慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	多次元空間計量によるユーザ•コンテクスト対応イベント情報抽出•配信システムの構成
Sub Title	
Author	柏木, 裕希(Kashiwagi, Hiroki) 佐々木, 史織(Sasaki, Shiori) 清木, 康(Kiyoki, Yasushi)
Publisher	慶應義塾大学湘南藤沢学会
Publication year	2013
Jtitle	交通運輸情報プロジェクトレビュー No.22 (2013.) ,p.40- 47
JaLC DOI	
Abstract	本稿では, ユーザの物理的状況や嗜好に適したイベント情報を抽出•配信するシステムを提案する。 本方式は, ユーザが過去に参加したイベント情報や現在の時間や場所といった状況をユーザ•コンテクストとして用い, 多次元空間からの統合的分析により各ユーザに合ったイベントを抽出する。 また抽出されたイベント情報を最適に配信する方法についても議論する。本方式により, 私たちが有意義なイベント選択を行うことができ, 新たな発見をすることができる情報環境の実現を目指す。
Notes	2013年度慶應義塾大学JR東日本寄附講座報告書 慶應義塾大学交通運輸情報プロジェクト その3 : 慶應SFC院生•学部生の研究
Genre	Technical Report
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO92001006-00000022-0040

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

多次元空間計量によるユーザ・コンテクスト対応 イベント情報抽出・配信システムの構成

柏木 裕希† 佐々木 史織‡ 清木 康†

†慶應義塾大学環境情報学部 〒252-0805 神奈川県藤沢市遠藤 5322
‡慶應義塾大学政策・メディア研究科 〒252-0805 神奈川県藤沢市遠藤 5322
E-mail: †‡{t10227hk, sashiori, kiyoki} @sfc.keio.ac.jp

あらまし 本稿では，ユーザの物理的状況や嗜好に適したイベント情報を抽出・配信するシステムを提案する．本方式は，ユーザが過去に参加したイベント情報や現在の時間や場所といった状況をユーザ・コンテクストとして用い，多次元空間からの統合的分析により各ユーザに合ったイベントを抽出する．また抽出されたイベント情報を最適に配信する方法についても議論する．本方式により，私たちが有意義なイベント選択を行うことができ，新たな発見をすることができる情報環境の実現を目指す．

キーワード 多次元空間計量，コンテクスト分析，イベント情報

1. はじめに

インターネットが発達しネット空間がなくてはならないものとなった近頃，リアル世界で行われるイベントに改めて注目が向けられている．イベントとは本稿においては，スポーツの試合や音楽コンサート・ライブ，ミュージカル等を指す．人々は，さまざまな感覚的もしくは物理的な要因によってイベントに参加する．その要因はイベントに参加する人それぞれであり，どのイベントに参加する傾向があるのかは一概に定型化しづらいものである．

日本では週末にもなると，各地であらゆる種類のイベントが開催されている．私たちは注目しているイベントに関しては事前にそのイベントに行くことを決めてスケジュールに組み込み，それに合わせた行動をとる．しかしながら，週末にたまたま時間が空いた場合，その時間を使って近くのイベントに参加することはほとんどないだろう．それでは，なぜイベントに参加しないのだろうか．それは，その時リアルタイムに近くでどのようなイベントがやっているかという情報を入手することが困難であるからだ．イベント情報を入手するためには，アーティストやチームのページにアクセスしたり，会場に問い合わせたりという手段が一般的となっている．

そこで，本研究においては，上記の課題を解決するために，イベントが持っている多次元空間から，ユーザの物理的状況と嗜好に合ったイベントを効果的に抽出しその情報を伝えるシステムを提案する．

本研究の目的は下記 2 点が挙げられる．

(1) あらゆる事象に適したイベントを抽出するシステムの提案

(2) 抽出結果をリアルタイムな情報として配信

これらの目的が満たされることで，私たちは日々の生活においてより効率的な行動をすることができると同時に，今まで機会のなかった情報に触れることができ，さらにはそれらを通じて新たな発見が生まれ，人々の行動の活性化を引き起こすことが可能となる．

2. イベント情報抽出システムの手法

2.1 システム基本構造

本システムは，ユーザの物理的状況からの情報の絞り込み，ユーザやイベントの特性からの情報抽出，そして抽出された情報のユーザへの提供の 3 つのサブシステムから構成される．(図 1)

まず，(1)当時点での時間以降のイベントのみを対象とし，(2)ユーザの場所と会場の場所を比べて物理的に行くことが可能なイベントのみに絞る．そして，(3)それらのイベントの中からユーザの嗜好に適したイベントを抽出する．分析方法はイベント類似度の分析とユーザ類似度の分析にそれぞれ分かれ，複合的に分析を行いユーザに適したイベントを抽出する．(4)以上のプロセスにおいて抽出されたデータをマトリックスとして表すとともに，マップ上に抽出されたイベントの会場の場所を表示する．

