

局所化指向テキストマイニングの実践と評価

2012年度

竹内 広宜

学位論文 博士(工学)

局所化指向テキストマイニングの実践と評価

2012年度

慶應義塾大学大学院理工学研究科

竹内 広宜

要旨

テキストマイニングでは、特定の文書集合においてキーワードの出現頻度を求め、傾向や規則を発見することが行われている。この時、分析に用いる観点や辞書を準備する必要があり、それらの初期設定や選択が実践上の課題となっている。本論文では、分析観点や辞書の設定や利用に局所化手法を適用した。そして、市場分析および会話分析に対して、局所化を利用した分析手法を提案し、実践例を通してその有用性の評価を行った。

以下、本論文の構成について述べる。

はじめに、第1章において、本論文の背景、課題、目的について述べる。

第2章では、本研究の関連技術としてテキストマイニングで用いられる自然言語処理技術およびテキストマイニングの活用手法について述べるとともに、これらの関連研究について述べる。

第3章では、会話分析を対象とし、タスクを持った会話からタスクの成功につながる発言パターンの抽出を行う。その際、局所化手法として、冗長な発言を含む会話データからタスクの成功に寄与する重要発言区間を同定する手法と、同定した重要発言区間からタスクの成功に関連するキーワードを抽出する手法を提案した。実践例としてコンタクトセンター受託企業で収集されたレンタカーの予約会話データを対象とした。そして、顧客が予約した車を取りに来る/来ないと結果が異なる予約会話間の差異分析を行った。長い会話の中から結果に影響を与える発言区間として顧客の最初の発言および提案時の発言を同定し、その中から結果に関連する発言パターンを抽出した。そして抽出した発言パターンを元にオペレーターへの教育に行い、予約された車の利用率を約3%向上することができた。

第4章では、市場分析を対象とし、自由回答および選択回答形式のアンケートデータから次期購買層の発見につながるルール抽出を行う。その際、分析するキーワードを限定する局所化手法として、順序関係を持った顧客属性に対して頻度が増加・減少する傾向を持つキーワードをランキングする手法を提案した。そして、データマイニングによるルール発見の結果から、テキストマイニングで関連があると分析したキーワードと顧客属性の組を含むルールをフィルタリングする手法を提案した。実践例として、生ごみ処理機の市場分析を目的とした購買者・非購買者へのアンケートデータから次期購買層の発見につながるルール抽出を行った。実践例では、提案手法により、マーケティング専門家が解釈・評価を行うルール数を、精度を保ちながら約1/3に削減することができた。

最後に第5章で、本論文のまとめと今後の課題および展望について述べる。

Title: Practice and Evaluation on Localization-Oriented Text Mining

Abstract:

In the text mining analysis, we usually try to get the frequencies of keywords in a selected document set. For such an analysis, it is needed to define view points and prepare dictionaries in advance. In this thesis, we considered to apply localization methods for preparing dictionaries and selecting analysis view points. We proposed localization-oriented text mining for marketing analysis and conversation analysis.

This thesis is organized as follows:

Chapter1 describes backgrounds, issues and purpose of this study.

Chapter2 introduces text mining technologies and researches on text mining applications as related technologies and studies around this thesis.

Chapter 3 describes the conversation analysis where we try to find utterances leading to insights that improve business from the conversation data. As a localization method, we proposed a method to identify important segments from the conversations which are often long and redundant, and extract effective expressions from the important segments to define the viewpoints. We applied the method to the conversation data from a car rental reservation center. We identified customers' first utterance and utterances in the proposal as important segments and extracted expressions relating to the reservation conversation where customers picked up the reserved cars. Through the education for the operators based on the extracted utterance patterns, we could improve the picked-up ratio of the reserved cars by about 3%.

Chapter 4 describes the market analysis to find rules for potential customers from a questionnaire survey data for a product. As a localization method, we proposed a method for ranking keywords correlating to segments that have ordering. We also proposed a method for filtering data mining results by using the trend analysis results by text mining. We applied the methods to the market analysis of a garbage disposal. In this analysis, the number of extracted rules that the marketing expert had to assess was reduced into about 1/3.

Finally, in chapter5, we conclude this thesis and point out future directions.