

気象条件と店舗特性を考慮した需要変動のモデル化に関する一考察

1X14C068-4 世古 裕都
指導教員 後藤 正幸

1 研究背景・目的

スーパーマーケットなどの小売店では、日々来店する顧客の需要変化に対応するため、蓄積された購買履歴データを活用した需要予測へのニーズが高まっている。こうした需要変化は、顧客自身の嗜好や外的要因の変化に影響される。その中でも特に、店舗の特性や、日々の生活と密接な関係にある気象条件の影響は大きいと考えられる。そのため、多数の店舗の購買履歴データと気象条件の統合的な分析により、店舗ごとの気象条件の変化に起因する商品（以下、アイテム）の需要変動を定量的に把握することは大変重要である。

一方で、複数の異なる特性を持つマーケティングデータを分析するモデルとして、潜在クラスモデルが知られている。このモデルは、ユーザとアイテムの背後に潜在的な変数の存在を仮定することで、異質なデータ間の関係性のモデル化が可能となり、対象とする膨大な数のユーザやアイテムを、潜在クラスという軸での解釈を可能とする。代表的な手法の1つに Probabilistic Latent Semantic Analysis [1] (以下、PLSA) がある。PLSA は、データが複数の潜在クラスから確率的に生起することを許容する確率モデルとして知られ、ユーザの潜在的な購買傾向を確率的に表現することができる。

そこで本研究では、気象条件および店舗特性と、アイテムの需要変動との関係性を表現する潜在クラスモデルを提案し、店舗特性と気象条件、並びにアイテムの需要変動の関係性を分析する手法を示す。また、提案モデルを実データに適用した分析実験を行い、実際に気象条件に影響されやすいアイテム（以下、気象感応度の高いアイテム）を抽出することでモデルの有用性を示す。最後に結果に対する考察を行う。

2 従来研究

購買履歴データから、ユーザの購買傾向を把握する潜在クラスモデルとして、PLSA がある。PLSA は、購買傾向の類似したユーザの集合である潜在クラスを仮定し、ユーザとアイテムの共起関係を潜在クラスによる条件付確率分布で表したモデルである。

ここで、 J 個のアイテム集合を $\mathcal{X} = \{x_j : 1 \leq j \leq J\}$ 、 I 人のユーザ集合を $\mathcal{Y} = \{y_i : 1 \leq i \leq I\}$ 、 K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ と定義する。PLSA の確率モデルは式 (1) で表される。

$$p(x_j, y_i, z_k) = p(z_k)p(x_j|z_k)p(y_i|z_k) \quad (1)$$

3 提案

3.1 提案への着想

従来の PLSA では、ユーザとアイテムを潜在クラスを通じて紐づけ、これらの共起関係を分析することにより、ユーザのアイテム購買傾向を分析する。その一方で、本研究では、気象条件および店舗特性によるアイテムの需要変動を分析することが目的である。そこで、これらを潜在クラスで紐づけ、共起関係をモデル化することを考える。これにより、気象条件によるアイテムの需要変動が類似した店舗の分析と、類似した店舗群での気象感応度の高いアイテムの抽出、そして、店舗群ごとに、気象条件の変化に伴うアイテムの潜在的な需要変動の表現が可能となると考えられる。

これは、アイテムの陳列方法や提供方法などに関するマーケティング施策に、直接的に繋げることができると考えられる。また、気象条件の変化がアイテム A の売上と関係があることを事前に把握できれば、気象予測に基づきアイテム A の適切な在庫管理が可能となる。さらに、他店舗の情報も加味することで、自店舗と同様の気象条件を持つ他店舗でアイテム B の気象感応度が高いと事前に分析できれば、自店舗でも潜在的に売上が変化するアイテムとしてマーケティングに活用することができる。

