1K2-1

ID-POS データによる来店行動・購買行動の潜在的時空間意味構造分析

Probabilistic Latent Special-Time Semantic Analysis of purchasing behavior using ID-POS data

山下和也[*1] 原田奈弥[*1*2] 黄冬陽[*1*3] 吉開朋弘[*4] 本村陽一[*1*3]
Kazuya Yamashita Nami Harada Toyo Ko Tomohiro Yoshikai Yoichi Motomura

[*1]産業技術総合研究所 National Institute of Advanced Industrial Science and Technology
[*2]株式会社豊田自動織機 TOYOTA INDUSTRIES CORPORATION
[*3]東京工業大学 Tokyo Institute of Technology
[*4] 日本気象協会 Japan Weather Association

Personal big data utilization will become more important from now on to increase individual's life satisfaction in population reduction society with protecting personal information. Using ID-POS data, we obtained probabilistic latent time semantic model by pLSA and Bayesian network in regard to purchasing behavior. We found the time structure of purchased items that referring period consisting of consecutive days. On the other hand, the time structure of customer which can mainly be explained by the day of the week was found out. And we found the relationship between the customer characterized by semantic time, customer's preference or dependence to this supermarkets, and it's purchased products. This showed the possibility of analyzing personal information by replacing it with time information of the individual. We think that the latent time semantic structure proposed in this study approaches the essence of time.

1. はじめに

人口が減少していく日本社会において、消費者や労働者など個々人の生活満足度を高めつつ、限りある物的・人的資源を活かしていく事が今後愈々重要になってくる。売れ残り等による食品廃棄量の削減や、適切な労働環境構築など、ある場所・ある時間に適切に物や人を配置する事により解決が図れる問題が多い。その達成には人間の行動をモデル化する事が必要であり、その元となる個々人の生活に関する様々な人間行動のパーソナルデータの利活用が益々重要になってくる。一方で今月末日(2017年5月30日)の改正個人情報保護法の全面施行を受けパーソナルデータを個人情報を保護しながらの利活用手法の確立が喫緊の課題となる。

そこで本研究では人がスーパー等の店舗に行って必要な物や買いたい物を選び商品を買うという行為を、それぞれ来店行動、購買行動として、人間の行動現象の一つの現れと捉え、それぞれの人や店が持つ潜在的時空間意味に注目した解析を行う。本解析は、個人情報を時空間情報に置き換える手法としての意味も持っている。

[石垣, 2010] では顧客の購買行動をモデル化し、ID-POS データを用いて顧客パーソナリティやライフスタイルを考慮した顧客と商品の同時分類をしている。その手法を用いて[石垣,2011], [本村,2011]ではそれぞれ来店人数予測、需要予測の精度を上げている。更に時間が持つ潜在的意味に注目した解析として一日のうちの来店時間帯で顧客を分類した[川島,2016]や、月毎の顧客の購買行動から季節感のあり/なしで顧客を分類した[原田, 2016]がある。一方、人の移動データを用いて空間が持つ潜在的意味で人を分類した[廣川,2015]がある。

本研究では、日や時間単位にも注目した潜在的時間意味解析に加え、時空間についての潜在的意味解析を提案し、過不足ない適切な商品の発注量や適切なレジ要員人数の決定支援、顧客満足の向上、ひいては社会の全体最適を図る事を目的とする。

このようにビッグデータ利活用のニーズの高まりを受け、本研究ではその一例として以下のビッグデータを扱う。関西に約 150 店舗を展開する総合スーパーの、2008 年 9 月から 2010 年 9 月までの 2 年強の ID-POS データである。総顧客 ID は 100 万を超えるが、2008 年 10 月から 2010 年 9 月までの 24 カ月のどの月にも少なくとも 1 回は来店した ID は約 35.4 万(顧客集合 A)である。その内、全期間の平均で週に約 2 回以上来店しているのは約 13.5 万 ID(顧客集合 B)である。その中にはプロフィールや購買行動に関するアンケートに回答した顧客のうちの 2122ID が含まれている。