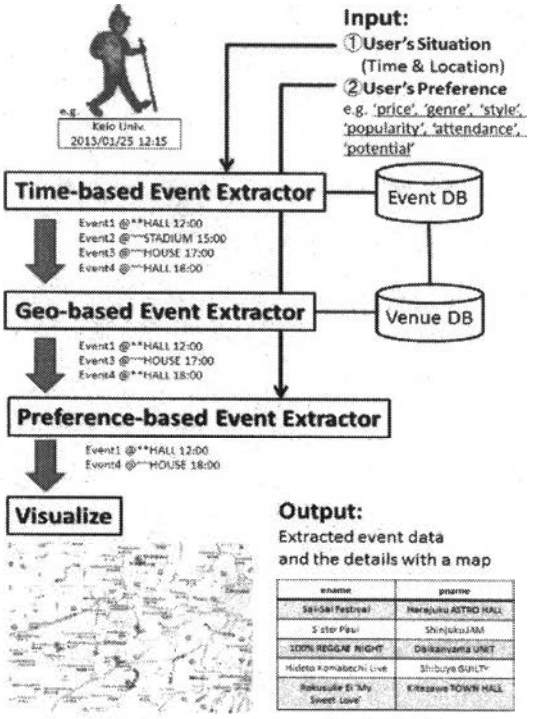

図 1 システム基本構造

2.2 本システムの特徴

本システムはユーザの物理的状況や求めている嗜好的条件から，各ユーザに適したイベントを抽出するものである．また，それらの抽出されたデータをリアルタイムにユーザに対して視覚的に情報を提供するものである．

本システムの特徴は以下の 3 点である．

(1) 対象データが動的で属性が多様であり，かつアーティスティックな対象

(2) リアルタイムに情報提供を行う

(3) 多次元データを複合してダイナミック情報として扱う

以下にて，これら 3 つの特徴について詳細に述べる．

2.2.1 対象データが動的で属性が多様であり，かつアーティスティックな対象

情報抽出・推薦の研究成果[1][2][3][4]は多く世間に公表されており，また実際のマーケティングに役立っている例も少なくない．これらのデータは協調型フィルタリングを主としたレコメンドアルゴリズムによりある種型にはまった形で行われ，それがそのまま良質な結果として反映することができる．

一方で，本システムで扱うイベント情報は必ずしも定型に沿った形で情報抽出をすることのできない対象である．ジャンルや出演者等の属性に加え，それぞれのイベントでユニークである開催日時や会場，そして特徴を示すタグ等動的でかつ多様な属性が存在する．また，人がイベントに参加するという選択は極めて感覚的な要素が強く，一概にどのイベントに行くのかというのは定めることが難しい．こうした，動的で属性が多様で，感覚的要素の強いアーティスティックなデータを本システムでは対象とする．

2.2.2 リアルタイムに情報提供を行う

イベント情報が蓄積されているサイトはある分野に絞ったものや特定の地域に絞ったもの，できるだけ多くのイベントを網羅するものまで様々なものが存在している[7][8]．それらのサイトは日付や場所ごとに分かりやすく表示されているものも多くイベント情報を調べることには適しているが，ユーザの状況や嗜好を分析してイベントを提案してくれるものとは言い難い．

本システムでは現在の時間・場所を基にした「今から行けるイベント」を抽出することができるので，人々が空いた時間でより流動的に，新しい発見ができるような機会を提供することができる．

2.2.3 多次元データを複合してダイナミック情報として扱う

本システムでは対象データとして大きく分けて 2 つ，イベントの情報とユーザの状況を扱う．なぜならイベントに参加するためにはこれら 2 つの要素が交互に絡み合ったうえで最終的に決断がなされるからである．

本システムは刻一刻と移り変わる時間や場所等のユーザコンテクストとイベントの特徴データを複合的に融合させることで，ユーザに対して最適なイベントを提案する．

また，ユーザが自分自身の好みによりダイナミックに条件を操作ができる手法の提案をする．

2.3 基本方式

2.3.1 実現プロセス

本システムはユーザの位置や時間という物理的情報と嗜好情報等の多次元空間から，ユーザに合ったイベントを抽出し提案するものであるが，はじめにその実現プロセスについて示す．本システムは以下の手順で実現する．

STEP 1: 対象データの決定

STEP 2: 対象データの整形

STEP 3: 時間と地理的空間情報からのイベントの抽出

STEP 4: 状況データと特徴データを組み合わせたイベントの抽出

STEP 5: 抽出イベントの視覚化

以下の節では，それぞれの STEP3-5 で行う詳細な計算方式について述べる．

2.3.2 時間計算基本方式

本システムはユーザが将来に参加するイベントを推薦するものであるので，過去のイベントは抽出結果の対象外とする．

時間計算プロセスにおいては，イベントの時間情報については開演時間を基に計算をする．それは開演時間や終演時間が定義されていないデータがあることと，多くのイベントにおいて開演時間に合わせて来客する人が多いことが理由である，

タイムスタンプ型のデータで“現在(ユーザが設定した)時刻≦開演時刻”の計算式を用いた，これに適さないイベントは抽出イベントとしてすべて除外する．

2.3.3 地理座標計算基本方式

本システムにおいては，以下の計算方式を用いて二点間の距離を計算する．

$$\text{距離}distance = \sqrt{\left(\frac{\text{現在地緯度} - \text{対象地緯度}}{0.0111}\right)^2 + \left(\frac{\text{現在地経度} - \text{対象地経度}}{0.0091}\right)^2}$$