3.2 対象とする気象条件

本研究では、アイテムの売上に最も影響を与える気象条件として、気温の変化に焦点を当てる。特に、前日の平均気温 (0 時から 1 時間おきに 24 回計測された気温の平均) と、その日の平均気温の差である気温前日差を用いる。例えば、ある日の平均気温が同じ 3°C であったとしても、前日が暖かければ「寒い」と感じ、寒ければ「暖かい」と感じるといったように、体感気温は前日との気温差に依存する部分が大きく、平均気温のみに着目した分析ではこのような気温変化の影響は考慮できない。また、気温前日差では、基準となる前日の平均気温が観測値であり、観測値からの上昇・下降に着目していることから、実際に各店舗で分析できる気象条件である。そのため、マーケティングに実用的な指標と考えられる。

3.3 提案モデル

アイテム、気温前日差、店舗特性の共起関係を表現し、提案モデルはどの地域の店舗で、どのような気温の変化が、どのようなアイテムに影響するのかに関する分析を可能にする。

いま、 J 個のアイテム集合を $\mathcal{X} = \{x_j : 1 \leq j \leq J\}$ 、 I 個の店舗集合を $\mathcal{S} = \{s_i : 1 \leq i \leq I\}$ 、気温前日差を t 、 K 個の潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ と定義する。このとき、提案する確率モデルを式 (2) で定義する。

$$p(x_j, s_i, t, z_k) = p(z_k)p(x_j|z_k)p(s_i|z_k)p(t|z_k) \quad (2)$$

このモデルの1つの生起事象は、気温前日差 t の日に、店舗 s_i でアイテム x_j が1つ購買されたときである。提案モデルのグラフィカルモデルを図1に示す。

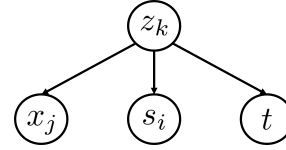


図1. 提案モデルのグラフィカルモデル

ただし、アイテムと店舗特性の確率分布 $p(x_j|z_k)$ 、 $p(s_i|z_k)$ には多項分布を仮定し、気温前日差 t には、式 (3) で表される正規分布を仮定する。

$$p(t|z_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(t-\mu_k)^2}{2\sigma_k^2}\right) \quad (3)$$

ただし、 μ_k 、 σ_k^2 はそれぞれ k 番目の正規分布の平均、分散である。

3.4 パラメータの学習

総購買アイテム数 N 件の購買データのうち、 n 番目に購買されたアイテムを $a_n (\in \mathcal{X})$ 、その購買が行われた店舗を $b_n (\in \mathcal{S})$ とする。また、その日の気温前日差を c_n とすると、 n 番目の購買データはこれらの共起 (a_n, b_n, c_n) で表現できる。このとき、全 N 件の購買データに対する対数尤度 LL は以下の式 (4) で表される。

$$LL = \sum_{n=1}^N \sum_{k=1}^K \log p(a_n, b_n, c_n, z_k) \quad (4)$$

このモデルのパラメータは EM アルゴリズムを用いて、対数尤度関数 LL を最大化するように以下の更新式により推定する。

E-step)

$$p(z_k|a_n, b_n, c_n) = \frac{p(z_k)p(a_n|z_k)p(b_n|z_k)p(c_n|z_k)}{\sum_{k=1}^K p(z_k)p(a_n|z_k)p(b_n|z_k)p(c_n|z_k)} \quad (5)$$

M-step)

$$p(z_k) = \frac{1}{N} \sum_{n=1}^N p(z_k | a_n, b_n, c_n) \quad (6)$$

$$p(a_n | z_k) = \frac{1}{N p(z_k)} \sum_{n=1}^N \delta(a_n, x_j) p(z_k | a_n, b_n, c_n) \quad (7)$$

$$p(b_n | z_k) = \frac{1}{N p(z_k)} \sum_{n=1}^N \delta(b_n, s_i) p(z_k | a_n, b_n, c_n) \quad (8)$$

$$\mu_k = \frac{\sum_{n=1}^N p(z_k | a_n, b_n, c_n) c_n}{\sum_{n=1}^N p(z_k | a_n, b_n, c_n)} \quad (9)$$

$$\sigma_k^2 = \frac{\sum_{n=1}^N p(z_k | a_n, b_n, c_n) (c_n - \mu_k)^2}{\sum_{n=1}^N p(z_k | a_n, b_n, c_n)} \quad (10)$$