2. 潜在的時空間意味構造分析

pLSA(Probabilistic Latent Semantic Analysis)は、2 つの変数 x,y が潜在変数 $z \in Z = \{z_1,...,z_k\}$ を介して生成されると仮定する。その場合 x,y の同時確率は式 1 で表される。

$$p(x,y) = \sum_k p(y \mid z_k)\, p(x \mid z_k) p(z_k) \qquad (式\ 1)$$

式 1 右辺の $p(x|z)$、$p(y|z)$、$p(z_k)$をパラメータとし、式 2 の対数尤度関数 L を 最大にするパラメータを EM アルゴリズムを用いて求める。N はデータの個数である。

$$L = \sum\sum N(x,y) \log p(x,y) \qquad (式\ 2)$$

この結果から、x,y の各値全てに各 z への所属確率

$p(x|z)p(z)/p(x)$、$p(y|z)p(z)/p(y)$を求められ、その中から最大確率値を取る潜在変数 z を付随させる事(ハードクラスタリング)もできる。

pLSA では、観測される変数の分布に対する事前の仮定が不要であり、異なる分布や形式の変数を同時に厳しい制約なく扱えるため、本研究のような人間行動に関する変数を扱う際にも

連絡先:山下和也, 国立研究開発法人 産業技術総合研究所, 135-0064 東京都江東区青海 2-4-7, 03-3599-8513, yamashita-kazuya@aist.go.jp

適している。pLSA の実行には産業技術総合研究所の知財ソフトウェア PLASMA を用いた。PLASMA では任意のクラス数を指定でき、それぞれのクラスタ数 k に応じたクラスタリング結果を得る事ができる。各解析において必要に応じて式 3 で定義される AIC が最小のクラスタ数のモデルを選択できる。式中の K はパラメータ数である。EM アルゴリズムには初期値依存性がある為、初期のパラメータに与える乱数値を複数試す事で、局所解に陥っていないモデルを選択できる。

$$\mathrm{AIC} = -2\ln L + 2\mathrm{K} \qquad \text{(式 3)}$$

本研究では、上記の x,y のうち 1 つの変数（仮に y にする）を時間（本研究では日や時）の変数にし、時間 y が x に対して潜在変数 z を介して生じていると捉える事で、潜在的に意味のある時間構造を抽出した。これにより x を潜在的意味時間で分類する事が出来、x が顧客の場合は顧客マスタにラベルとしてクラスタリングの結果を付随でき、そのクラスタリングラベルを変数としてベイジアンネットワークを構築する事で、その潜在的時間意味を説明する変数を探す事が出来る。更に、y を時空間（店舗の地理的位置と時間）とする事で、潜在的に意味のある時空間構造を抽出できる。

3. 小売り分野への潜在的時空間意味構造分析の適応

3. 1. 商品による潜在的時間意味構造

まず、アンケートに回答した顧客 2807ID が 2009 年 9 月から 2010 年の 8 月までの 1 年間に購入した商品を x とし、1 年間の各日付 365 日を y とした pLSA を行った。AIC が最小となった 19 クラスタリングの結果を図 1(a)に示す。1 年が 2~4 週間程度の連続した日にちに分かれた。これは暦より意味がある基準期間のようなものと考える事ができる。各期間の分かれ目（新しいクラスタの初日）は火曜日が多かった（連続した日にちで分かれた 17 クラスタのうち 8 クラスタ）。また異なる季節にまたがるほぼ火曜日だけから成る 1 クラスタがあった。火曜は本総合スーパーでポイントが多く付く曜日でもある。残る 1 クラスタにはどの日にもハードクラスタリングされておらずそのクラスタに含まれる商品は酒、肉類、化粧品、ペット用品等であり通年で購入される物品と考えられる。他のクラスタではそれぞれの季節に関連した商品が含まれていた。図 1(b)は上記解析の x はそのままで、y を日時（例：2010 年 5 月 23 日 14 時）としてより細かく分けた際の結果である。結果は日の分類と大きくは変わらず連続した期間で分かれた。期間の分かれ目を詳しくみると、日で分けた際の初日のクラスタの前日の夕方から新しいクラスタに分かれている場合が多くみられる。その時間帯には既に翌日用の商品が陳列され始めている事実との関連が考えられる。

