

人々の経験を活かすための経験マイニング

倉島 健

日本電信電話株式会社 NTT サイバーソリューション研究所

Mining Experiences for Reusing People's Past Experiences

Takeshi KURASHIMA

NTT Cyber Solutions Laboratories, NTT Corporation

1. 背景と目的

ブログに代表される CGM(Consumer Generated Media)は、個人の行動履歴や、行動に基づく主観的な記述を多く含むという他メディアとは異なる性質を持っている。本来、経験とは個人レベルで年月の経過とともに少しずつ蓄積されていくものである。CGMの爆発的な普及は、他人の過去の経験を利用するという方法で、自らの経験不足を補い、賢い行動選択を行う可能性をもたらした。特に、他人の過去の失敗経験から学ぶことの意義は大きい。しかし、CGMの大半は、自然言語で記述された非構造データである。さらに、日々膨大な量のコンテンツが発信されるため、そのほとんどが再利用される機会を持たず捨てられてしまっている現状がある。

本研究においては、人間が他人の過去の経験を活かしながら賢く行動する世界を目指し、大量の経験情報集合の中から、ユーザにとって“興味深い”知識を発見する経験マイニング技術の検討を行っている。本稿では、これまでに筆者らが行った取り組みと今後の課題について述べる。

2. 経験マイニング

経験マイニング技術とは、人々によって日々発信されるブログから経験情報を抽出して保持し(構造化)、大量の経験情報集合から、ユーザにとって潜在的に“興味深い”知識を発見する技術である。人間の経験は、経験情報: $E = \{時間, 空間, 行動, 評価, 事実, 感情\}$ で表現できると定義する。また、経験に関する知識は、相関ルール形式: $X \rightarrow Y$ (X, Y はそれぞれ経験要素に属する語の集合であり、 $X \cap Y = \emptyset$) で表現できると定義する。例えば、“昨日、嵐山で紅葉を見ました。きれいでしたが、道が渋滞していて困りました”という経験は、 $E = \{昨日, 嵐山, 紅葉を見る, きれい, 道が渋滞, 困る\}$ という経験情報で表現できる。経験情報集合の中から得られる知識の例としては、“5月、蛍を見る→銀閣寺”といった時間、空間、行動間のルールや、“ディズニーランド、5月→困惑”のような動作主の主観(感情)を考慮したルールがある[1][2]。図1は経験マイニング技術の概要を述べた図である。

経験マイニングにおいて重要な課題は、得られたルールの潜在的な“興味深さ”を如何にして測るかである。ルールの“興味深さ”の指標は、経験情報を利用するユーザの背景知識を考慮するか否かで客観的指標と主観的指標とに分けられる[3]。筆者らは実際に約4800万件の大規模ブログデータから相関ルールを抽出した。得られたルール群を lift, χ^2 乗値, J-measure 等のユーザの背景知識を考慮しない客観的指標を用いてランキングすることで、これ

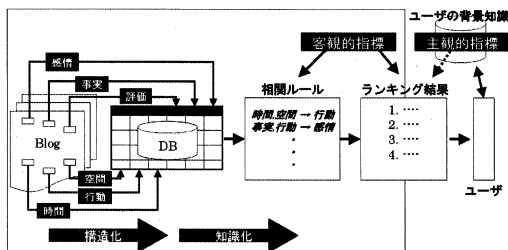


図1. 経験マイニング技術の概要

らの指標の有用性を検証した結果、支持度や確信度といった指標と比較し、特に lift が興味深いルールの発見に効果的であることを示した[2]。しかし、これら客観的指標を用いた上位のランキング結果は、多くのユーザにとって既知である場合も多く、依然として、ユーザに、新たな、意外な発見をもたらす機会は少ないのが現状である。

3. 今後の課題

情報の“興味深さ”は、ユーザ自身の経験や、状況、目的によって異なる。今後、ライフログの普及により個人の行動履歴をはじめとした、個人情報の蓄積が進むことを考えると、ユーザの背景知識を考慮する主観的指標を用いたルールのランキングが重要になってくると考えている。新奇性(novelty)や意外性(unexpectedness)の高い情報を、人や状況や目的に応じて提示することで、蓄積した経験情報を活かす機会も増えるであろう。また、経験情報集合から得られるもうひとつの知識として、シーケンシャルパターンを経験情報集合から発見することも視野に入れている。この結果は、人間の未来の行動、思考パターンの予測に用いることができると考えている。

参考文献

- [1] Tezuka, T., Kurashima, T. and Tanaka, K.: Toward Tighter Integration of Web Search with a Geographic Information System. Proc. of WWW2006, pp.277-286, 2006.
- [2] 倉島健, 藤村考, 奥田英範: 大規模テキストからの経験マイニング, 電子情報通信学会第16回データ工学ワークショップ(DEWS2008), 2008.
- [3] Geng, L., et al.: Interestingness measures for data mining: A Survey, ACM Computing Surveys, vol.38, No.3, Article 9, 2006.