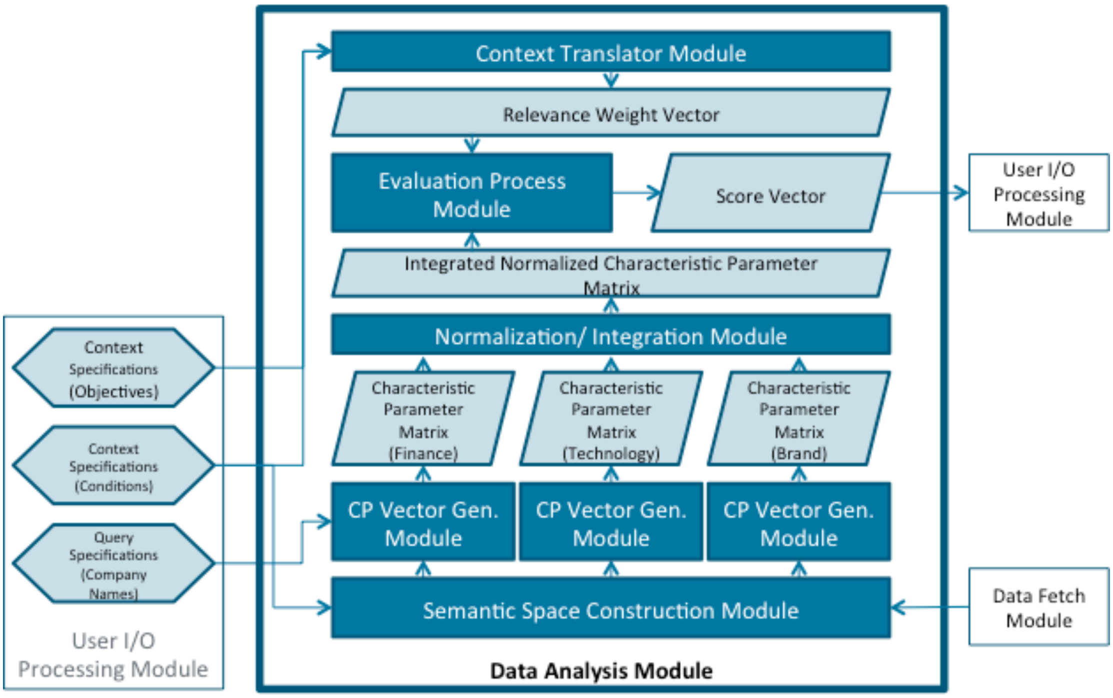



**Figure 34.** Data Analysis Module

**Semantic Space Construction Module**

Referring to the user context and target companies, Semantic Space Construction Module functions in two steps – first selects and manages domains to describe companies, and second selects appropriate Characteristic Parameters to construct a semantic space for each of domain. This module requires general domain intelligence and expertise for the aforementioned two-step procedure. Care is taken in selecting a set of Characteristic Parameters such that they sufficiently represent companies but simultaneously they do not overcrowd the semantic space causing wasted computing power and potential malfunction. Semantic space Construction Module assures mutual orthogonality in the final set of domain-specific Characteristic Parameters. The module may conduct eigenvalue decomposition to reduce dimensions and ascertain orthogonality.

**Characteristic Parameter Vector Generation Module**

Based on the set of Characteristic Parameters produced by the Semantic Space Construction Module, Characteristic Parameter Vector Generation Module generates vector representation of each company in way of the selected Characteristic Parameters. The module extracts company data passed by the Data Fetch Module to create such vector representation. As an illustration, when a Characteristic Parameter is a financial ratio, the module appropriately selects the correct line items from a set of financial statements of the correct publication date, possibly collects market related data, and proceed with appropriate calculations to create the financial ratio value for a company. The process is repeated for each Characteristic Parameter for all target companies.

**Normalization-Integration Module**

The Normalization-Integration Module creates the integrated Semantic Space from multiple domain- specific semantic spaces. To create an integrated space, Characteristic Parameter Vectors for all companies need to be normalized for comparable measurement across different measurement system in each of the Characteristic Parameters. Once normalization is complete, domain specific semantic space is united to create an integrated semantic space.

