

ユーザモデルとコンテキストが共に動的に変化する 場合における情報推薦手法の提案

松山 学[†] 大園 忠親[†] 伊藤 孝行[†] 新谷 虎松[†]

名古屋工業大学大学院工学研究科情報工学専攻[†]

e-mail: {manabu, ozono, itota, tora}@ics.nitech.ac.jp

1 はじめに

近年、ユーザ個人に特化した情報を推薦するパーソナライゼーションに関する研究が盛んに行われている。ユーザ個人に特化した情報を推薦するためには、ユーザが持つ複数のドメインに対する興味とユーザの興味変化を捉える必要がある。ユーザの興味変化に関する議論 [1] では、ユーザの興味には大きく分けて2つの興味があると考えられている。1つは長期的興味と呼ばれ、ユーザが潜在的に持つ興味であり変化が少ないものであり、もう1つは短期的興味と呼ばれ、ニュースの最新性や世間の動向に左右され変化するものとされている。本研究では、まずニュース中のメタデータを利用したユーザモデルの構築を行う。そして、世間の動向とユーザモデルとの相関関係を利用したニュース記事推薦手法を提案する。ここで、本研究では世間の動向をコンテキストの動的な変化と呼び、ポータルサイトで配信されるカテゴリ毎の記事数に着目して定義する。本論文の構成を以下に示す。第2章ではユーザモデル構築に関してを述べる。第3章では動的に変化するコンテキストについて述べる。第4章で評価実験及び考察し、第5章で本論文をまとめる。

2 メタデータを用いたユーザモデルの構築

本研究では、ニュース中に付加されるメタデータであるタイトルとカテゴリを利用する。ユーザモデルを構成するキーワードベクトルの重み $W_{i,j}$ とキーワードベクトルの更新式を式 (1), (2) に示す。

$$W_{i,j} = \begin{cases} 1 & (w_i \in p_{j_title}) \\ 0.5 + 0.5 \frac{freq_{i,j}}{\max_j freq_{kj}} & (freq_{ij} > 0) \\ 0 & (freq_{ij} = 0) \end{cases} \quad (1)$$

$$V_c^{t+1} = \sum_{p \in P_c^{new}} V_p + \alpha * V_c^t \quad (2)$$

[†]Information Recommendation Technique in Case Both a User Model and a Context Change Dynamically
Manabu Matsuyama, Tadachika OZONO, Takayuki ITO, and Toramatsu SHINTANI

Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya, 466-8555 JAPAN

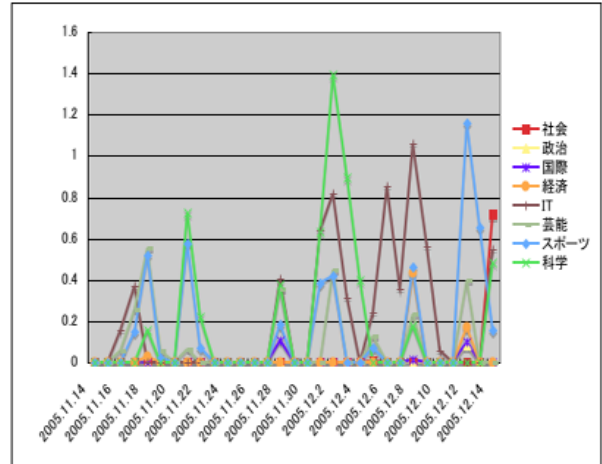


図 1: エネルギー値 E_c による推移図

$W_{i,j}$ はユーザが閲覧したニュース記事 p_j 中の語 w_i の重みを示し、 $freq_{i,j}$ は、 p_j 中の語 w_i の頻度を示し、 p_{j_title} は記事 p_j のタイトルを示している。閾値 α はカテゴリの更新の調整を行なうための値であり、0 から 1 の値をとる。 $p \in P_c^{new}$ は新規閲覧履歴中の記事を示す。 V_c^t はカテゴリ c における過去の閲覧履歴を示す。本研究では、ユーザの興味をカテゴリ毎のエネルギー値 E_c^t により表現する。 E_c^t の更新式は式 (3) により計算される。ここでカテゴリ c に関して新規更新が無い場合、興味の衰退を表現するため閾値 β を E_c^t から引いた値を E_c^{t+1} とする。ここで、閾値 β は 0 から 1 の値をとる。図 1 に 1 ヶ月間のエネルギー値の推移図を示す。

$$E_c^{t+1} = E_c^t + \sum_{p \in P_c^{new}} \cos(V_c^t, V_p) \quad (3)$$

3 動的に変化するコンテキスト

動的に変化するコンテキストの定義

本研究では goo ニュースを対象にカテゴリ毎に配信される記事数の調査を行った。その結果、カテゴリ毎に配信数に偏りがあるものの週単位で記事配信数に対して規則性を伺うことができた。必要な変動成分を取得するためこの規則性を取り除いた値を動的に変化するコンテキスト EC_c^t と定義する。ある時点 t におけるカテゴリ c の記事配信数を O_c^t 、 t における曜日の平均