ここで、 $\delta(\alpha, \beta)$ は、 $\alpha = \beta$ である場合に 1、そうでない場合には 0 の値を示すインジケータ関数である。

4 実データを用いた分析

4.1 分析条件

分析データの対象期間は、2013 年 3 月 1 日から 3 月 31 日とした。長期間を対象とすると、季節性のあるアイテムの売上が気温などと強い相関を持つため、気象感応度の高いアイテムの抽出が困難になると考えられる。そこで、1 カ月という短期間のデータを分析対象とする。

分析対象とする購買データは、中部地方を中心に展開する某小売チェーンにおいて購買された、生鮮カテゴリに属するアイテムとする。生鮮カテゴリのアイテムの多くは、店舗ごとに加工方法を変えるなど、オペレーションレベルで提供方法を変更可能なアイテム群である。このため、需要の変化に柔軟に対応することができ、各店舗での施策に直接的に繋がる有用な情報となるアイテムである。

総店舗数は $I = 174$ 、総アイテム数は $J = 3,745$ 、全データ数は $N = 37,166,162$ である。また、潜在クラス数については、気象と店舗の解釈性の高いクラス数を探索的に求め、 $K = 8$ と設定した。

4.2 評価方法

本研究では、アイテム、気温前日差、店舗特性の共起関係を分析し、気象感応度の高いアイテムを抽出することを目的としている。一方で、事前分析から、気温前日差に関係なく、日常的に一定量購買されるアイテムの存在が確認された。そこで、「気温前日差が影響しないクラスが存在する」という仮説のもと、気温前日差が影響するクラスと、影響しないクラスのアイテムを比較することで、気象感応度の高いアイテムを抽出する。すなわち、前日との気温差の影響がないクラスでは出現せず、影響があるクラスでは出現するアイテムを気象感応度の高いアイテムとする。

4.3 実験結果と考察

気温前日差と店舗特性について、結果を表 1 に示す。ここで、店舗所在地の特徴は、各クラスの店舗所属確率上位 15 店舗の所在地をまとめたものである。また気温前日差は、各クラスの気温前日差の平均 μ_k と分散 σ_k^2 に基づき、気温前日差が上昇したときは +、下降したときは -、どちらの場合もとりうるときは ± と表記している。

表 1. 各クラスに属する店舗と平均気温前日差の傾向

z_k	店舗所在地の特徴	気温前日差	$p(z_k)$
z_1	北陸地方	±	0.0223
z_2	中部地方西側	±	0.2358
z_3	愛知県南部・静岡県南部	±	0.2529
z_4	岐阜県南部・愛知県西部	+	0.2357
z_5	岐阜県南部・愛知県南部	+	0.0400
z_6	愛知県西部	+	0.0059
z_7	中部地方西側	-	0.0050
z_8	中部地方東側	-	0.0074

表 1 の結果から、クラスごとに気温前日差と店舗特性について特徴がみられる。クラス $z_1 \sim z_3$ では、気温前日差が上昇、下降のどちらもとりうる可能性があることから、それ

ぞれの地域において気温前日差の影響が少ない店舗やアイテムが出現するクラスと考えられ、前節で述べた仮説に相当するクラスといえる。また、クラス $z_4 \sim z_6$ は気温前日差が上昇したときに、クラス z_7, z_8 では気温前日差が下降したときにアイテムの売上が伸びるクラスと解釈できる。