The Module also ascertains orthogonality among Characteristic Parameters in the integrated semantic space.

**Context Translator Module**

Context Translator Module is a module that converts user contexts to Characteristic Parameter based representation in the integrated semantic space. Such translation is accomplished by assigning weight to each of the Characteristic Parameters in the integrated semantic space. The general domain intelligence and expertise in the Semantic Space Construction Module may be utilized in translating user contexts into Characteristic Parameters. The resulting Context Vector will be used to select an appropriate semantic subspace for analyses and evaluation of entities in the Evaluation Process Module.

### Evaluation Process Module

The Evaluation Process Module selects a subspace and conducts analyses to evaluate the entities. Entity Vectors for the companies under analyses are projected onto the subspace and the Euclidean norm is calculated to generate the relevance score for each company. The scores are ranked and output to the User I/O Processing Module.

## CHAPTER SUMMARY

In this chapter, I presented a typical context-dependent integrated multi-domain corporate evaluation system. The core module is the Evaluation Process Module and variations in this module allows flexible analyses and evaluation of entities. A similar system configuration may be utilized to implement a market return predictive model with time series data as described in earlier chapters.

# CHAPTER 9. SUMMARY AND FUTURE WORK

I presented new methodologies for analyzing and evaluating companies and financial assets using integrated multi-domain semantic space. The methodologies empower the user and provide flexibility in customizing the evaluation of companies and assets. The economic activities have become a new area of application for the Mathematical Model of Meaning.

I demonstrated a viable market return predictive model as an application of the semantic space analysis method. I also presented a new methodology to solve inverse problem for the Mathematical Model of Meaning. By identifying the relationship between the user context and the outcome of the evaluation method, the user context can be found when the outcome is given. This discovery opens a new research areas that further develops applications of the Mathematical Model of Meaning.

I identify the following areas for further research.

### INTANGIBLE ASSET REPRESENTATION

Intangible assets become a major performance driver for companies but corporate use of such assets have been suboptimal due to difficulty in properly inventorying features for different use. Appropriate identification, evaluation, and representation of intangible assets that reflect use objectives are highly demanded for better asset management. Representing the intangibles by Characteristic Parameters and providing users with the flexibility to evaluate the assets in their context will allow more effective use of the assets. Identification of Characteristic Parameters and defining subspace for numerous contextual needs are application research areas for the context-dependent asset evaluation system using semantic space.

### HUMAN CAPITAL MANAGEMENT

Human capital management is difficult and is a highly human-centric task. Though there exist certain database systems that store and manage features about a person, such features tend to be low-context and are not sufficient in managing for organizational use. Thus, employee task assignment is often made by managers by their past experiences and reputations with the employees. A context-dependent human capital management system may effectively identify employees most suited for given tasks. Employees are represented in an integrated semantic space and tasks are expressed as contextual needs. Such system is particularly beneficial in large organizations where managers' knowledge about employees and tasks are limited to their own locales. The research area is to how to identify high-context features about a person and continually enhancing such data for much improved representation about a person in the form of Characteristic Parameters.

### APPLICATIONS FOR SOLVING INVERSE PROBLEMS WITH THE MMM

Inverse problems can be solved when the relational patterns are obtainable between the outcome of the evaluation and the contextual setting. I showed historical market price moves as an experimental example. There are numerous factual records from which such relational pattern can be obtained and thus the inverse problems can be solved. Solving an inverse problem shall allow higher precision model making by discovering the optimal subspace of a semantic space. New research areas are how to efficiently discover the relational patterns between the output and the user context, and how to create robust domain-specific models using the solution to MMM inverse problems.

# References


[1] T. Kitagawa and Y. Kiyoki, ``A mathematical model of meaning and its application to multidatabase systems," Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993.

[2] Porter, M. E. "How Competitive Forces Shape Strategy." Harvard Business Review 57, no. 2 (March–April 1979): 137–145.