0.0111 と 0.0091 については，日本においては緯度 1 度当たり平均 111km，経度 1 度当たり平均 91km によるもので，本計算式にあてはめると km 単位として二点間の距離を算出することができる．

本計算式を SQL 関数として定義し，原則として現在地(もしくは設定した場所)とイベントが行われる会場間の距離を計算する．

2.3.4 イベント類似度計量方式

特徴データに関する計量方式は 2 種類あり，それはイベント類似度計量方式と，ユーザ類似度計量方式である．

イベント類似度計量方式は，例えば何かのイベントを問い合わせとして発行した時に，その問い合わせ条件のイベントに近いイベントを抽出する．近いイベン

トとはジャンルやタグが一緒もしくは相関の強いジャンルであったり，イベントに参加している人の相関性の強いもの，さらには会場の収容人数や場所から判断したりする等の方法がある．

2.3.5 ユーザ類似度計量方式

ユーザ類似度計量方式は，ユーザ同士の特徴の近さを判断するものである．対象となったユーザに対して参加しているイベントが近いユーザを抽出し，その抽出されたユーザが参加したことのあるイベントから対象となったユーザに対してイベントを提案する．ユーザ相関については，それぞれのユーザのオーダデータから判断し，イベント類似度計量方式も合わせて，総合的に判断する．

2.3.6 統合計量方式

統合計量方式は上で述べた 4 つの方式を統合しイベントを抽出するものである．時間と地理座標を複合してこれから行くことのできるイベントを絞り，その絞ったイベントの中からイベント類似度計量方式またはユーザ類似度計量方式またはその両方を用いて，ユーザの嗜好に合ったイベントを総合的に抽出する．それぞれの計量方式においては都度，重要視する条件を変更することによって，よりユーザに合ったイベントを抽出することができる．統合計量方式は本システムのイベント抽出プロセスにおける最終段階である．

2.3.7 抽出結果出力方式

上で述べたイベントの抽出方式において抽出されたイベントは，SQL 上でのものであり，一般ユーザが容易に用いることは困難である．

抽出結果出力方式においてはそれらの抽出されたイベントをウェブブラウザ上で視覚的に提案するものある．本方式においては Google Maps API[10]を用いて，マップ上に現在地(もしくは設定した場所)とともに，抽出された複数のイベント情報をその会場の座標においたマーカとともにイベント情報を配置する．Google Maps API のコードは JavaScript として設定し，ウェブブラウザに表示する形は HTML，SQL との連携は PHP でそれぞれ行う．

3. 実現方式

3.1 基本データ構造

本システムにて使用するデータは，プレイガイド「e+(イープラス)[6]」の運営会社である株式会社エンタテインメントプラスから提供を受けたものである．主に用いるデータテーブルとしてイベントデータ，会場データ，そしてユーザのイベントチケットオーダデータがある．また，これら 3 つに付随した形でイベントのタグ情報とジャンル情報，会場のエリア区分情報と都道府県区分情報，オーダのステータス区分情報を用いる(図 2)．すべてのデータはオープンソースのオブジェクト関係データベース管理システムである PostgreSQL[9]にデータベースとして格納しており，SQL に対して問い合わせを発行することでデータを抽出する．以下，それぞれのデータベース定義を述べる．

※本稿においてはユーザ ID やイベント名称は権利上，加工したうえで提示する．また，その他一部のデータについても加工・修正した後のデータを提示しているが，分析に使用するデータはすべて生のデータである．

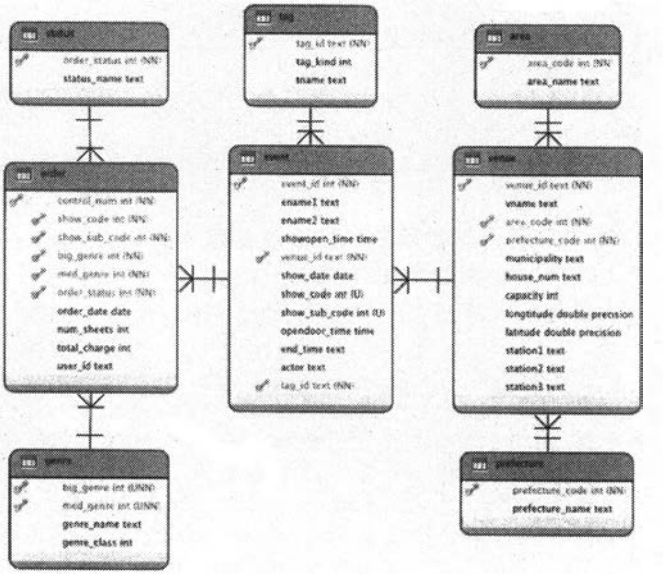

図 2 基本データ構造(システム ER 図)

3.1.1 イベントテーブル

イベントテーブルの構造とデータ例は表 1 の通りである．イベント数は計 177,729 データが本テーブルに存在する．本テーブルにはイベント名称の他，そのイベントの開場・開演・終演時刻や出演者情報等が含まれている．

