室蘭工業大学
学術資源アーカイブ
Muroran Institute of Technology Academic Resources Archive

可変精度ラフ集合モデルを用いたレコメンデーションシステム

メタデータ	言語: Japanese
	出版者: 日本知能情報ファジィ学会
	公開日: 2013-08-22
	キーワード (Ja):
	キーワード (En):
	作成者: 天野, 庄平, 工藤, 康生, 村井, 哲也
	メールアドレス:
	所属:
URL	http://hdl.handle.net/10258/2205

室蘭工業大学
学術資源アーカイブ
Muroran Institute of Technology Academic Resources Archive

可変精度ラフ集合モデルを用いたレコメンデーションシステム

著者	天野　庄平，工藤　康生，村井　哲也
雑誌名	ファジィシステムシンポジウム講演論文集
巻	23
ページ	809-810
発行年	2007-08
URL	http://hdl.handle.net/10258/2205

可変精度ラフ集合モデルを用いたレコメンデーションシステム

A Recommendation System Using Variable Precision Rough Set Models

天野 庄平　工藤 康生　村井 哲也
Syohei Amano　Yasuo Kudo　Tetsuya Murai
室蘭工業大学　北海道大学
Muroran Institute of Technology　Hokkaido University

Abstract: We introduce variable precision rough set models to a recommendation system using Pawlak's rough set theory we have proposed. By extracting reducts and decision rules based on user's queries, the proposed recommendation system recommends some goods that the user may like though the characters are different with the queries. However, too many goods may be recommended, and some of recommended goods have low connection with user's queries. To avoid these problems, we propose a new recommendation method using variable precision rough set models, and evaluate the proposed method by experiments.

1 はじめに

ラフ集合 [2] による縮約計算および決定ルールの抽出の応用として，著者らは商品検索システムにおいて指定された検索条件から，検索条件に該当する商品以外にもユーザが好むと思われる商品を提案するレコメンデーションシステムを試作した [3]．このシステムでは，検索条件から決定ルールと上近似を構成し，上近似に含まれ，かつ，検索条件に該当しない商品を提案する．しかし，このシステムでは多くの商品が上近似に含まれ，検索条件と関係性が低い商品も提案されてしまうという問題が生じる場合がある．この問題を解決するために，本研究では可変精度ラフ集合モデル [4] を用いて，著者らの従来手法の改良を試みる．

2 背景

2.1 決定表と縮約

ラフ集合によるデータ解析では，複数の属性とその値からなる決定表を対象とすることが多い．決定表は 4 項組 $(U, C \cup D, V, \rho)$ で表される．ここで，U は対象の有限集合，C は条件属性集合，D は決定属性集合で $C \cap D = \emptyset$ を満たす．V は値の集合，ρ は要素 x の属性 a での値 $\rho(x, a) \in V$ を表す関数である．決定表の例を表 1 に示す．属性集合 B に基づく識別不能関係 R_B を

$$xR_By \Longleftrightarrow \forall a \in B,\ \ \rho(x, a) = \rho(y, a)$$

と定義する．識別不能関係は同値関係となる．xR_By は B に含まれる属性のみでは x と y を区別できないことを表す．決定属性は全体集合 U の分割 $\{D_1, \ldots, D_m\}$ を与え，各 D_i は決定クラスと呼ばれる．決定表の縮約とは，各決定クラス D_i の下近似と同じ分割を与える必要最小限の条件属性の集合である．

表 1: 決定表

	形態	駐車場	敷金	礼金	風呂
s1	アパート	あり	あり	あり	独立型
s2	アパート	あり	なし	なし	一体型
s3	マンション	なし	なし	なし	一体型
s4	マンション	なし	あり	あり	独立型
s5	マンション	あり	なし	なし	独立型
s6	アパート	あり	なし	あり	一体型

2.2 可変精度ラフ集合モデル

可変精度ラフ集合モデルとは，矛盾を含む情報をラフ集合の枠組みで扱う手法である．任意の集合 $X, Y \subseteq U$ に対して，誤判別度 $c(X, Y)$ を以下の式で定義する．

$$c(X, Y) = \begin{cases} 1 - \frac{|X \cap Y|}{|X|}, & |X| > 0. \\ 0, & |X| = 0. \end{cases}$$

ここで，$|X|$ は集合 X の要素数である．誤判別度 $c(X, Y)$ は，X の全ての要素を Y に分類した場合に生じる誤分類の割合を表す．精度 (誤差を許す度合い) β $(0 \leq \beta < 0.5)$ を用いて，識別不能関係 R_B による X の β -下近似 $\underline{B}_\beta(X)$ および β-上近似 $\overline{B}_\beta(X)$ は次のように定義できる．

$$\begin{aligned} \underline{B}_\beta(X) &= \{x \in U \mid c([x]_{R_B}, X) \leq \beta\}. \\ \overline{B}_\beta(X) &= \{x \in U \mid c([x]_{R_B}, X) < 1 - \beta\}. \end{aligned}$$