[3] Porter, M. E. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press, 1980.

[4] Bensanko, David, David Dranove, and Mark Shanley. *The Economics of Strategy*. John Wiley & Sons, Inc. 1996

[5] Day, George S., David J. Reibstein. *Wharton on Dynamic Competitive Strategy*. John Wiley & Sons, Inc. 1997.

[6] Mintzberg, Henry, Bruce Ahlstrand, and Joseph Lampel. *Strategy Safari: The Complete Guide through the Wilds of Strategic Management, 02 Edition*. 2009

[7] White Gerald I., Ashwinpaul C. Sondhi, and Dov Fried, *The Analysis and Use of Financial Statements*, John Wiley & Sons, Inc, 1994.

[8] Bodie, Zvi, Alex Kane, and Alan J. Marcus. *Investments, Second Edition*. Irwin, Inc. 1993.

[9] Richard A. Brealey, and Stewart C. Myers. *Principles of Corporate Finance, Fourth Edition*. McGraw-Hills, Inc. 1991.

[10] Richard H. Thaler. The Winners Curse. The Free Press. 1992

[11] Richard H. Thaler, and Cass R. Sunstein. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press. 2008.

[12] Erich A. Helfert. *Techniques of Financial Analysis: A Practical Guide to Managing and Measuring Business Performance. Eighth Edition*. Irwin Professional Publishing. 1994

[13] Kiyoki, Y. and Kitagawa, T., "A metadatabase system for supporting semantic interoperability in multidatabases," Information Modeling and Knowledge Bases, IOS Press, Vol. V, pp. 287-298, 1994

[14] T. Kitagawa and Y. Kiyoki, ``A new information retrieval method with a dynamic context recognition mechanism," Proceedings of 47th Conference of International Federation for Information and Documentation, pp.210-215, Oct. 1994.

[15] Y. Kiyoki and T. Kitagawa, ``A semantic associative search method for knowledge acquisition," Information Modelling and Knowledge Bases (IOS Press), Vol. VI, pp.121-130, 1995.

[16] Y. Kiyoki, T. Kitagawa and T. Hayama, ``A metadatabase system for semantic image search by a mathematical model of meaning," Multimedia Data Management -- using metadata to integrate and apply digital media --," McGrawHill(book), A. Sheth and W. Klas(editors), Chapter 7, 1998.

[17] Y. Kiyoki, T. Kitagawa and T. Hayama, ``A metadatabase system for semantic image search by a mathematical model of meaning, ACM SIGMOD Record, (refereed as the invited paper for special issue on metadata for digital media), Vol.23, No. 4, pp.34-41, Dec. 1994.

[18] Y. Kiyoki and S. Ishihara: "A Semantic Search Space Integration Method for Meta-level Knowledge Acquisition from Heterogeneous Databases," Information Modeling and Knowledge Bases (IOS Press), Vol. 14, pp.86-103, May 2002

[19] Bodie, Z., Kane, A., Marcus, A., "Investments Second Edition," Ch 19 Financial Statement Analysis Richard D. McGraw-Hill, Inc. 2009

[20] Brealey, R., Myers, S., "Principles of Corporate Finance Fourth Edition", Chapter 29 Financial Analysis and Planning, McGraw-Hill, Inc. 2003

[21] White, G., Sondhi, A., Fried, D., "The Analysis and Use of Financial Statements," 1994 John Wiley and Sons, Inc.

[22] Ito, Kiyoki "A Context-based Multi-factor Corporate-analysis Method Using Finance-Related Information" 2012.

[23] Damodaran, A., "Investment Valuation," John Wiley and Sons, Inc. 1996.

[24] Bodie, Ziv, Alex Kane, and Alan J. Marcus. Investments, Second Edition. Richard D. Irwin,Inc. 1993.

[25] White, G., Sondhi, A., Fried, D., "The Analysis and Use of Financial Statements," John Wiley and Sons, Inc. 1994.