属性 tag_id は 1 つのレコード内それぞれにタグ ID(数量制限なし)が配列として格納されている．タグ数は計 36,936 データ存在する．

表 1 イベントテーブルの構造

属性名	属性の意味	データ型
event_id	イベントの ID	Integer
ename1	イベント名称 1	Text
ename2	イベント名称 2	Text
showopen_time	開演時間	Time
venue_id	会場の ID	Text
show_date	開演日	Date
show_code	公演コード	Integer
show_sub_code	公演サブコード	Integer
opendoor_time	開場時間	Time
end_time	終演時間	Text
actor	出演者	Text
tag_id	タグ	Text[]

3.1.2 会場テーブル

会場(venue)テーブルではイベントが行われる表 2 に示す通り各会場の情報を格納している．会場とはライ

ブハウスや競技場，劇場等のイベントが行われる施設や場所を指し，日本全国各地計 10,808 のデータが存在する．

属性 area_code では全国を 10 区分した地域をセットし，属性 prefecture_code では 47 都道府県をセットしている．

表 2 会場テーブルの構造

属性名	属性の意味	データ型
venue_id	会場の ID	Text
vname	会場名称	Text
area_code	エリアコード	Integer
prefecture_code	都道府県コード	Integer
municipality	市区町村名	Text
house_num	番地その他	Text
capacity	収容人数	Integer
longitude	経度	Double precision
latitude	緯度	Double precision
station1	最寄り駅 1	Text
station2	最寄り駅 2	Text
station3	最寄り駅 3	text

3.1.3 オーダテーブル

本システムにおいて最も重要な対象データの一つがオーダ(order)テーブルである．オーダとは「e+」において実際にチケットが注文されたものである．order テーブルのデータ例を表 3 に示す．本データは主に年間一定回数以上オーダのあるユーザのオーダを対象とし，期間は 2010-08-01 ～ 2013-07-31 の 3 年間で，テーブルに格納しているオーダ数は約 400 万である．

本テーブルでは属性 show_code と属性 show_sub_code によって event テーブルとリレーションすることができる．

属性 big_genre と属性 med_genre ではそれぞれのイベントにユニークに付与されているイベントのジャンルを表す．ジャンルは大ジャンルとそれに連なる中ジャンルを合わせた形で計 119 存在する．属性 order_status においては status テーブルで定義するオーダの状況を設定している．本オーダテーブルにおいては予め status で発券もしくは発券状態まで至っているオーダのみを絞って格納している．

表 3 order テーブルデータ例

show_code	show_sub_code	name1	name2	big_genre	med_genre
34483	3	柔道大会 A	2010	400	404
51833	51	球団 A 主催試合	在京球団 A ×在阪球団 A	400	401
3977	115	50 代男性ソロアーティスト A	NULL	100	101

order_status	order_date	control_num	num_sheets	total_charge	user_id
36	2010-MM-DD	562051	1	2700	aaa
36	2011-MM-DD	581622	1	1900	bbb
36	2012-MM-DD	589727	1	8000	ccc

計 8 つのテーブルをリレーションすることができ，複数の条件を加えて問い合わせを発行することが可能である．これらのテーブルとダイナミックに移り変わるユーザの情報をそれぞれリレーションさせ複合させた結果がイベント情報抽出システムとして成立することになる．

4. 実験結果とその検証

4.1 実験概要

実験の最終目的はユーザに最も合ったイベントを抽出するための手法を見出すことである．本実験においてはそれぞれの実現方式について実際にデータを用いて実証を行い，その実現性を検証する．本稿においては(1) 時間と地理座標の関係性，(2)イベント類似度計量実験，(3)ユーザ類似度計量実験，(4) 統合システム実験の 4 パターンの実験検証を行う．以下にてそれぞれの実験をおこなうが，4 パターンの実験検証の中でも異なった条件を用いた実験を行い実現性の検証を行う．

4.2 実験 1: 時間と地理座標の関係性

最初の実験はユーザの時間・座標とイベントの行われる開演時間・会場座標から，ユーザが物理的に行けることが可能なイベントを抽出するものである．本システムにおいては対象データがすべて過去のものであり，リアルタイム時刻でユーザの状況を設定するのが難しいので，以下を仮想ユーザの状況として設定し，このデータを用いて実験を行う．

User: Kashiwagi
Date: 2012-12-23
Time: 14:50:00
Location: 35.690921,139.700257(新宿駅)

4.2.1 イベントの開演時間・座標からのイベント抽出

現在のユーザの座標とイベントの開演時間を考慮に含めてイベント抽出を行う．ここでは，ユーザは平均 20km/h(0.333km/min)で移動できるものとして計算を行い，本実験では現在地から 15 分以内に行けるものを対象とする．サンプルユーザの状況とイベントの開演時間・座標からのイベント抽出結果が表 4 の通りである．

4.3 実験 2: イベント類似度計量実験

本実験においてはイベントというコンテンツをベースとした，その類似度を計量し，ユーザに対して適したイベントを抽出する方式を検証する．ここではユーザに視点をおかないイベントのみを対象とした実験であるので，あるイベントをクエリーとして，そのイベントに近いイベントを抽出する．本実験においてはイベントの近さを測る検証として(1)ジャンル，(2)イベント参加者，(3)複合的な条件からの 3 つのパターンを用いる．

