

業務連携を支援する社内ブログ記事推薦システム

Enterprise Blog Recommender System to Support Collaborative Work

佐藤政寛
Masahiro Sato

出雲 英剛
Hidetaka Izumo

富士ゼロックス株式会社 研究技術開発本部
Research and Technology Group, Fuji Xerox Co., Ltd.

Enterprise SNS and blog systems have been introduced into many companies in order to promote collaborative works. However, it takes significant man-hours to find useful information among a massive amount of post data. In this presentation, we propose a novel enterprise blog recommender system which extracts articles helpful for each employee to conduct collaborative works.

1. はじめに

業務連携の促進を目的に、社内 SNS や社内ブログで情報共有を行う企業が増えている。しかしながら、大量の投稿情報から有用な情報を見つけるには、多大な工数を要するという課題がある。一方でプロジェクト単位の情報共有システムでは、プロジェクトに密接に関係する情報のみが共有されるので情報検索に要する時間は短いが、意外性のある発見や新たな業務連携につながりにくい。そこで、多様な情報を含む社内 SNS や社内ブログから業務に役立つ情報を適切に取捨選択して自動的に配信する、情報推薦システムが望まれる。

一般的な情報推薦では、協調フィルタリングや内容ベースフィルタリングが用いられる[Ricci 11]。しかし協調フィルタリングでは、評価値が与えられていない新着投稿記事の推薦は難しい。内容ベースフィルタリングでは、協業が難しい社員の情報も配信されてしまったり、非常に密に協業しているので当たり前の情報が配信されてしまったり、といった問題がある。近年はユーザー間のつながりにもとづく Social Recommendation[Guy 11, Tang 13]の研究も盛んであるが、社員間のつながりのみにもとづく推薦では、社員が複数のプロジェクトを抱えていた場合、業務上関連の低いプロジェクトの情報まで配信されてしまうことになる。

従来の業務内容や社員間のつながりによる推薦では、個々の社員がどの業務でどの社員と協業しているかという、社内業務を考える上で重要な観点が見えにくい。そこで業務内容ごとに社員間のつながりを分析し、業務単位の協業度を指標とした推薦モデルを考案した。この協業度が高すぎず低すぎない適切な範囲のブログ記事を推薦することで、既存の協業範囲を超えた新たな業務連携を促進することが可能となる。

2. 提案手法

2.1 システムフレームワーク

提案システムのフレームワークを図 1 に示す。システムはブログサービスに記事が投稿されたことを逐次検知する。人物辞書内の人物名 B が記事内で言及されている場合、そのブログ投稿者 A は記事内の人物 B と協業していると判定する。次に業

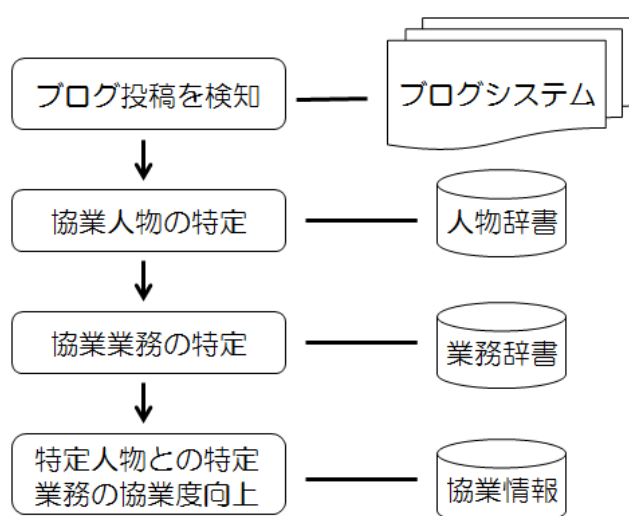


図 1 ブログ推薦システムのフレームワーク

務辞書内の各キーワード $\{w\}$ の有無から、協業業務を特定する。そして投稿者 A と特定人物 B との業務キーワード w_1 における協業度 $CV(A, B; w_1)$ を増加させる。投稿者 A による別の投稿記事において、同じ人物 B と同じキーワード w_1 が言及されていた場合、協業度 $CV(A, B; w_1)$ をさらに増加させる。

このようにあらかじめ特定人物との業務単位協業度ベクトル $CV(A, B; \{w\})$ の値を算出しておき、新しい記事が投稿された時に推薦を行う。新しい記事が B による投稿であるとき、人物 A に対する記事の関連度は、業務単位協業度ベクトルと記事内に含まれるキーワードベクトルとのコサイン類似度で求められる。この関連度が一定値の範囲内であれば、人物 A にこの記事が推薦される。協業度の算出段階では投稿記事内の人物言及関係を用いたが、推薦記事の抽出段階ではキーワード単位の協業度が既知なので、投稿記事内に人物が直接言及されている必要はない。

2.2 システム動作

実際にこのシステムがどう振る舞うか、具体例を用いて説明する。図 2 の投稿 1、投稿 2 で「S さん」と「I さん」は「ブログ」を含む記事を相手の名前の言及つきで 2 回投稿しており、「形態素」を含む記事は 1 回投稿している。よって「ブログ」に関する協業度は「形態素」に関する協業度より大きいことがわかる。

連絡先: 佐藤政寛, 富士ゼロックス株式会社 研究技術開発本部, 横浜市西区みなとみらい 6 丁目 1 番, sato.masahiro@fujixerox.co.jp

