

行動履歴を用いたデータマイニングによる人行動のモデル化

井口祐喜[†] 和田雄次[‡] 内藤和樹[‡]

東京電機大学情報環境学研究科[†] 東京電機大学情報環境学部情報環境学科[‡]

1. はじめに

近年、人の嗜好や周囲の環境などを考慮した情報配信サービスとして、コンテキストウェアサービスの研究が始まっている[1]。これには「人のコンテキストをいかに推定するか」と「コンテキストとコンテンツの関連付けをいかに行うか」という技術課題があり、本研究では人の行動履歴に対しデータマイニングを行い、相関ルールを抽出する方法でこの二つの問題に取り組んできた。また、サービス利用者の選択内容を使い推定結果の順位を変更する実験を行ったので報告する。

2. コンテキストウェアサービス

コンテキストとは人のニーズ、行っている行動などを指す。コンテキストウェアサービスとは利用者のコンテキストを様々な方法で推定し、それに適切なコンテンツを提供するサービスである。コンテキストウェアサービスの全体像を図1に示す。

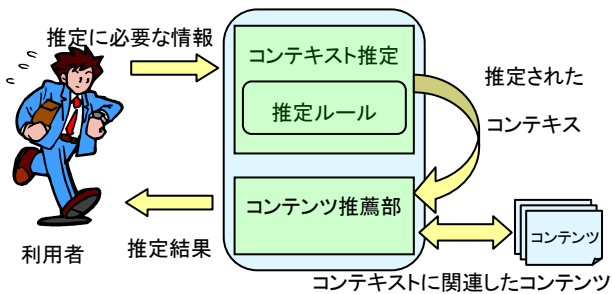


図1. コンテキストウェアサービスの例

3. 行動履歴を用いたデータマイニング

3.1. 提案手法

本稿では人の行動履歴から抽出した相関ルールを使用し、コンテキストの推定及びコンテンツの推定を行う。行動履歴データベースには曜日、時間帯、

Modeling Human Behavior by Data Mining Behavior Histories

[†] Graduate School of Information Environment, Tokyo Denki University

[‡] School of Information Environment, Tokyo Denki University

行動内容、行き先の情報が含まれる。これらの内、行動内容をコンテキストとして、行き先を推薦するコンテンツとして扱う。また、コンテキストの推定に曜日と時間帯を、コンテンツの推定に曜日、時間帯、そして行動内容を使用する。最終的な推定結果は確信度で順位を付ける。

3.2. データの収集

被験者2名の日常生活をGPSとヒアリングを行うことで収集した。収集は被験者Aについては28日間、Bについては73日間行った。その結果、それぞれ93件と258件のレコードを収集した。

3.3. 行動内容のカテゴリライズ

収集したデータを分析した結果、被験者等が行った行動には繰り返し行われた行動と一度しか行われなかった行動の2種類が存在することが分かった。そこで繰り返し行われ、関連性が高いと考えられるものでカテゴリを作成した。表1は被験者等がとった行動の回数を示す。

表1.被験者等が行った行動の一部(回数)

行動内容	被験者A	被験者B
帰宅	19	64
勉強	19	45
飲食	16	41
買物	13	58
立ち読み	0	4
ゲームセンター	0	2
忘れ物の回収	1	1

3.4. 行動履歴データベース

収集したGPSデータとヒアリング結果より被験者毎に行動履歴データベースを生成した。1レコードは被験者がある場所に行ったときの曜日、時間帯、行動内容と行き先の名称によって構成されている。

3.5. ルールの抽出

行動履歴データベースから相関ルールの抽出を行った。同じ行動を一ヶ月に2回以上行ったものをル

ールとして扱うため、データ収集日数の 7%を最低サポート値として使用した。ルールは 2 種類、コンテンツ（行動内容）を推定するルールと、コンテンツ（行き先）を推定するルールである。それぞれを表 2 と表 3 に示す。

表 2. コンテキスト推定ルールの一部

条件部	結論部	確信度(%)
夕, Fri	勉強	31.8
夕, Fri	買い物	27.3
夕, Fri	飲食	22.7

3. コンテンツ推定ルールの一部

条件部	結論部	確信度(%)
夕, Fri, 飲食	牧の原モア	60
夕, Fri, 飲食	アルカサル	20
夕, Fri, 飲食	夢庵千葉ニュータウン店	20

4. ルールの検証

抽出したルールの精度を検証した。各行動履歴データベースを 7:3 で分割し訓練データと検証データを作成した。本研究では利用者がサービスより提示されたコンテキスト、コンテンツ群より自分の行う行動を選択してもらう方法を想定している。このため、正しい推定結果が提示される上位 n 項目以内であれば十分である。表 4、5 にコンテキストとコンテンツがそれぞれ何位で推定できたかを示す。

表 4. コンテキスト推定ルールの精度

	第1候補	第3候補以内	第4候補以内	予測不能
被験者A	37%	69%	26%	5%
被験者B	32%	55%	33%	12%

表 5. コンテンツ推定ルールの精度

	第1候補	第3候補以内	第4候補以内	予測不能
被験者A	58%	58%	11%	32%
被験者B	53%	59%	26%	15%

コンテキストを第一候補で予測できたのは約 3 割であった。しかし第 3 候補以内で見るとコンテキスト、コンテンツともに約 6 割の精度で推定できている。また、コンテンツ推定精度はコンテキストに比べ低いことが分かる。これはコンテキストに比べコンテンツは種類が多いためだと考えられる。また、被験者等が訓練データで利用しなかったコンテンツを実験データでは利用していたため、予測不能の割合も増加していた。

5. 利用者のフィードバックの利用

抽出した相関ルールの精度を評価した際、コンテンツの推定はコンテンツの種類が多いため適切なコンテンツを上位にすることが難しいという問題があった。本稿ではサービス利用者の選択内容を使い推定結果の順位を変更する実験を行ったので報告する。

5.1. 実験

コンテキスト「飲食」について利用者の選択内容を使い推定結果の順位変更を行った。一般的に、一日に同じ料理を 2 度食べるという事はあまりない。そこでこの特性を利用し一度利用したコンテンツは同じ日の推薦内容から削除する、という方法を取る。このとき各飲食店に洋食、和食、中華料理などジャンルを付け、同じジャンルのコンテンツを推薦候補から削除した。これにより相対的にその他のジャンルのコンテンツの順位を上げることができる。

5.2. 結果

表 6 に順位変更を行った結果を示す。実験の結果、推定されたコンテンツの内 56%で順位の上昇が見られた。一方、31%の推定結果については順位の変化は起きず、13%は逆に適切なコンテンツが削除され予測不能となった。

表 6. 順位の変化

順位が上昇	変化なし	予測不能
56%	31%	13%

6. まとめ

本稿では人の行動履歴に対しデータマイニングを行うことで、コンテキストウェアサービスにおけるコンテキスト及びコンテンツの推定を行うルールの抽出を行った。また、ルールの評価を行うとともに利用者の選択内容を使い推定結果の順位変更実験について報告した。今後は実際のコンテキストウェアサービスの実装が課題となる。

この研究は東京電機大学先端工学研究所の重点課題研究費の援助を戴いて実施しました。

7. 参考文献

[1] 林智天、ほか “小型モバイルセンサを用いたコンテキスト適応型コンテンツ配信サービスの設計と実装” 電子情報通信学会技術研究報告 NS2004-226, March 2005