表 4 サンプルユーザの状況とイベントの開演時間・座標からのイベント抽出結果

ename1	ename2	show_date	open_time	vname	distance
在京交響楽団A	特別演奏会A	2012-12-23	15:00:00	Bunkamura オーチャードホール	2.74
在京バレエ団A	2012 Xmas 公演	2012-12-23	15:30:00	ゆうぽうとホール	6.61
ビジュアル系バンドA	NULL	2012-12-23	15:30:00	渋谷 REX	3.03
ビジュアル系バンドB	ビジュアル系フェス	2012-12-23	15:30:00	高田馬場 AREA	2.05
海外男性アーティストA	NULL	2012-12-23	16:00:00	パシフィコ横浜 国立大ホール	22.0

4.3.1 ジャンルを用いたイベント類似度計量

イベント類似度を測る最初の実験はジャンルを用いたものである．ジャンルはそれぞれのイベントに付与されているものであるが，ここでは中ジャンルで2つ以上のイベントに参加している人を抽出し，それらのジャンルのみを配列として持つ．そしてその配列に対して統計解析向けのプログラミング言語R言語においてアプリオリアルゴリズム[11]を実行することで，それぞれのジャンル間の相関の近さを抽出する[5]．以下に，その例を示す．

①ジャンル"野球"に参加したことのあるユーザ

表 5 "野球(401)"とのジャンル相関結果

	lhs rhs	support	confidence
1	{401} => {101}	0.03088	0.3129
2	{401} => {102}	0.01446	0.1465
3	{401} => {402}	0.01212	0.1228
4	{401} => {301}	0.01190	0.1206
5	{401} => {128}	0.00881	0.0893

表5は中ジャンル"野球"の野球ジャンルを対象とし Confidence 値の昇順に並べたものである．lhs は対象のデータとなる条件部を表し，rhs は抽出データとなる結論部を表す．Confidence 値とは信頼度とも呼びルール間の結びつきの強さを表し，Support 値とは支持度とも呼びルール同士の相関を表す．ここでは最低限の信頼性を担保するために最小 Support 値を 0.01 と指定した．こうしてみると，"野球"ジャンルに近いジャンルは 101 のジャンルである"J-POP／邦楽ポピュラー"であることが分かる．

②ジャンル"野球"，"J-POP／邦楽ポピュラー"の両方に参加したことのあるユーザ

表 6 "野球(401)"&"J-POP／邦楽ポピュラー(101)"とのジャンル相関結果

	lhs rhs	support	confidence
1	{101,401} => {102}	0.00642	0.2079
2	{101,401} => {128}	0.00335	0.1087
3	{101,401} => {301}	0.00297	0.0962
4	{101,401} => {125}	0.00197	0.0640
5	{101,401} => {402}	0.00196	0.0635

表 6 では，条件をジャンル"野球"，"J-POP／邦楽ポピュラー"のイベント2つに参加したことのあるユーザが他にどのようなジャンルのイベントに行っているかを表している．ここでは，102 のジャンル"邦楽ロック"や 128 のジャンル"フェスティバル"が Support 値の上位となっている．ある程度のサンプル数を確保したうえで複数のジャンル条件を指定することができると，より類似度の信頼性も深くなることを期待することができる．

4.3.2 イベントへの参加者からのイベント類似度計量

本実験では，各ユーザのオーダデータを用いて，単純にイベント自体の相関を分析する．所謂「このイベントに行ったことのある人はこのイベントに行っています」というものを抽出するものである．

ここでは属性 show_code が男性アカペラグループ A を指す 1163 を対象として，男性アカペラグループ A のイベントに参加している人が，他にどのようなイベントに参加しているのかを抽出する．その結果が表 7 の通りである．

表 7 show_code"1163"とのイベント類似度計量

	lhs rhs	support	confidence
1	{1163} => {19270}	0.15674603	0.16934620
2	{1163} => {3977}	0.06646825	0.07181136
3	{1163} => {376}	0.06349206	0.06859593
4	{1163} => {3629}	0.05753968	0.06216506
5	{1163} => {12570}	0.05158730	0.05573419

本実験は 1163 が存在するトランザクションのみを対象としてアプリオリアルゴリズムを適用している．抽出された結果のそれぞれの名称は1番目から"音楽フェス A"，"50代男性ソロアーティスト A"，"50代男性ソロアーティスト B"，"40代男性ソロアーティスト A"，"音楽フェス B"の順であり，男性アカペラグループ A とそれぞれにおいて相関が強いことが分かる．

これより，男性アカペラグループ A (1163)に参加しているユーザに対しては将来に行われる上のイベントを抽出すると適している可能性が高いことが期待される．

4.3.3 複合的要素からのイベント類似度計量

ここでは要素を複合的に計量することでより適切で詳細条件からのイベント類似度を抽出する実験を行う．今回サンプルイベントとして以下のデータを用いて，サンプルイベントのジャンルである"野球"に相関の強いジャンル"J-POP／邦楽ポピュラー"を対象とし，"J-POP／邦楽ポピュラー"の中でサンプルイベントに相関の強いイベントを抽出する．

show_code: 16107
ename1: 在京球団B
ename2: 【公式戦】在京球団B×在京球団A
big_genre: 400(スポーツ)
med_genre: 401(野球)

表 8 “J-POP／邦楽ポピュラー(101)”の中でサンプルイベント(16107)の相関結果

	lhs rhs	support	confidence	抽出イベント
1	{16107} =>{12987}	0.07461810	0.07461810	ポップスグループA
2	{16107} => {376}	0.05405405	0.05405405	50代男性ソロアーティストB
3	{16107} =>{12263}	0.04054054	0.04054054	40代男性デュオ
4	{16107} => 15474}	0.02820212	0.02820212	30代女性ソロアーティスト
5	{16107} => {8764}	0.02643948	0.02643948	音楽フェスG

サンプルベントに参加している人の中で“J-POP／邦楽ポピュラー”のイベントにも同時に参加している人を対象としたデータ数は1702であった.表8の抽出結果を見ると，サンプル数が少ないために，Support値とConfidence値が同値になってしまっているが，少なからずとも相関の強いイベントがあることが分かる.

