

CGM からのユーザ行動マイニングの提案

A Proposal for Human Activity Mining from CGM

川村 隆浩 山崎 智弘 長野 伸一 溝口 祐美子 飯田 貴之
 Takahiro Kawamura Yamazaki Tomohiro Shinich Nagano Yumiko Mizoguchi Iida Takayuki

(株) 東芝
 Toshiba Corporation.

Recently, context-dependent recommendation system for cellular phone has been publicly available, which presents any of useful information related to user's outdoor situation. In this paper, we propose an approach which first creates activity state maps based on the users' behavior log on CGM, then selects the information relevant to a specific topic by searching the states.

1. はじめに

近年、携帯電話を用いて屋外でネットにアクセスし、ユーザの状況に関連する情報を提示する試みが進められている。ここでは、携帯電話や RFID、各種センサーを活用し、イベントの記録を行動履歴として記録し、そこから何らかの規則性（頻出するイベント連鎖）を発見し、ユーザに関連する情報や次の行動を推薦する試みが行われている。

例えば、情報大航海プロジェクトの1つ、NTT ドコモの「マイ・ライフ・アシストサービス」や KDDI の「ケータイ版ライフログ」といったライフログ系の研究では、携帯電話や RFID、各種センサーを活用し、イベントの記録を行動履歴として、そこから何らかの規則性（頻出する連鎖）を発見する試みが行われている [1]。しかし、

1. イベントデータにはノイズが多く、意味のあるイベントの連鎖を見つけるのは難しい。
2. また、個人の履歴からだけでは、その人にとって意外性のある発見ができない。しかし、他人の履歴を用いるには処理データ量やセキュリティ、プライバシーなどに問題がある。
3. 一方で、規則性（推薦すべき行動）を手動でルール化するには、構築、保守に多大なコストが必要となる。

といった問題点が挙げられる。

そこで本論では、まずブログや SNS といった CGM (Consumer Generated Media) を対象に、特定のトピックキー（検索語）と行動オントロジーを参照して類似の行動を結合することで行動関連マップを作成する。そして、マップ上を探索することでユーザに関連情報や次の行動を推薦する手法を提案する。

2. CGM からの行動関連マップ作成手法

図 1 にシステムの概要を示す。今回提案のアプローチでは、まず、ブログや SNS といった CGM (Consumer Generated Media) を対象に、特定のトピックキー（検索語）と行動オントロジー（ユーザ行動の種類をグラフ上に体系化したもの、図 2 左参照）、商品オントロジー（商品のカテゴリーや名称などを体系

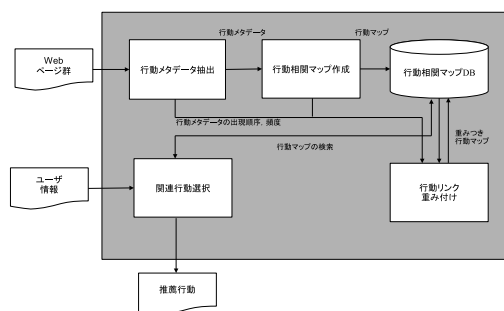


図 1: システム構成

化したもの、図 2 右参照) を用いてそれぞれの係り受け関係を抽出し、行動メタデータ (S, {O, C}, V からなるトリプル) を作成する。そして、それらの行動メタデータをオントロジーを参照して結合することで、行動関連マップを自動的に構築する。更に、CGM 上に書かれている行動の連続性と出現順序を考慮した共起関数を用いて関連行動間に重みを付けを行う。

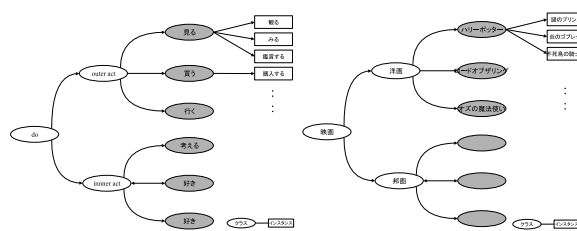


図 2: 行動オントロジー (左) と商品オントロジー (右)

以下に、トピック「ハリーポッター」に関して行動マップを作成する例を示す。まず、Web ページ群として、通常のキーワード型検索エンジンを用いて「ハリーポッター」という単語を含むページを収集する。そして、トピック「ハリーポッター」、および行動オントロジーに含まれるクラスやインスタンス（見る、借りるなど）や商品オントロジーに含まれるハリーポッターの副題などをトリガーとして、上記のページ内のセンテンスを対象に、係り受け関係のあるワードのリスト（行動メタデータ）を抽出する。抽出されたメタデータには、例えば以下のようなものが含まれる。

