

情報爆発時代に向けたコンテキストウェア情報推薦方式 のための特徴パラメタ最適化に関する考察

奥 健太[†] 中島 伸介[†] 宮崎 純[†] 植村 俊亮[‡] 加藤 博一[†]

奈良先端科学技術大学院大学情報科学研究科[†] 奈良産業大学情報学部情報学科[‡]

1. はじめに

情報爆発時代に向けて、大量の情報から嗜好に合った情報をユーザに提供する情報推薦技術が注目されている。我々は、これまでにコンテキストウェア情報推薦を目的とした Support Vector Machine (SVM) [1]に基づくユーザ嗜好モデル化方式を提案している [2]。ここで、モデルの構築に必要な学習データは、特徴ベクトルにより表現する必要があるが、我々の手法では、推薦対象のアイテムに由来する特徴パラメタと、ユーザの状況 (コンテキスト) に由来する特徴パラメタとを合わせたものを特徴ベクトルにより表現している。しかし、ユーザの意図を正しくモデルに反映させるためには、これら特徴パラメタの表現方法を最適化する必要があるが、この最適化に関する議論は十分に行えていない。そこで、本稿では、ユーザ嗜好のモデル化に際し、必要/不要なパラメタを踏まえた特徴パラメタの最適化について考察を行う。

2. 特徴パラメタの設計

表 1 は、飲食店情報を推薦対象アイテムとしたときに用いる、アイテムおよびコンテキストに由来する主な特徴パラメタを示している。それぞれ表 1 の上段、下段にあたる。また、これら特徴パラメタは、そのとりうる値のタイプによって、表 2 のように分類できる。

しかし、これら特徴パラメタの中には、ユーザ嗜好のモデル化にとって、不要なものも含まれている可能性がある。また、一つ一つの特徴パラメタの表現方法についても適切に正規化する必要がある。そこで、次節では、これらを踏まえた特徴パラメタの最適化に関する考察を行う。

3. 特徴パラメタ最適化に関する考察

アイテムに由来する特徴パラメタとコンテキストに由来する特徴パラメタとでは、それぞれ性質が異なるため、最適化の方針もそれぞれで異なる。したがって、これらの特徴パラメタを別々に考察する。

表 1 飲食店情報推薦を対象とした主な特徴パラメタ設計

特徴パラメタ	パラメタ値	タイプ
料理ジャンル	[和食], [洋食], [中華] など	カテゴリ型
店の平均予算	0~10,0000 円	連続値型
設備	[個室あり], [禁煙席あり] など	2 値型
時刻	0:00~23:59	周期型
休日	[休日], [休日前]	2 値型
ユーザの予算	0~10,0000 円	連続値型
同伴者	[1人], [家族], [恋人] など	カテゴリ型

表 2 特徴パラメタのタイプ

パラメタタイプ	説明
2 値型	設備の有無などを {0/1} の 2 値で表現する。
カテゴリ型	n 次元 ($n \geq 3$) のパラメタにより表現し、そのうち該当する一つのパラメタに 1 を、それ以外のパラメタに 0 を与える。
連続値型	量や程度などを [0,1] の連続値で表現する。
周期型	周期的な属性を持つ情報を 2 次元 ([0,1], [0,1]) で表現する

3.1 アイテムに由来する特徴パラメタ

アイテムに由来する特徴パラメタは、アイテムの特徴を必要かつ十分に表現するものでなければならない。ここでは、以下の場合における最適化方針について考察する。

(1) ほとんどすべてのアイテムにおいてほぼ同じ値をとるパラメタの場合

例えば、「ライブあり/なし」という 2 値型の特徴パラメタを考える。いま、このパラメタの値が、ほとんどすべてのアイテムにおいて「ライブなし」であった場合、このパラメタでは、アイテムの特徴を表現するには不十分であると考えられる。したがって、このようなパラメタはモデル構築にとって不要であると判定し、排除するという検討をする。

また、図 1 は、カテゴリ型の特徴パラメタである [料理ジャンル] を示したものである。このカテゴリは、[和食], [洋食], [中華], [居酒屋] の四つのパラメタで構成されるものとし、それぞれの度数を示している。2 値型の例と同様に、図 1(a) のように、ほとんどのアイテムが [和食], [洋食], [中華] のいずれかであった場合、[居酒屋] というパラメタは不要であると判定され、それを排除する。また、図 1(b) のようにパラメタを排除した結果、そのカテゴリを構成するパラメタ数が 2 になった場合、これは 2 値型のパラ

A Study of Optimization of Feature Parameters for Context-Aware Recommendation Method in Info-Plosion Era.