4.4 実験 3: ユーザ類似度計量実験

4.4.1 実験概要とサンプルユーザの設定

イベント類似度は多くの人が参加しているイベントを抽出する相関分析を用いるので，動員数が多い人気のイベントが多く抽出され，なかなか新しい発見が得られるという目的の達成まで持ち込むことは難しい.そこで，対象ユーザに近いユーザを新たな対象ユーザとすることで，そのユーザが参加しているイベントの特徴から，今後参加する可能性の高いイベントを抽出し，その抽出されたイベントを元の対象ユーザに対して提案するという手法を取る.

最初に，ユーザ類似度計量実験を行うため，サンプルユーザをその参加したイベントとともに設定する.なお，参加したイベントは実際のデータに即したものを一部修正したものである.

4.4.2 ジャンルからのユーザ類似度計量

ジャンルからのユーザ類似度では同じジャンルのイベントに参加している回数が多いユーザを対象ユーザに近いユーザとみなして，その抽出されたユーザが参加しているイベントの傾向を結果として抽出する.

サンプルユーザが参加しているイベントのジャンルは“125,304,401*2,402*2”であり，これらのジャンルのイベントにより多く参加しているユーザを抽出する.表9は4つのジャンルすべてに参加したことのあるユーザの内訳である.

表 9 “125,304,401,402”すべてのジャンルのイベントに参加したことのあるユーザの参加イベントジャンルの内訳

user_id	med_genre
GUser1	101*2,125,201,304,401,402*2 (計 8)
GUser2	125*2,201*8,204*2,205,304*6,311*9,401*13,402*12,408*2,411,507*2 (計 58)
GUser3	101*3,125*8,127*5,203,301*4,304*2,316*11,401,402*2 (計 37)

サンプルユーザが参加しているイベントのジャンルすべてに適合するユーザは3人存在した.この中からさらに近いユーザを測るために以下の計算式を対象ユーザが保有しているそれぞれのジャンルに適用し，その和をもってユーザの相関の強さを測る.

$$\frac{\text{対象ユーザの同ジャンル数} \times \text{対象ユーザに一致する同ジャンル数}}{\text{抽出ユーザの参加イベント数}}$$

表 10 ユーザ相関を測るための計算結果

	125	304	401*2	402*2	
1	1	1	2	4	8/8 = 1.00
2	2	6	26	24	58/58 = 1.00
3	3	2	2	4	11/37 = 0.30

表10の計算結果よりGUser1とGUser2が同順で並んだが，総イベント参加数の少なさから，ここではGUser1がサンプルユーザに最も近いユーザであると判断し，ここでイベント類似度計量方式等を用いながら，最も近いユーザの嗜好に合ったイベントを抽出することで，サンプルユーザにイベントを提案する.

4.4.3 参加イベントからのユーザ類似度計量

対象ユーザが参加したイベントと同じもしくは同等のイベントに参加したことのあるユーザを見つけ出し，より多くの同じイベントに参加しているユーザを最も近いユーザとみなす実験が参加イベントからのユーザ類似度計量実験である.

サンプルユーザは“23,133,6724,6733,16107,70211”のコードが振られているイベントにそれぞれ参加しており，これらと同じコードのイベントに参加しているユーザを抽出する.しかしながら，イベントの開催日を考慮しないで抽出するとユーザの現在の嗜好にそぐわない可能性が出てきてしまうので，最新のイベントから一致するユーザを絞り込んでいく形でイベントを抽出する.まず，直近で参加している2つのイベント6724と70211に参加しているユーザはサンプルユーザを含めないと計6名いる.この中から同じ要領で直近のイベントから参加しているものが同じユーザを絞っていくと，表11の通り“70211,6724,6733,16107”のイベントに同じく参加している3名に絞られる.同じイベントに参加しているユーザをこれ以上絞ることはできないので，サンプルユーザに最も近いユーザは，抽出ユーザの中で最も参加しているイベント数が少ないSUser2とする.