2.3 ラフ集合を用いたレコメンデーションシステム

従来手法ではユーザが入力した検索条件から，縮約を計算し決定ルールと上近似を構成することで，上近似に含まれ，かつ検索条件に該当しない商品を提案する．以下に手順を示す．

(i) 検索条件 ([属性=値] の連言) の入力．
(ii) 検索条件に該当するデータ集合を Q，検索条件に条件に該当しないデータ集合を Q^c とする．
(iii) 検索条件を決定属性として，縮約 B を求める．
(iv) 同値関係 R_B による上近似 $\overline{B}(Q)$ を求める．
(v) $x \in \overline{B}(Q) \cap Q^c$ となる x を提案する．

3 可変精度ラフ集合を用いたレコメンデーションシステム

従来手法では一度すべての縮約を求め，上近似を構成し，その上近似に含まれる商品を提案していた．しかし，この方法では縮約計算に時間がかかる，入力された検索条件とは関係性の低い商品が多く提案されるなどの問題点がある．これらの問題を解決するために本研究では，可変精度ラフ集合モデルにおける β-上近似を用いる．

β-上近似を計算する際に，以下の 2 条件を満たす上近似の構造を保存する縮約 $A \subseteq C$ を用いる [1].

(i) $\overline{A}_\beta(X) = \overline{C}_\beta(X)$,
(ii) 任意の $B \subset A$ に対して $\overline{B}_\beta(X) \neq \overline{C}_\beta(X)$.

条件 (i) より，上近似の構造を保存する縮約は，条件属性集合 C に基づく β-上近似と同じ上近似をもたらす．よって，そのような縮約 A を具体的に求めなくても，C に基づく β-上近似 $\overline{C}_\beta(X)$ を求めれば十分である．

この性質を用いて，検索条件に該当するデータの集合 Q に対して C に基づく β-上近似 $\overline{C}_\beta(Q)$ を求め，$x \in \overline{C}_\beta(Q) \cap Q^c$ となる x を提案する．

4 実験

前節で提案した手法を用いたシステムを試作し，実験を行った．実験に用いるデータは仮想の賃貸住宅に関するデータ 600 件 (属性数 12) とした．実験では，11 名の被験者がシステムを実際に使用した．精度 $\beta = 0$ で被験者が検索条件を入力し，提案された各商品に対して「A:良い」「B:どちらともいえない」「C:悪い」の 3 段階で評価した．また，β を 0.05 刻みで増加させ提案される商品の変化を調査した．提案された商品の個数の平均を表 2 に，提案された商品における A 評価および C 評価の割合を表 3 に示す．ただし，表 3 では 2 名の被験者について $\beta = 0.1$ で商品が出力されなくなったため分析から除外している．

表 2: β に対する提案商品の個数の平均

精度 β	0	0.05	0.10	0.15	0.20
平均値	21.4	18.5	10.5	8.4	5.9

表 3: 提案商品における A 評価および C 評価の割合

精度 β	0	0.05	0.10	0.15	0.20
A の割合	0.37	0.37	0.38	0.47	0.45
C の割合	0.25	0.25	0.25	0.22	0.19

表 2 より，β の値を増加させると提案商品の個数が減少する傾向が見られた．このことから，β-上近似を用いることで多くの商品が提案されることを回避できることが示唆される．一方，表 3 より A 評価と C 評価の割合に大きな変化は見られなかった．理由として，精度の増加によって β-上近似から取り除かれる同値類に，検索条件に該当するデータと該当しないデータが混在すること，および一部の検索項目 (駐車場の有無など) で検索条件に該当しないことが低評価の要因になりやすいことなどが考えられる．

5 おわりに

本研究では，ラフ集合を用いたレコメンデーションシステムに可変精度ラフ集合モデルを導入し，多くの商品が提案される問題の解決を試みた．今後の課題として精度 β の動的な調整，β-下近似の利用などが挙げられる．

参考文献

[1] Inuiguchi, M.: Several approaches to attribute reduction in variable precision rough set model, *Proceedings of MDAI2005*, Springer-Verlag, pp.215–226, 2005.

[2] Pawkak, Z.: *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers,1991.

[3] Yasuo, K., Syohei, A., Takahiro, S. and Tetsuya, M.：A Simple Recommendation System Based on Rough Set Theory, *Kansei Engineering International*, Vol.6, No.3, pp.19–24, 2006.

[4] Ziarko, W.: Variable Precision Rough Set Model, *Journal of Computer and System Science*, Vol.46, pp.39–59, 1993.

連絡先
工藤　康生
〒 050-8585 北海道室蘭市水元町 27-1
室蘭工業大学工学部情報工学科
TEL 0143-46-5469 FAX 0143-46-5499
E-mail: kudo@csse.muroran-it.ac.jp