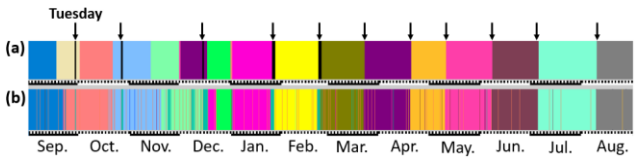


図 1．商品からみた潜在的時間意味構造

横軸は 1 年間　色別で各クラスタ期間を表示　上の矢印は火曜日を示す
(a)は日単位で分類　(b)は時単位で分類

次に、多数の商品を揃える大型店と、扱う商品がある程度限られている小型店を、それぞれ代表 1 店舗ずつ選び（この特定大型店を L、この特定小型店を S とする。なお L 店と S 店は約 1.2 ㎞＝徒歩約 16 分の距離である）、それぞれの店舗で同様の解析をした。毎月利用の顧客が（顧客集合 A のうち、L 店利用は約 3.3 万 ID，S 店利用は約 1 万 ID）それぞれの店舗で購入した商品を x、日を y として pLSA を行った結果を図 2 に示す。多店舗での結果と同様、特定店舗でも連続した日にちでクラスタが分かれた。AIC 最小のモデルを選択すると、L 店では 2 年間が 44 期間に（平均約半月で 1 期間）、S 店では 16 期間（平均 1 月半で 1 期間）となった。共に期間の始まりは火曜か金曜が多く、L 店では期間の約 8 割、S 店では約 6 割が火曜か金曜

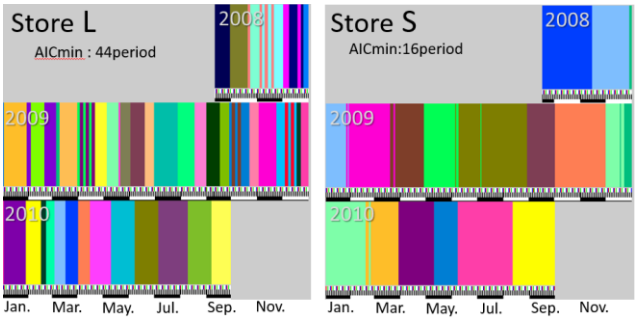


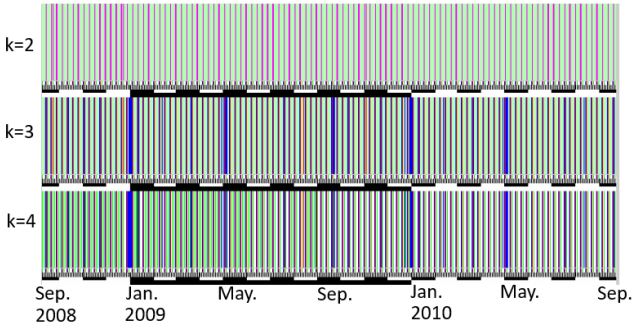
図 2. 店毎の潜在的時間意味構造

左図：大型店　右図：小型店　　横軸は 1 年間　2 年分のデータを同じ月日が横位置に揃うように並べた　色で各クラスタ期間を表示　左図と右図との間では色の対応付けはしていない