表 11 show_code “70211,6724,6733,16107”のすべてのイベントに参加しているユーザとその内訳

user	show_code
SUser2	70211,19949,16107,42196,6724,6733,6633,1060
SUser3	40515,93542,51833,70211,14896,70209,90092,91583,15084,19949,66722,80882,14917,12302,16107,6722,88563,6739,51431,6724,6752,6714,6844,6733,6750
SUser6	14896,51833,70211,90092,16246,12620,16107,51431,6724,6718,6739,6750,6733

4.5 実験 4: 統合システム実験

統合システム実験として，以下の物理的状況，ユーザ類似度計量実験で用いたサンプルユーザを仮想ユーザとして，これから参加することのできるイベントを抽出する．

Date: 2013-08-03
Time: 14:00:00
Location:35.690921,139.700257(新宿駅)

まず，前節のユーザ類似度実験によってサンプルユーザに近いユーザは表 11 における SUser2 であると判断した．SUser2 が参加しているイベントの詳細な内訳は表 12 の通りである．

表 12 SUser2 が参加しているイベント内訳

show code	ename1	ename2	show date	medgenre
6633	２０１２ 米国プロ野球リーグ開幕戦	プレシーズンゲーム 米国球団 A×在阪球団 A	2012-03-25	401
6724	在京サッカーチーム B ホームゲーム	在京サッカーチーム B×在京サッカーチーム C	2010-09-18	402
6733	在京サッカーチーム A ホームゲーム	在京サッカーチーム A×在京サッカーチーム C	2011-03-13	402
6844	在京サッカーチーム C ホームゲーム	在京サッカーチーム C×在阪サッカーチーム A	2013-07-27	402
16107	在京球団 B'主催試合	【公式戦】在京球団 B'×在阪球団 A	2011-05-07	401
19949	全日本サッカー大会	在京サッカーチーム C×在京サッカーチーム B	2011-12-17	402

ここからどのように分析してイベントを抽出するかは，各ユーザの判断に委ねることができるが，本実験においては最も近いユーザの特徴がジャンル“401”と“402”のイベントに頻繁に参加していることから，ジャンル相関からのイベント類似度計量を行い，サンプルユーザが行くことのできるイベントを抽出する．結果として表 13 の値が抽出され，ジャンル“J-POP／邦楽ポピュラー(101)”との相関が強いことが分かる．

表 13 “野球(401)” & “サッカー(402)” とのジャンル相関結果

	lhs rhs	support	confidence	抽出ジャンル名称
1	{401,402} => {101}	0.0019626199	0.16192412	J-POP／邦楽ポピュラー
2	{401,402} => {128}	0.0007636973	0.06300813	フェスティバル
3	{401,402} => {102}	0.0007554855	0.06233062	邦楽ロック
4	{401,402} => {301}	0.0005748259	0.04742547	演劇

そこで，サンプルユーザの状況と組み合わせて，「今から行くことのできるジャンル“J-POP／邦楽ポピュラー”のイベント」としてイベントを抽出することができる．表 14 はその抽出イベントの開催時間の早い順に並べた上位 5 イベントであり，図 3 は同上位 10 のイベントを抽出結果出力方式によりマップ上にマーカとして配置したものである．

表 14 サンプルユーザが行くことのできるジャンル“J-POP／邦楽ポピュラー(101)”のイベント

ename1	date	open time	vname	distance	latitude	longitude
20代男性アーティスト	2013-08-03	15:00:00	Zepp Tokyo	10.748	35.625961	139.782298
ガールズフェス	2013-08-03	16:00:00	shimokitazawaGARDEN	4.447	35.660448	139.66842
男性ボーカルユニット	2013-08-03	16:30:00	新宿 Glamstein	1.297	35.701806	139.707987
男性二人組音楽ユニット	2013-08-03	17:00:00	江ノ島虎丸座	41.864	35.307618	139.484871
音楽フェスG	2013-08-03	17:00:00	なかの ZERO 大ホール	3.371	35.704299	139.671599

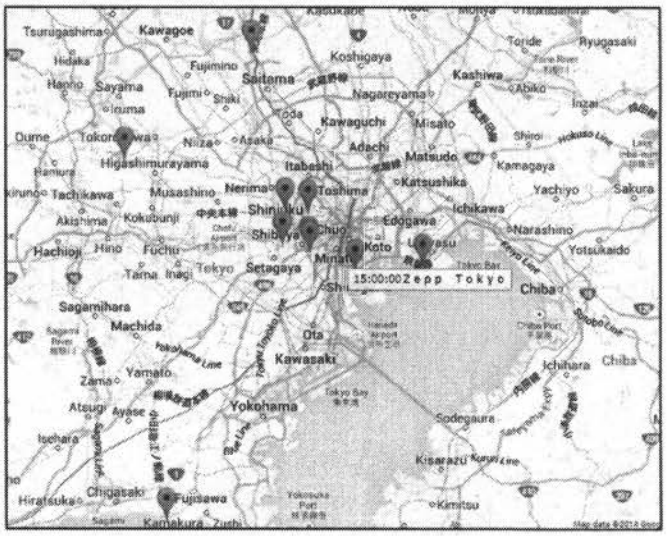

図 3 サンプルユーザが行くことのできるジャンル“J-POP／邦楽ポピュラー(101)”のイベントの視覚結果

5. 考察

ユーザの物理的状況からのイベント抽出をする実験，イベントの類似度を計量する実験，ユーザの類似度を計量する実験，そしてこれらを複合してイベントを抽出する実験という 4 つの方式で実験を行った．

物理的状況からのイベント抽出では，実際にユーザがこれから行けるものと考えることができるイベント

を抽出することができ，これはユーザのリアルタイムの行動に影響を及ぼすものと考えられる．本システムにおいてはユーザの移動速度が平均 20km/h と設定をしたが，この設定に関しては徒歩での移動であったり，列車の時刻表を考慮したりする等，さまざまな応用設定をすることができるだろう．

