

周期性のあるデータで生活形態を判別

水野貴之[†] 大西立顕[‡] 清水千弘[#] 渡辺努[§]

国立情報学研究所情報社会相関研究系, 総合研究大学院大学複合科学研究科, JST さきがけ[†]

東京大学大学院情報理工学系研究科[‡]

麗澤大学大学院経済学研究科[#]

東京大学大学院経済学研究科[§]

1. はじめに

近年, センサー端末の普及に伴い電子マネーを使った商品の販売数や顧客の来店数の把握, 携帯電話の GPS データによる施設の利用者数の推定, スマートメーターによるエネルギー使用量など, 様々な日中の 24 時間周期性を持つビッグデータが集まってきている.

本稿では, 24 時間活動する Web 店舗における価格改定や販売などの店舗活動データを利用して, 様々な店舗の活動パターンを抽出し, そのパターンを生み出す店舗属性を発見する手法について示す. このような手法を開発することにより, 各商品における店舗間の競争状況を把握し, 商品の過剰な値崩れ防止に向けた, 市場のルール作りに役立つ.

店舗属性に関する統一的なデータは価格比較サイトから集めることができる. 価格比較サイトには, 各店舗の取扱商品数や取り扱っている主要な商品カテゴリ, 安い店なのか高い店なのか, 店舗の定休日などの情報が表示されている. また, 店舗の価格改定の時系列から, コンピュータによって他店の価格を自動的に追尾しているのか, 人手による定期的な価格改定をおこなっているのかを推定することができる.

本研究では, 約 1000 店舗のそれぞれについて, 日々 30 分間ごとに価格改定回数と Web で購入ボタンを押された回数が記録されているデータを用いる. はじめに, 各回数の時系列データから特徴的な日中の店舗活動パターンを抽出するのに, 特異値分解とテンソル分解が有効であることを述べる. 次に, 決定木により, 各特異値に対応する特異ベクトルが, どのような店舗属性を反映した結果であるのかを推定できることを述べる.

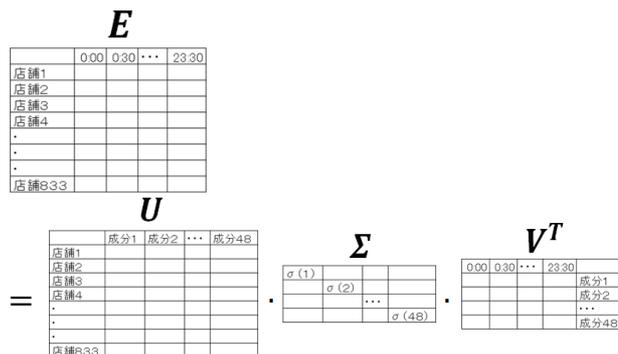


図1 価格改定回数行列の特異値分解

2. 日中パターンの抽出

はじめに, 特異値分解を利用して特定のある日の価格改定回数行列の情報を圧縮し, 日中の店舗活動パターンを抽出することから始め, それを全ての曜日や全ての日に対応できるように順次拡張していく.

店舗*i*の*d*日の時刻*t*の正規化した価格改定回数 $e_d(i, t)$ を価格改定回数行列 E_d の (i, t) 成分と定義する. 店舗数が N 件, 1 日の時間分解能が 30 分間であれば, この行列のサイズは $N \times 48$ である. この行列を,

$$E_d = U_d \Sigma_d V_d^T \quad (1)$$

と特異値分解する. 図1は店舗数を $N = 833$ とした場合の特異値分解の概念図である. 特異値分解した各行列の解釈としては, 行列 V_d^T の各行は価格改定回数の d 日における様々な 24 時間周期成分 (店舗活動の日中パターン) を表し, 行列積 $U_d \Sigma_d$ の列はそれぞれの 24 時間周期成分を各店舗が d 日においてどれぐらい持っているかを表す. そして, 行列 Σ_d の成分は, そもそも, d 日において各 24 時間周期成分が全店舗トータルでどれぐらい存在するかを表す. ここで, 便宜上, 特異値を表す行列 Σ_d の成分は大きい順に $\sigma_d(1) > \dots > \sigma_d(48)$ と書く.

