1L3-OS-17b-2

Contextに基づいたID付きPOSデータの分析方法

Context-Aware ID-POS Data Analyzing

吉田真[*1] 藤居誠[*2] 佐々木憲二[*2] 本村陽一[*3*1]
YOSHIDA Makoto FUJII Makoto SASAKI Kenji MOTOMURA Yoichi

[*1]東京工業大学 Tokyo Institute of Technology
[*2]東急エージェンシー Tokyu Agency
[*3]産業技術総合研究所サービス工学研究センター
National Institute of Advanced Industrial Science and Technology

At retail stores, such as a supermarket, there exists a context that is change of goods sold well with time. In order to analyze of such a context of the distribution in ID-POS data, we extracted the topic about the costomers' buying behavior using PLSA. The analysis for one month in May, 2010, and results to obtain the time dominant topics about purchase goods are shown.

1. はじめに

スーパーマーケットなどの小売店では，時間によって売れる商品が変化するという特徴を持つ．例えばお弁当類は昼食の時間帯に集中して購入され，野菜や精肉は夕方に最も購入される．このことから，ID 付き POS データを単純な個人の購入履歴としてではなく，商品を購入したときの状況に注目して POS データを分析する必要がある．

顧客がスーパーマーケットから購入する商品は，様々な要因によって変化する．例えば，スーパーマーケットの顧客はその日の気温や気候，また季節などによって購入する商品が変化する．本研究では，スーパーマーケットの ID 付き POS データから，このような顧客の購買行動の背後に存在する潜在的な要因の変化をコンテキストと考える．本稿では，時間に関するコンテキストを抽出した結果について報告する．

2. Context-Aware Recommendation

ユーザの嗜好はその時の状況に応じて変化するものである．そこで，コンテキストに基づいて，「現在の」ユーザが興味を示す推薦を行うといった研究が行われている．

Hariri らは，ユーザが聞いている楽曲のプレイリストをもとに，ユーザのコンテキストを考慮し，それに基づいて新しい楽曲を推薦する手法を提案した [Hariri 2012]．それぞれの楽曲には，last.fm[*1]から入手したタグをもとに Latent Dirichlet Allocation を使って生成した潜在トピック (dominant topics) が割り当てられている．パターンマイニングを使って，データセット中のプレイリストから出現頻度の多い dominant topics のパターンを抽出した．システムは，ユーザから新しいプレイリストが入力されたときに，パターンマッチングを使って次に求められるトピックを推定し，そのトピックに適した楽曲をユーザに推薦する．

Hariri らの推薦手法は，従来の履歴に基づく推薦と比較して，過去数件の楽曲が推薦結果に与える効果が大きいことである．そのため，その人が感じている楽曲の雰囲気を邪魔することのない楽曲の推薦が可能となっている．またパターンマッチングに際して，楽曲を潜在トピックを使って抽象化したことにより，順列の組み合わせの数が爆発することを防いでいる．コンテキストに基づくことで，嗜好の変化を検出することが容易になるほか，新しい楽曲にも柔軟に対応することができる．これらの成果を使って，LDA をもとにユーザの嗜好とコンテキストを統一したトピックモデルを Hariri らは提唱した [Hariri 2013]．

3. 分析手法

本研究では，Hariri らと同様にコンテキストに注目して，スーパーマーケットの ID 付き POS データを分析する．コンテキストに注目することで，その日の天候や気温，季節などといった情報を考慮した商品の推薦に応用することが可能となる．そのために，ID 付き POS データからコンテキストを生成する方法について考える．

ID 付き POS データには，誰が，いつ，何を購入したかが記録されている．顧客 *user* が商品 *item* を購入した時間を *time* とすると，ID 付き POS データは表 1 のような情報を持つ．

本稿では，Probabilistic Latent Semantic Analysis (PLSA)[Hoffman 1999] を使って顧客の購買行動に対して時間に基づくコンテキストを生成する方法を提案する．PLSA は Hoffman により提案された分析方法で，この手法を用いることで顧客の購買行動をいくつかの潜在変数によって説明することができる．潜在変数とは本来は確認することのできないものであるが，これを購買パターンとして見なすことで購買行動を複数のパターンの重ね合わせとして表現することができるようになる．商品 *item* が時刻 *time* で購入される確率，PLSA では潜在変数 z を使って次のように表現する．

$$p(item, time) = \sum_z p(item|z)p(time|z)p(z) \qquad (1)$$

確率 $p(item|z)$, $p(time|z)$ はそれぞれ潜在変数 z が生起したときに顧客 *item*, *time* が生起する確率であり，z が変われば *item* や *time* の出現確率も変化する．

商品 *item* が与えられたとき，その商品によって購買パターン z が行われる確率はベイズの定理を使って次のように書ける．

$$p(z|item) = \frac{p(item|z)p(z)}{\sum_z p(item|z)p(z)} \qquad (2)$$

連絡先: 吉田真，東京工業大学総合理工学研究科，神奈川県横浜市緑区長津田町 4259，m.yoshida@sp.dis.titech.ac.jp

*1 http://www.last.fm/

Hariri らの提案した手法に則り，確率 $p(z|item)$ がある値 l を超えたのであれば，購買パターン z を商品 $item$ のコンテキストとして与える．

4. 実験結果

2010 年 1 月から 2010 年 12 月まで 12ヶ月を対象に，1ヶ月ずつに分けて，時間コンテキストの生成を行った．赤池情報量規準を基にモデルを選択した結果，いずれの月も潜在変数は 7 か 8 つになった．2010 年 5 月のデータについて，それぞれの潜在変数の出現確率をまとめたものを表 1 に示す．表 1 において，塗りつぶされているマスは出現確率が 0.3 を超えるものであり，太字は 0.2 を超えているものである．