店舗所在地の特徴も加味すると、クラス z_2, z_7 は、中部地方西部の店舗の所属確率が高く、クラス z_2 は気温前日差と関係なく売上が伸びるアイテムの出現確率が高いクラスである。また、クラス z_7 は気温前日差が下降した際に売上が伸びるアイテムの出現確率が高いクラスと解釈できる。

続いて、各クラスの気象感応度の高いアイテムを抽出する。表 2 に、クラス $z_1 \sim z_3$ で出現確率の高いアイテムを、表 3 にクラス $z_4 \sim z_8$ での気象感応度の高いアイテムを示す。ただし、クラス $z_4 \sim z_8$ でのアイテム出現確率上位アイテムのうち、クラス $z_1 \sim z_3$ でのアイテム出現確率上位 30 品を気温前日差に関係なく売上が伸びるアイテムとして除いたものを、気象感応度の高いアイテムとして示している。

表 2. 各クラスの代表的なアイテム Top3

z_k	No.1	No.2	No.3
z_1	もやし	きゅうり	ホウレン草
z_2	PB 惣菜	もやし	ブロッコリー
z_3	きゅうり	PB 惣菜	しめじ

表 3. 各クラスの気象感応度の高いアイテム Top3

z_k	No.1	No.2	No.3
z_4	舞茸	豚肉しゃぶしゃぶ用	牛小間切れ
z_5	バナナ	菜の花	刺身わかめ
z_6	惣菜	エリンギ	たこ刺身
z_7	まぐろ刺身	水菜	大葉
z_8	ねぎ	刺身盛合せ	豚肉生姜焼き用

表 2 に示されているアイテムはクラス $z_1 \sim z_3$ での出現確率が高いことから、気温前日差に関係なく売上が伸びるアイテムといえる。「もやし」や「きゅうり」は複数のクラスを横断的に出現しやすく、気象条件だけでなく店舗所在地とも関係なく購買されやすいアイテムであると考えられる。一方で、「ホウレン草」や「ブロッコリー」などのアイテムは、クラスを横断的に出現していないことから、店舗所在地特有のアイテムであると考えられる。

表 3 から、クラス z_4 では「舞茸」が、クラス z_8 では「ねぎ」が気象感応度の高いアイテムとして抽出されていることがわかる。また、クラス z_4 では「豚肉しゃぶしゃぶ用」、クラス z_8 では「豚肉生姜焼き用」が気象感応度の高いアイテムとして抽出されている。一般に豚肉などはブロックで仕入れられ、各店舗で加工されるものであり、店舗ごとの需要に応じて提供方法を変化させることができる。このように、同じ食材であっても、気象条件に対応するクラスの特徴に応じて加工方法を変更することで需要に対応できると考えられる。

以上の結果から、具体的な施策として、例えば北陸地方の店舗では、気温前日差に関わらずホウレン草の売上が伸びるため、店頭での陳列を前にするなどの施策が打てると思われる。同様に、同じ豚肉であっても、岐阜県南部・愛知県西部にある店舗で、気温前日差が上昇した店舗では、豚肉をしゃぶしゃぶ用とすることで、需要の変化に対応できると考えられる。

以上より、気象条件による需要傾向が類似した店舗の分析、および気象感応度の高いアイテムの抽出ができることを確認し、具体的な施策が可能となることを示した。

5 まとめと今後の課題

本研究ではアイテム、気象条件、店舗特性の共起関係を分析するモデルを提案し、実データによる実験を通じてモデルの有用性を示した。また、実験結果から、クラス間の比較を行うことで、気象感応度の高いアイテムの抽出を行った。

今後の課題として、AIC などを用いた適切な潜在クラス数の決定、他の気象条件も考慮したモデルの考案、気象感応度の定量化、得られた結果の活用方法などが挙げられる。

参考文献

- [1] T. Hofmann, “Probabilistic Latent Semantic Analysis,” *Proc. of UAI’99*, pp.289-296, 1999.