Lifestyle analysis based on High-frequency data

[†]T. Mizuno · National Institute of Informatics

[‡]T. Ohnishi · The University of Tokyo

[#]C. Shimizu · Reitaku University

[§]T. Watanabe · The University of Tokyo

	成分1	成分2	...	成分48
店舗1 金曜日				
店舗1 土曜日				
店舗1 日曜日				
店舗1 月曜日				
.				
.				
.				
店舗833 水曜日				
店舗833 木曜日				

図2 週周期を加味した価格改定回数行列

1日の時間分解能が30分間の場合、特異値は48個算出される、つまり、価格改定回数（店舗活動）の48種類の日中パターンが抽出される。しかしながら、全てのパターンが統計的に有意なわけではない。我々は、ランダム行列理論[1, 2]を応用し、統計的に有意な日中の店舗活動パターンを見出す。

価格改定回数行列 E_d から全ての日中パターンを取り除くために、全ての成分 $e_d(i, t)$ に対して、ランダムに選ばれた別の成分 $e_d(j, k)$ との配置変えをおこなう。このようにして作られたランダム価格改定回数行列 E'_d について同じく特異値分解をおこなう。このようにして得られた行列 V'^T_d の各行が示す48個の24時間周期成分は、サンプルが有限なことにより生じてしまった統計的に有意でない擬似パターンである。ランダム価格改定回数行列 E'_d の最大特異値 $\sigma'_d(1)$ は、ランダム価格改定回数行列に含まれる最大の擬似パターンの大きさを表す。

従って、特異値分解は以下のように、 $\sigma'_d(1)$ よりも大きい統計的に有意な r 番目までの成分と、 $\sigma'_d(1)$ 以下の統計的に有意でない r 番目以降の成分に分割できる。

$$E_d = \sum_{i=1}^r \sigma_d(i) \mathbf{u}_d(i) \mathbf{v}_d(i)^T + \sum_{i=r+1}^{48} \sigma_d(i) \mathbf{u}_d(i) \mathbf{v}_d(i)^T \quad (2)$$

価格改定回数行列 E_d の特異値 $\sigma_d(i) > \sigma'_d(1)$ に対応する特異ベクトルが、統計的に有意な店舗活動の日中パターンである。上記で示した手順により統計的に有意な複数の店舗活動の日中パターンを抽出することができる。

次に、各店舗の曜日ごとの店舗活動の日中パターンを抽出するために、価格改定回数行列 E を図2のように拡張する。ここで、成分 $e(i_w, t)$ は店舗 i の曜日 w の時刻 t の標準化された平均価格改定回数を表す。この行列を特異値分解することにより、店舗活動の日中パターンの曜日依存性

を明確にできる。

季節変動まで加味した、店舗活動の日中パターン抽出には、特異値分解を拡張したテンソル分解（CP分解）が利用できる。店舗 i の d 日の時刻 t の価格改定回数 $e(i, d, t)$ を成分に持つ価格改定回数テンソル E をテンソル分解すれば良い。

3. 日中パターンを生む店舗属性の推定

k 番目に大きい特異値に対応する価格改定回数の24時間周期成分（日中パターン）が、どのような店舗属性によって作られているのかを知るために、各店舗のそのパターンの含有量を表す行列 U_d の k 列目の成分 $u_d(1, k), u_d(2, k), \dots$ を目的変数に設定し、各店舗の属性リストを説明変数とした回帰木による解析をおこなうことにより、日中パターンを生み出す店舗属性を絞り込むことができる。

4. まとめ

本稿では、Web店舗における価格改定回数の時系列データを例に取り、店舗活動の日中パターンの抽出と、そのパターンを生み出す店舗属性の関係の推定手法について述べた。これにより、出入りの激しいWeb市場における店舗構成の状況を把握することができ、その構成が変化した場合に、商品にどのような価格変動（特に値崩れの発生確率の変化）が生じるかを調べることが可能となる。

このような分析は、店舗活動以外にも、GPSデータによる各施設の利用者数の時系列データと、各施設の利用会員のアンケートベースのサンプルデータを教師データとして、施設利用者数の日中パターンから施設利用者の属性推定をおこなうことができる。これは、店舗経営に役に立つ。また、エネルギー使用量の時系列データから、顧客の属性推定をすることも可能であり、この結果は、顧客に対するきめ細かいサービスの提供に役立つ。

参考文献

[1] V. Plerou, et al., PRL 83, 1471-1474, 1999.
 [2] T. Mizuno, et al., PTP Supplement 194, 181-192, 2012.