始まりであった。ただ L 店と S 店共通の期間の始まりは 2009 年と 2010 年それぞれの 1 月 3 日、2010 年 5 月 11 日（火）、2010 年 6 月 8 日（火）の 4 日のみであった。大型店と小型店では基準区間はずれていて固有の時間構造を持っている事がわかった。両店とも異なる年で全く同じ時間構造というわけではなく、期間の変わり目がずれている場合も多い。

3. 2. 顧客による潜在的時間意味構造

次に毎月利用かつ平均週約2回以上利用の顧客（顧客集合 B）の全店舗での購買を対象とし、x:顧客 ID y:日　重みを購買個数　として pLSA を行った。その結果を図 3 に示す。連続した日にちで分かれた商品の場合と異なり、概ね曜日で説明できるクラスタに分かれた。2 クラスタに分けた場合は、平日クラスタと週末（休日）クラスタに分かれた。3 クラスタの場合は、2 クラスタ分けの際の休日クラスタが土曜クラスタと日曜クラスタに分離した。4 クラスタ分けの場合は平日クラスタが火曜・木曜クラスタと月曜・水曜・金曜クラスタに分離した。この土クラスタ(15%の ID が所属)・日クラスタ(19%)・火木クラスタ(20%)・月水金クラスタ(46%)の 4 クラスタ分けの結果を以下で利用する。興味深い事に年末年始や 5 月のゴールデンウィークは土曜クラスタに属する日が多く、土曜クラスタとはより本質的には“当日休日かつ翌日休日クラスタ”、そして日曜クラスタは“休日最終日クラスタ”と言い換え得る。



横軸は 1 年間　k=2 (2 クラスタ分け結果) のピンクは “休日” クラスタ
k=3, k=4 の青色は” 土曜” クラスタ

なお以上の分離の仕方は、重みとして購買金額にした場合も同様であった。よりシンプルな重みとして来店の有無で 1 と 0 を与えた場合は AIC 最小が 7 クラスタ分けとなりやはり主に曜日で説明できる。

さらに、購買時間帯 (1 時間毎) を y とし、上記と同様の x:顧客 ID でも pLSA を行った。また 商品と顧客 ID との pLSA も行った。

上記に示した x が顧客 ID の 3 種類の pLSA (y が日の 4 クラスタ分けと、y が時間帯の 12 クラスタ分け、y が商品の 16 クラスタ分け) の結果のクラスタラベルを、各顧客 x にそれぞれ付与した。結果的に顧客集団 B に含まれるアンケートに回答した 2122ID にも顧客集団 B 全体での pLSA クラスタラベル (全体データで付けたラベルでグローバルラベルとも呼べるもの) が付与される事になる。そこにアンケート結果を変数として加えてベイジアンネットワークを構築した。その結果を図 4 に示す。

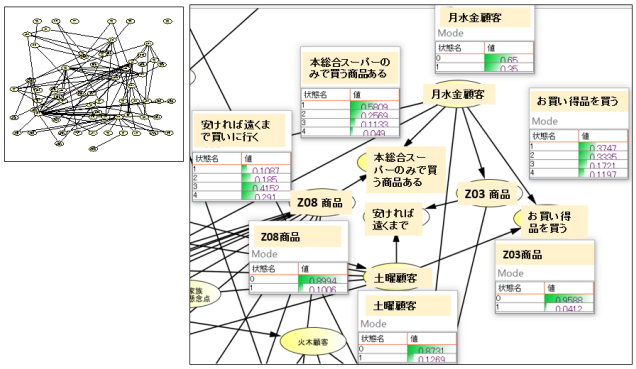


図 4. 潜在的時間意味構造モデル
左上図: 全体図　右: 一部の拡大図

月水金クラスタでない顧客 (土日火木に購買多い傾向の顧客) は、この総合スーパーのみで買う商品がある・高くても健康重視 (特に日曜顧客)、お買い得品を買うに positive 傾向で、安ければ遠くまで買いに行くには negative 傾向 (特に土曜顧客) であった。また商品で分類した Z03 クラスタの商品を買う傾向があり、Z08 クラスタの商品は買わない傾向があった。Z03 には余暇用品 (高級文具、園芸用品、野菜用肥料、水素充電池、植物の種) やこだわりのある生活用品などが多い。一方、Z08 は菓子パンや菓子、惣菜、炭酸飲料などすぐに飲食出来る加工食品が多い。