配信記事数を S_c^t とすると、ある t におけるカテゴリ c の EC_c^t は式 (4) により計算される。

$$EC_c^t = O_c^t - S_c^t \quad (4)$$

動的に変化するコンテキストに関する検証

上述で定義した EC_c^t の有意性を検証するため、対象期間中に行われた大きなイベントである日本プロ野球日本シリーズと定期的に行われる Jリーグに関するイベントと EC_c^t との相関関係を検証を行った。 EC_c^t の推移図を図 2 に示す。結果、規則的なイベントである Jリーグに関しての相関係数 (検証期間 60 日) は 0.54、検定値が $4.88 > 2.66$ (有意水準 1% 棄却値)、日本シリーズに関しては相関係数 (日本シリーズ期間付近 13 日間) は 0.75、検定値が $3.79 > 3.11$ (有意水準 1% 棄却値) であったため十分有意であると考えられる。また、他のカテゴリに関しても EC_c の高い期間とイベントとの関係性を確認できた。これらの検証結果、考察より本研究では $EC_c^t > 0$ の時、世間においても何かしらのイベントが発生していると考えられる。

4 相関関係を利用したニュース記事推薦

本研究ではユーザモデルとコンテキストが共に動的に変化する場合において双方の相関関係を利用しニュース記事推薦手法を提案する。相関関係を式 (5) により評価する。ここで V_{E_c, EC_c} は対象期間中の E_c と EC_c 値と共分散を示す値であり、 S_{E_c} 、 S_{EC_c} は各値の標準偏差を示す値である。 $R(E_c, EC_c) > \gamma$ の時、カテゴリ c においてコンテキストの動的な変化とユーザモデルの相関関係があるとする。本提案手法は時系列 t において $E_c > 0$ の場合、そのカテゴリを選択する。次に選択されなかったカテゴリに対して $R(E_c, EC_c) > \gamma$ かつ $EC_c > 0$ の場合、カテゴリ c に関して推薦を行う。

$$R(E_c, EC_c) = \frac{V_{E_c, EC_c}}{S_{E_c} S_{EC_c}} \quad (5)$$

評価実験

1ヶ月間の実験データを元にユーザモデルを構築し、その後ユーザに 10 日間ニュースを閲覧してもらった。ここでユーザが閲覧した記事を正解例とし、同時に各カテゴリ 20 件ずつランダムに記事を取得し、混在した実験対象データから記事推薦を行った。そして、再現率 (recall) と適合率 (precision) により評価を行う。比較対象として、本手法と同様にニュース記事におけるメタデータを利用しユーザモデルを構築する PVA[2] を用いる。類似度計算にはベクトル空間モデルを共に利用する。類似度が閾値よりも高いものを推薦記事とする。 $\alpha = 0.95, \beta = 0.5, \gamma = 0.1$ とする。実験結果を図 3 に示す。本手法と PVA の評価値の平均値は recall

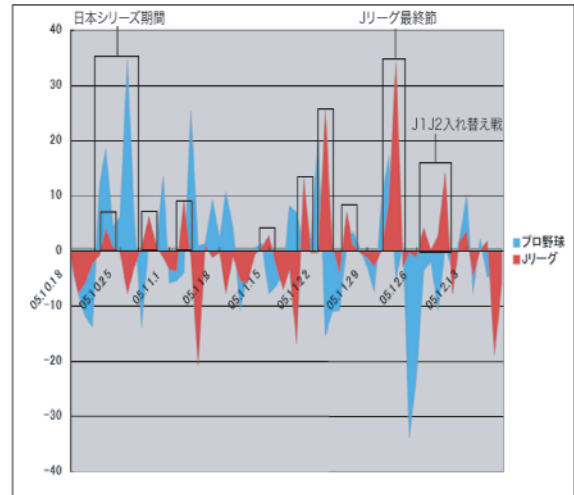


図 2: EC_c とイベントの関連性

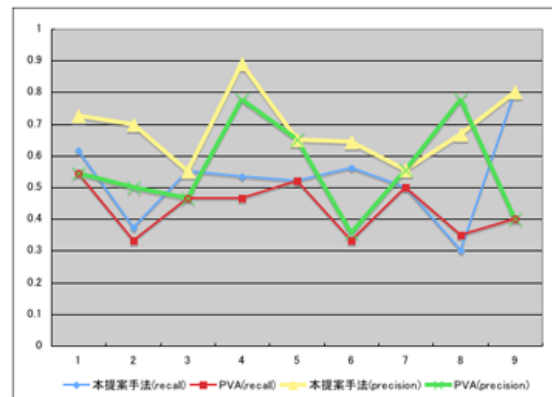


図 3: recall と precision

に関しては 0.53, 0.45, precision は 0.69, 0.57 であり PVA よりも上回った結果を得た。理由として被験者が 3 つカテゴリに関して相関関係があったことが言える。

5 おわりに

本論文では、ユーザモデルとコンテキストが共に動的に変化する場合における情報推薦手法の提案について述べた。相関関係を利用することにより短期的な興味関しても表現する可能性を伺うことができる。

参考文献

- [1] Daniel Billsus, Michael J. Pazzani : “ A Personal News Agent that Talks, Learns and Explains ”, Third International Conference on Autonomous Agents, (1999).
- [2] Chien Chin Chen, Meng Chang Chen and Yeali Sum : ” A Web Document Personalization User Model and System, ” Workshop on Machine Learning, Information Retrieval and User Modeling, (2001).