表 2 には，各潜在変数においてそれぞれ購入機会が多かった商品の例を挙げた．潜在変数によって時間毎に購入される商品に違いがでることが発見できた．

大きく 3 つ，朝，昼，晩の基本的な購買行動を取得できている．朝の時間帯 (z_1) では生活用品が買われる傾向があり，昼の時間帯 (z_4) では麺類や米飯類といった昼食用の商品が買われている．夕方の時間帯 (z_7) では，その日の夕食のための食材が購入されている．

3 つの基本的な購買行動の他に，特徴的なコンテキストも取得することができた．17 時から 19 時 (z_2) では，ビールと惣菜を購入するコンテキストが割り当てられている．購入されている惣菜はローストチキンやまぐろ，竜田揚げなど温めればおかずとして利用できるものが多く含まれている．z_7 のように食材ではなく加工済みの商品を購入していることから，自炊しない顧客が商品を購入する時間帯であると予想される．

お昼過ぎにあたる 15 時から 17 時 (z_7) では，果物や菓子類が購入されている．間食用のお菓子として購入されていると予想される．

19 時以降では，2 つのコンテキスト (z_3，z_6) が割り当てられている．ともに菓子類が主な購入品であるが，z_3 ではポテトチップスやサラダせんべいなどが惣菜や米飯類とともに購入される傾向がある．惣菜や米飯類などが購入されていることから，z_3 は何かしらの理由で夕食をとることができなかった人たちによる購買行動であると考えられる．一方で z_6 ではほとんど惣菜や米飯類は購入されておらず，菓子類が多くを占めることから，夜食用として商品が購入されていると考えられる．

5. 考察

時間に応じて，購入される商品に大きな違いがあることが今回の実験結果から判明した．特に昼食や夕食の食材を購入する，17 時以降にビールと惣菜を購入するといった特定の目的を持って商品が購入される時間帯をコンテキストとして取得することができた．このことは，野菜や肉／魚のような食材は夕方の時間帯に，果物や菓子類はお昼頃に推薦するといった戦略に応用できる．一方で今回の実験では，顧客ごとにどの時間，どの商品が売れたかまでは考慮されていない．そのために，このコンテキストは全体の消費行動を表すことはできても，必ずしもこの結果がすべての顧客を記述できるコンテキストであるとは限らない．

6. おわりに

本稿では，スーパーマーケットの ID 付き POS データを対象に，商品の売れ行きの変化をとらえるための PLSA によるコンテキストの抽出を行った結果について報告した．顧客の一日の購買行動は 7 から 8 個のコンテキストを使うことで記述できることがわかった．

表 1: 2010 年 5 月における各時間ごとの潜在変数の出現確率表 (潜在変数の順序を入れ替えていることに注意)

時間	z_6	z_1	z_4	z_5	z_7	z_2	z_3
8	1.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.38	0.57	0.05	0.00	0.00	0.00	0.00
10	0.08	0.49	**0.24**	0.02	0.17	0.00	0.00
11	0.02	**0.24**	0.44	0.12	0.15	0.03	0.00
12	0.01	0.02	0.58	**0.27**	0.04	0.09	0.00
13	0.01	0.09	0.33	0.39	0.06	0.12	0.00
14	0.00	**0.22**	0.06	0.38	**0.23**	0.11	0.00
15	0.00	**0.21**	0.01	**0.29**	0.38	0.12	0.00
16	0.00	0.13	0.03	0.15	0.51	0.18	0.00
17	0.00	0.03	0.09	0.04	0.51	0.31	0.02
18	0.00	0.03	0.09	0.00	**0.29**	0.42	0.16
19	0.00	0.14	0.03	0.00	0.00	0.30	0.54
20	0.10	0.08	0.00	0.00	0.02	0.01	0.80
21	0.44	0.00	0.00	0.00	0.00	0.00	0.56
22	0.48	0.00	0.00	0.00	0.00	0.00	0.52
23	0.49	0.00	0.00	0.00	0.02	0.00	0.49

表 2: 2010 年 5 月における潜在変数ごとの購入機会の多い商品

z_1	**生活用品** (洗剤，石けん・入浴剤，カイロ，防虫剤，キッチン用品，歯磨き粉，スリッパ)
z_2	**ビール，惣菜** (ローストチキン，まぐろ，焼豚，かつおのたたき)
z_3	**総菜** (唐揚げ，チルド惣菜)，**菓子類** (ポテトチップス，サラダせんべい，あられミックス)
z_4	**麺類** (うどん，うどんつゆ，和そば，天ぷら，かき揚げ)，**米飯類** (巻き寿司，お弁当，むすび)
z_5	**果物** (いちご，みかん)，**菓子類** (おはぎ，メロンパン，アイスクリーム，菓子パン)
z_6	**菓子類** (あられ，せんべい，チョコレート，ビスケット)
z_7	**食材** (ネギ，豆腐，牛肉，豚肉，鶏肉，野菜各種)

今回は時間に関するコンテキストのみを抽出したが，このコンテキストでは顧客それぞれの嗜好については無視されている．今後は，顧客それぞれがよく購入する商品などを踏まえていきたい．

参考文献

[Hariri 2012] Hariri, N., Mobasher, B. and Robin, B.: Context-Aware Music Recommendation Based on Latent Topic Sequential Patterns, *Proc. RecSys'12*, pp. 131-138 (2012).

[Hariri 2013] Hariri, N., Mobasher, B. and Robin, B.: Query-Driven Context Aware Recommendation, *Proc. RecSys'13*, pp. 9-16 (2013).

[Hoffman 1999] Hoffman, T.: Probabilistic Lantent Semantic Analysis, *Proc. UAI'99*, pp. 289-296 (1999).