イベントの類似度計量実験ではジャンル，イベント参加者，それらを複合的に扱った実験を行った．

ジャンル相関の実験に関しては，1 つのジャンル，2 つのジャンルで実験を行ったが，少なからず相関が強いジャンルが抽出されることが分かる．また，対象ジャンルを複数にした分析はサンプル数が小さくなり，希少な結果抽出に繋がる可能性があるが，一方で必ずしも適した相関結果が得られるとは限らないだろう．

イベント参加者のオーダデータを用いた実験においては，設定したイベントと並んで参加したことのあるイベントを抽出した．ここでは，2 つ同時に参加している率が 5%以上の結果が 5 つ発見され，これは非常に相関の強い結果であると考えることができる．複合的要素からの実験ではサンプルイベントを設定し，そのサンプルイベントのジャンルとイベント参加者を複合した要素から類似度を計量した．ここではより綿密な条件設定を行うことができることを証明できた．

ユーザの類似度計量実験では，サンプルユーザがよく参加しているイベントのジャンル，そして参加イベントそのものから近いユーザを抽出する実験を行った．

同じジャンルやイベントに参加しているユーザを見つけ出すことができるのは非常に有意義であると考えられるが，近いユーザとして抽出されたユーザはあるイベントに偏った抽出結果となる例もあり，必ずしもユーザに合ったイベントの推薦に適するユーザが抽出されるとは限らないことが分かった．この部分は，ユーザの近さを判断する指標について再考の余地がありそうだ．

統合システム実験においては，ユーザ類似度計量，抽出されたユーザが参加しているイベントからのイベント類似度計量，そして物理的状況も加味したうえで応用となるアプリケーションシステムに最も近い形での実験を行った．

ユーザ類似度計量からイベント類似度計量，そして最終的な抽出結果の出力まで一連のプロセスを示すことができたが，抽出された結果がサンプルユーザにとって必ずしも適切なものであるかと言えばあまりそうとは考えられず，より一層の詳細な組み合わせのアルゴリズムが必要であろう．

イベントという対象は本稿で扱った属性以外にも様々な属性を付与することができ，更なる条件式，新たな角度からのイベント抽出も様々な方法で行うことができる．今回行った実験から更に発展した分析が必要となるだろう．

6. おわりに

本稿では，多次元空間のユーザ・コンテクストからユーザの物理的状況や嗜好に適したイベント情報を統合的に分析・抽出・配信するシステムを提案した．

本システムはユーザが自らに適したイベントを抽出することを可能にし，またその時の状況からこれから行くことのできる自分に合ったイベントを抽出することに有効である．その他，イベント運営側にとってもイベントのプロモーション方法や，ユーザのより活発的な流動を期待することができるので，本システムは有効である．

今後の展望としては，本稿においてはイベントの抽出をイベントの類似度とユーザの類似度のそれぞれからダイナミックな条件付与のもと行えることを示したが，これらの分析抽出配信のプロセスをユーザが自ら行うことができるアプリケーションの作成が不可欠である．本アプリケーションの作成のためには，どんな状況どんな嗜好条件のもとでも適切なイベント情報が抽出されるより精度の高いアルゴリズムを確立し，またユーザが簡単に自ら条件を指定できる環境も必要であろう．

参 考 文 献

[1] Oppermann, R., Specht, M., A Nomadic Information System for Adaptive Exhibition Guidance. Journal of Archives & Museum Informatics, Vol.13,No.2,pp.127-138, 1999

[2] Koyo Yoshida, Satoshi Endo, Naruaki Toma, Yuhei Akamine and Koji Yamada, *Event Recommend System based on Preference Extraction Using Tags*, IPSJ SIG Technical Report, Vol.2010-MPS-80 No.15, 2010

[3] C. Ono, M. Kurokawa, Y. Motomura, H. Asoh, *A Context-Aware Movie Preference Model Using a Bayesian Network for Recommendation and Promotion*, Proc. Of 11th Conf. of User Modeling (UM2007), 2007

[4] Koyo Yoshida, Satoshi Endo, Naruaki Toma, Yuhei Akamine and Koji Yamada, *A Proposal of Event Recommend Sytstem based on Similarity. Information Processing Society of Japan*, 2010

[5] 平山巧馬，小柳滋，“協調フィルタリングにおける相関係数法の予測性能向上，”電子情報通信学会論文誌，Vol.J90-D, No.2,pp.223-232, 2007.

[6] e＋(イープラス)，http://eplus.jp/

[7] 東京アートビート，http://www.tokyoartbeat.com/

[8] Events in America, http://www.eventsinamerica.com/

[9] PostgreSQL http://www.postgresql.org/

[10] Google Developers, Google Maps API, https://developers.google.com/maps/

[11] 柏木 裕希，清木 康，“ベースボールにおける投球パターン分析を対象としたデータ・マイニング方式とその投球予想への応用”，DEIM Forum 2012, March 2012.

[12] Kashiwagi, H., Sasaki, S., and Kiyoki, Y., "An Event Information Provision System with Multi-Dimensional Computations," Knowledge Creation & Intelligent Computing 2013, March 2013.