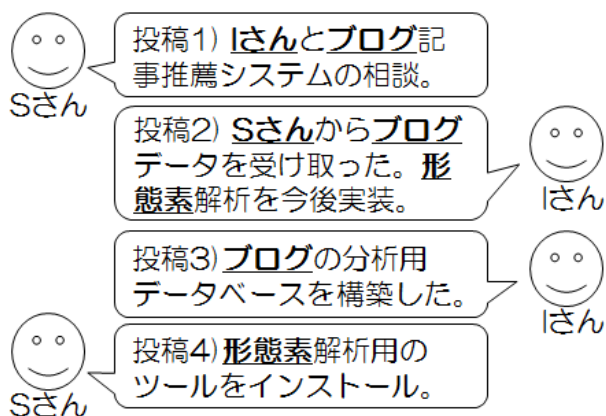


図2 投稿記事例

関連度が thresh_L 以上 thresh_H 以下の記事が推薦され、 $\text{CV}(S, I; \text{ブログ}) > \text{thresh_H} > \text{CV}(S, I; \text{形態素}) > \text{thresh_L}$ 、であるとする。このとき「ブログ」を含む「Iさん」の投稿3は「Sさん」に推薦されず、「形態素」を含む「Sさん」の投稿4は「Iさん」に推薦される。「ブログ」に関しては「Sさん」と「Iさん」の協業度は高いため、互いに既知の内容が多いと考えられる。「形態素」に関しては協業度が中程度なので、未知の情報である可能性が高い一方、互いに無関係な情報である可能性も低く、推薦に適していると考えられる。

3. 実験

3.1 実装

人物辞書は、所属部署の社員約 50 名の姓名を用いた。業務辞書は、ブログ記事内の名詞を MeCab[Kudo 04]を用いて抽出し、名詞が連続するものは複合名詞として扱った。抽出された単名詞、複合名詞のうち、文字数や出現頻度が一定範囲内のもをフィルタリングして、最終的な業務辞書(約 2000 語)として用いた。

ユーザーによって投稿される記事の量は異なるため、ユーザーごとに協業キーワードの出現頻度の和が 1 になるよう規格化を行った。

$$\text{CV}^*(A, B; w_k) = \text{CV}(A, B; w_k) / \sum \text{CV}(A, B; w_k)$$

またキーワードによって出現頻度が異なるため、他のユーザーの協業キーワード出現頻度平均との比を取った。

$$\text{CV}^*(A, B; w_k) = \text{CV}^*(A, B; w_k) / \text{mean}(\text{CV}^*(A, B; w_k))$$

3.2 評価

提案システムの効果検証のため、ユーザー5名に対して官能評価を行った。評価者には、「新たな業務連携につながる情報か否か」(投稿者とこれまで協業してこなかった内容に関して、情報交換したいと感じたものなら○)、という観点で評価を依頼した。

3.3 結果

図3に、関連度上位5件の記事推薦に対する、比較評価結果を示す。協業相手を考慮しない業務内容ベースの推薦(業務内容)、業務内容を考慮しない社員間のつながりによる推薦(つながり)、業務単位の協業度による推薦(提案手法)、の適合率を比較した。従来手法に比べて、業務単位の協業度を用いた手法では推薦精度の大きな向上が見られた。また図4には、提案手法で関連度が一定値範囲の記事を推薦した評価結果を示

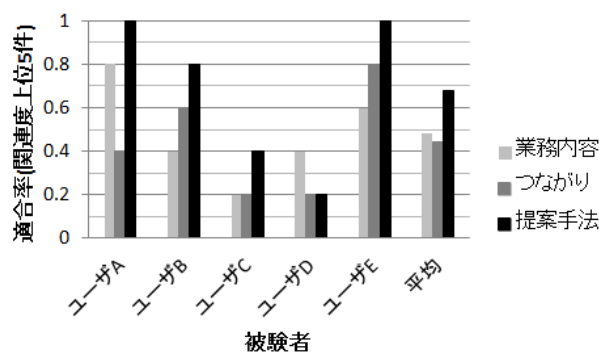


図3 推薦手法による比較

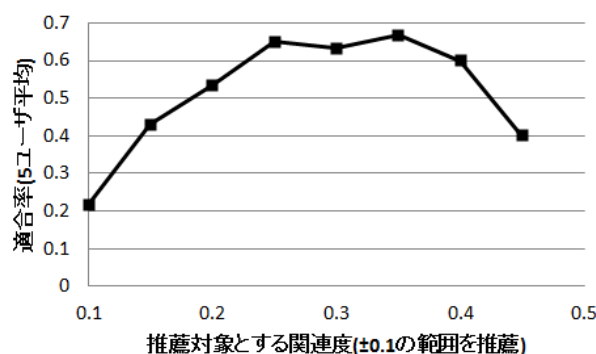


図4 関連度と適合率

す。推薦対象とする関連度の範囲は、提案システムにおいて任意に設定できる。関連度が低すぎても高すぎても適合率は低下しており、関連度 0.3 付近の適切な範囲の記事を推薦すべきであることが分かる。

4. 考察

業務単位協業度を用いたブログ記事推薦は、従来手法に比べて社内の業務連携支援に有効であることが分かった。今後はこのシステムの社内活用を進めていく予定である。

業務内容の抽出にクエリ拡張やトピックモデルなどを適用することで、より一層の性能向上が期待できる。また推薦対象として適切な関連度の範囲はユーザーによって異なるため、適合性フィードバックなどによる、推薦記事閲覧履歴からの学習も有効と考えられる。

参考文献

- [Ricci 11] Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor (Eds.): Recommender Systems Handbook. Springer (2011)
- [Guy 11] Ido Guy, David Carmel: Social recommender systems. WWW (Companion Volume) 2011:283-284
- [Tang 13] Jiliang Tang, Xia Hu, Huan Liu: Social recommendation: a review. Social Netw. Analys. Mining 3(4): 1113-1133 (2013)
- [Kudo 04] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis. EMNLP 2004:230-237