[26] Capital IQ online information services, a Standard & Poor's Business: https://www.capitaliq.com/home.aspx

[27] Mergent Online, Mergent Inc., 2008: http://www.mergentonline.com/

[28] Yahoo! Finance: http://biz.yahoo.com/r/

[29] Ito Shin, Kiyoki Yasushi, "A Context-based Multi-Dimensional Corporate Analysis Method" INFORMATION MODELING AND KNOWLEDGE BASES 2012, Vol. XXIV, IOS Press.

[30] Ito Shin, Kiyoki Yasushi, "A Multidimensional Market Analysis Method Using Level-Velocity-Momentum Time-Series Vector Space" INFORMATION MODELING AND KNOWLEDGE BASES 2014, Vol. XXV, IOS Press.

[31] Bettman, Jenni L. et. al., "Fundamental and technical analysis: substitutes or complements?", Accounting & Finance; Mar2009, Vol. 49 Issue 1, p21-36, 16p.

[32] Hickman, Kent, "A Comparison of Stock Price Predictions Using Court Accepted Formulas, Dividend Discount, and P/E Models", The Journal of the Financial Management Association; Summer1990, Vol. 19 Issue 2, p76-87, 12p.

[33] McCurtain, Robert, “Getting Technical with Economic Data”, Futures: News, Analysis & Strategies for Futures, Options & Derivatives Traders; Nov2010, Vol. 39 Issue 11, p44-50, 4p

[34] Boni, Leslie. et. al., “Analysts, Industries, and Price Momentum”, Journal of Financial & Quantitative Analysis; Mar2006, Vol. 41 Issue 1, p85-109, 25p.

[35] Bulkley, George., et. al., “Can the Cross-Sectional Variation in Expected Stock Returns Explain Momentum?”, Journal of Financial & Quantitative Analysis; Aug2009, Vol. 44 Issue 4, p777-794, 18p.

[36] Treynor, Jack L., et. al., “In Defense of Technical Analysis”, Journal of Finance; Jul1985, Vol. 40 Issue 3, p757-773, 17p.

[37] Lo, Andrew W. et. at., “Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation”, Journal of Finance; Aug2000, Vol. 55 Issue 4, p1705-1765, 61p.

[38] Lai, Hung-Wei, et. al., “Technical Analysis, Investment Psychology, and Liquidity Provision: Evidence from the Taiwan Stock Market”, Emerging Markets Finance & Trade; Sep/Oct2010, Vol. 46 Issue 5, p18-38, 21p.

[39] Varadharajan, P., “Effectiveness of technical analysis using candlestick chart for selection of equity stock in Indian capital market”, Journal of Contemporary Management Research; March 2011, Vol. 5 Issue 1, p12-23, 12p.

[40] Huang Wei, et. al., Neural Networks in Finance and Economics Forecasting”, International Journal of Information Technology & Decision Making; Mar2007, Vol. 6 Issue 1, p113-140, 28p.

[41] Zuo Yi., et. al., “Up/Down Analysis of Stock Index by Using Bayesian Network”, Engineering Management Research; Vol. 1, No. 2; 2012.

[42] Tsai Chih-Ling, et. al., “Does a Bayesian approach generate robust forecasts? Evidence from applications in portfolio investment decisions”, Published on line - The Institute of Statistical Mathematics, Tokyo 2009.

[43] Otrok, Christopher and Whiteman, Charles H., “Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa”, The NBER/NSF Seminar on Forecasting and Empirical Methods in Macroeconomics, July 1996.

[44] Wikipedia, http://en.wikipedia.org/wiki/Williams_%25R

[45] Innography: http://www.innography.com/

[46] Interbrand Best Global Brands 2006, 2007, 2008, 2009, 2010, 2011: http://www.interbrand.com/en/best-global-brands/BGB-Interactive-Charts.aspx

[47] Interbrand Best Global Green Brands 2011: http://www.interbrand.com/en/best-global-brands/Best-Global-Green-Brands/2011-Report/BestGlobalGreenBrandsTable-2011.aspx