[†]Kenta Oku, Shinsuke Nakajima, Jun Miyazaki and Hirokazu Kato : Nara Institute of Science and Technology

[‡]Shunsuke Uemura : Nara Sangyo University

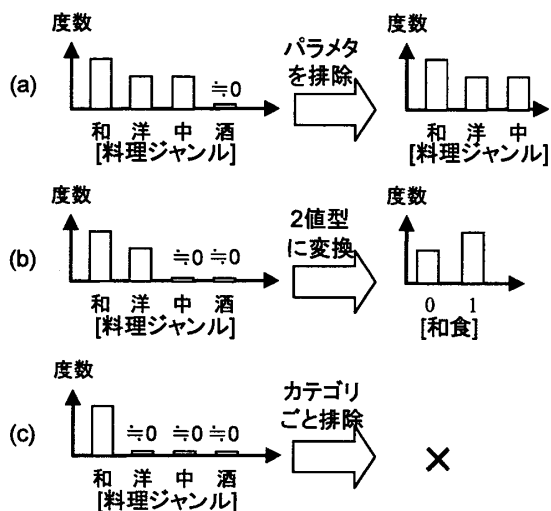


図1 カテゴリ型パラメータの最適化方針

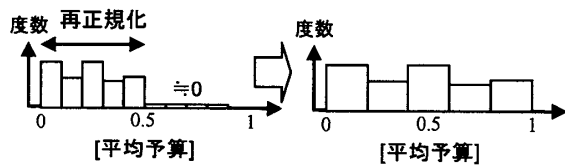


図2 連続値型パラメータの最適化方針

メタに一致するため、これら二つのパラメータを一つの2値型パラメータとして統合する。さらに、図1(c)のように、ほとんどのアイテムが[和食]であった場合、そもそもこのカテゴリは不要であると判定し、このカテゴリ自体を排除する。

図2は、連続値型の特徴パラメータである[店の平均予算]の度数分布を示している。このような場合には、度数が限りなく低い領域は、ほとんど不要であると考えられるため、度数が高い領域において、パラメータ値を再度[0,1]に正規化する。

(2) パラメータ間の相関が極端に高い場合

例として、[昼の平均予算]と[夜の平均予算]という二つの特徴パラメータを考える。いま、これらの相関が極端に高かったとする。このような場合には、片方のパラメータは、アイテムの特徴を表現するのに、冗長であると判断できるため、これら二つのパラメータを一つのパラメータ[平均予算]として統合することを検討する。

3.2 コンテキストに由来する特徴パラメータ

我々は、コンテキストを“人の価値判断基準に影響を及ぼす状況”と定義している[2]。この定義に則り、ユーザの価値判断基準への影響の大きさというものを尺度とし、それに基づいた最適化方針を検討する。

図3のように、SVMでは与えられた学習データ

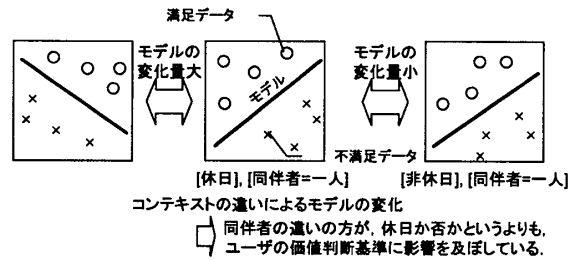


図3 モデルの変化量と価値判断基準への影響

集合に適合するようなモデルを構築する。いま、図3のように、ある一つのパラメータのみの値を変化させたとき、このモデルの変化量が大きければ、そのパラメータはユーザの価値判断基準に影響を及ぼすと考えられる。この考えに基づいて、影響を及ぼさないと判定されるパラメータを排除することで最適化を行う。なお、モデルの変化量を求める方法の一つとして、SVMの特徴空間上に無作為に配置されたデータに対する、各モデルによる判別結果の違いをみる方法がある。

しかし、この方法では、一つのパラメータによる影響の判定しか行っていない。一つのパラメータだけでは、ユーザの価値判断基準には影響を及ぼさないと判定されたとしても、別のあるパラメータと組み合わせることにより、影響を及ぼす場合があると考えられる。したがって、場合によっては、二つ以上のパラメータも組み合わせることで影響を調べることも必要である。

4. まとめと今後の課題

本稿では、SVMによるユーザ嗜好モデルの設計において、アイテムとコンテキストの特徴パラメータの性質の違いを踏まえながら、特徴パラメータ最適化に関して考察を行った。

今後は、3.節で述べた内容を踏まえて、具体的に最適化手法を設計する。さらに、実データによる検証実験を通じて、定量的に最適化手法の評価を行う。

謝辞

本研究の一部は、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」(A01-34, 課題番号 19024058) および経済産業省「情報大航海プロジェクト モデルサービスの開発と検証」(プロジェクト名: プロファイルパスポート事業) による。ここに記して謝意を表します

参考文献

- [1] Cortes, C. and Vapnik, V.: Support-Vector Networks, Machine Learning, Vol.20, No.3, pp.273.297 (1995).
- [2] 奥 健太, 中島 伸介, 宮崎 純, 植村 俊亮: 状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム, 情報処理学会論文誌データベース, (TOD34), 情報処理学会, 2007年06月.