すなわち以上をまとめ考察すると、土日火金曜に購買しやすい顧客は、安ければ他のお店に行くというわけではなく、この総合スーパーに密着していて、たとえ高くても良い品を購入し余暇用品等、生活全体に関わる物品を購入している傾向があることが分かる。同時にお買い得品は買うなど価格にこだわっていないわけではなく、ポイントが多く付く曜日を選んでいることからもこの総合スーパーを積極的に利用している顧客と言える。一方身体の都合や時間の制限などで、他のお店に行く選択肢が無い顧客も含まれていると思われる。

3. 3. 顧客による潜在的時空間意味構造

時間に空間を加えた時空間に対しての解析のため、空間として店舗の位置情報を加える事で、本発表では潜在的時空間意味構造を見出し、時空間が定まれば一意に決まる環境情報、たとえば気象との関係を示す。

4. まとめ

ID-POS データより人の購買行動から pLSA とベイジアンネットワークを用いて潜在的に意味を持つ時間構造を抽出した。購買された商品からみると、連続した日にちから成る基準期間とも言える時間構造が見られた。顧客からみると、結果として主に曜日で説明出来る時間構造が見られた。そしてこの時間構造で特徴付けられた顧客と、本総合スーパーへの密着度、及びその購買商品との関係を見出す事ができた。これにより個人をその個人が持つ潜在的時間意味に置き換えて解析出来る可能性を示した。この結果は個人情報保護法改正の下でのデータ活用にも貢献出来るのではないか。カレンダーや時計で表される暦や時刻は物質世界における一つの目盛りであり、人間行動を対象とする際、本研究で提案している潜在的時間意味構造がより時間の本質に迫るものと考えられる。更には潜在的時空間意味構造分析を提案した。

謝辞

データを提供して頂き、現場でのニーズを詳しく教えて頂いた総合スーパーの皆様に厚くお礼申し上げます。

本研究は NEDO 委託事業「人間と相互理解できる 次世代人工知能技術の研究開発」の支援を受けて行いました。

参考文献

[Hoffman 1999] Hofmann, Thomas. ”Probabilistic latent semantic analysis.” Proceedings of the Fifteenth conference on

Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1999.

[石垣,2010]石垣司, 竹中毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係の自動抽出法, オペレーションズ・リサーチ, 56, 2, (2010) .

[石垣,2011] 石垣司, 竹中毅, 本村陽一: 潜在クラスモデルによる流通量販店舗の来店人数予測の精度改善の評価, 人工知能学会全国大会, (2011)

[川島,2016] 川島健佑:価値観とライフスタイルを考慮した確率的潜在意味構造モデルの活用技術 平成 27 年度東京 工業大学知能システム科学専攻修士論文 (2016)

[原田, 2016] 原田奈弥, 山下和也, 本村陽一: ID 付 POS データによる購買行動の季節変化の分析と視覚化, 人工知能学会社会における AI 研究会 27 回研究会,2016.

[廣川,2015] 廣川典昭 ,村山敬祐,本村陽一:パーソントリップデータからの確率的潜在時空間意味構造モデリング ～地域活性化や観光サービスへの応用を目指して～, 人工知能学会全国大会(第 29 回)

[本村,2011] 本村陽一, 竹中毅, 石垣司: 条件付層別差分モデルによる需要予測の高精度化, 人工知能学会全国大会, 1B3-1, (2011)

[本村,2006] 本村陽一, 岩崎弘利:ベイジアンネット技術, 東京電機大学出版局, (2006).