連絡先: 〒 105-8001 港区芝浦 1-1-1, e-mail: takahiron.kawamura@toshiba.co.jp

元の文章「ハリーポッターのDVDが発売日だったってこと今日なんか気が付いて、仕事終わってから、ツタヤとか行って不死鳥の騎士団借りてきた」

抽出されたメタデータ「ハリーポッター, DVD, ツタヤ, 借りてきた」

そして、オントロジーのクラスを参照し、各メタデータ内で同義のワードを1つのノードとして表し、ワード間の係り受け関係をノード間のリンクとして表すことで、行動関連マップを作成する。最後に、ワードの出現回数やワードの出現順序、オントロジー上でのワードの意味的な近さなどから行動関連マップ内のリンクに重み付けを行う。

尚、行動推薦にあたっては、マップ内のトピックの周辺ノードを辿ることでユーザに関連性のある行動を選択する(情報推薦を状態探索問題に置き換えているともいえる)。

3. 行動関連マップの構築実験

3.1 実験概要

図3に行動関連マップの出力例を示す。また、図4にトピック「ハリーポッター」に関して抽出した行動関連マップを抜き出したものを示す。

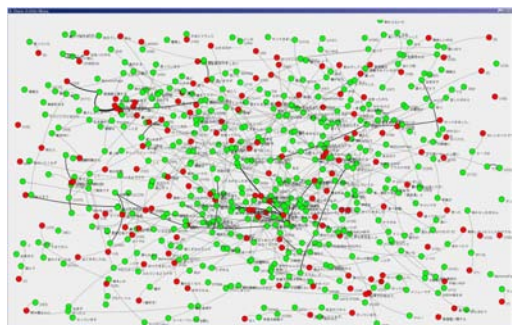


図 3: 行動関連マップの実際の出力例

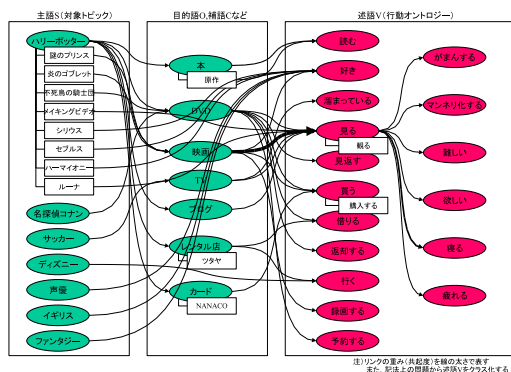


図 4: ハリーポッターに関する行動関連マップ

これは、07/11/22にGoogle Blog Searchで得られた「ハリーポッター」に関するブログ26件の内、149センテンスを対象に、CGMからの評判分析エンジン「ユビ de コミミハサンダー」[2]の一部機能を流用して行動メタデータを抽出したものである。尚、実験用に作成した行動オントロジーの規模は高々85ノードである。また、現在のところ重み付けは単純な総和となっている。

厳密な評価はこれからだが、マップを見る限り、

- ユーザが実際にしたいと思うであろう実用性の高い行動ルール(例えば、ハリーポッターの映画を観る, DVDを借りる, 原作を読む, など)や、
- 少し意外に感じる面白いと思える行動(例えば、見ると疲れる, 寝る(ハリポタ映画は尺が長いものが多いから?)や、声優に注目が集まっている(洋画にしては字幕より吹き替えが人気?), 名探偵コナンの話題が多い(視聴者, 読者層が被るから?)など)

がある程度、抽出できていることが確認できる。

3.2 考察

本アプローチでは、前述した以下の3つの問題点に対して以下のようなアプローチを行っている。

- CGMをベースとすることで、イベントデータに比べてノイズの少ないデータを抽出対象としている。
- 他人のブログなどから意外性のある行動を抽出できる可能性がある。
- 行動関連マップに関連行動が現れるため、ルールの手動作成が不要となる。

また、商品オントロジーや行動オントロジーを参照して、行動関連マップを意味クラスによってグループ化することで、対象トピックにどんぴしゃりな行動メタデータが存在していなくても、類似のトピックから何かしら推薦できるようにすることができる(マイナーな映画などについてはCGMから十分な行動メタデータが得られない問題への対応)。

更に、関連行動間に重みを付けを参照することで、役に立つありがたい行動と、?だが意外性のある行動を識別し、ユーザに実用性と面白さを提示することができることが期待される。

4. あとがき

今後は行動関連マップを用いて、ユーザのコンテキストやWebコンテンツ毎にマップを作成しておき、現在の行動やページビューから次の行動を推薦したり、Web閲覧中やTV視聴中に特定の単語が選択された際にユーザに関連のある行動を推薦するシステムを構築したい。

また、行動の連続性と出現順序を考慮した独自の共起関数の開発、および行動関連マップの辿り方(特定のトピックが与えられた際に推薦する行動の決定方法)についての検討を行う予定である。

参考文献

[1] レコメンドーションの虚実 ~ ソーシャルとライフログの交差点を目指す, <http://www.itmedia.co.jp/anchordesk/articles/0711/26/news026.html>, 2007.

[2] T. Kawamura, S. Nagano, M. Inaba, Y. Mizoguchi: Mobile Service for Reputation Extraction from Weblogs - Public Experiment and Evaluation, Proceedings of Twenty-Second Conference on Artificial Intelligence (AAAI-07), 2007.