

平成 24 年度
修 士 論 文

セレンディピティな情報推薦システム

Serendipity Oriented Recommender System

田中 純

神戸情報大学院大学
情報技術研究科 情報システム専攻

内容梗概

近年、情報化技術の進展により、個人や団体が容易に情報発信を行えるようになった。また、計算機の記憶媒体の大規模化や通信の高速化により、多くの人が大量に発信された情報を取得できる状況が生じた。しかし、扱う情報量が個人の情報処理能力を圧倒しており、必要な情報が埋もれてしまうという状況が生じている。この状況を打破する為、膨大な情報の中から利用者の嗜好に合った情報を提供する情報推薦手法の研究が盛んに行われている。

初期の情報推薦の仕組みとして、利用者の過去の行動履歴からなる学習データを分析し、利用者の嗜好を推定の上、それらの情報を推薦するといった研究がなされていた。一方、こういった学習データに基づく推薦は、精度を重視しすぎる余りに、似通った情報を推薦する傾向がある。結果として利用者はすぐに推薦結果に飽きてしまい、推薦を受けなくなってしまう場合がある。

これらの問題を解決するために、近年では従来から重視されてきた推薦精度に加えて、セレンディピティなどの推薦精度以外の評価指標により推薦システムを評価する事が重要であると指摘されている。筆者が作成したプロトタイプシステムでは、人手で Web コンテンツに付加したタグを解析する事でセレンディピティな情報推薦を試みた。しかし、タグの付与を手動で行う場合、(1) 付加する人によって付けられるタグの傾向が変わる可能性がある (2) 大規模データへの対応が困難である、といった課題が顕在化した。

そこで本研究では、これらの問題を解決する事を目的として、システムによって自動的に Web コンテンツに付与したカテゴリ情報(タグ)と閲覧履歴を使用するセレンディピティな情報推薦システムを提案する。提案システムでは出現頻度と接続頻度に基づく手法によって、一様な尺度で自動的にタグ付けを行っている為、タグ付けの尺度が変化する事はない。前述したプロトタイプシステムの課題を克服することが期待できる。

本システムの有効性を検証する為に、システムの推薦する情報を実際に利用してもらい、アンケート形式での評価を行った。その結果、提案システムがセレンディピティをもたらした事を示唆する結果が得られ、提案システムの基本コンセプトの裏付けをとることができた。しかし、同時に推薦精度とセレンディピティの両立や、処理速度など、実装面に関してはまだ課題が多く残されている事がわかった。

目次

1. はじめに	1
1.1. 研究の背景	1
1.2. 本研究の目的	2
2. 関連研究	3
2.1. 情報推薦手法に関する関連研究	3
2.1.1. 協調フィルタリング	3
2.1.2. 内容に基づくフィルタリング	4
2.2. セレンディピティな情報推薦	4
2.3. 本論文の研究的位置づけ	5
3. プロトタイプシステムの設計	6
3.1. プロトタイプシステムの要件	6
3.2. プロトタイプシステムの概要	6
3.2.1. RSS 受信・解析機能	7
3.2.2. 記事閲覧機能	8
3.2.3. 閲覧履歴登録機能	8
3.2.4. 「隠れた興味」解析機能	8
3.2.5. 記事推薦機能	10
4. プロトタイプシステムの実装	11
4.1. 開発環境	11
4.2. プロトタイプシステムの利用例	12
4.3. プロトタイプシステムの推薦結果	13
5. タグの自動付与による改善	18
5.1. 人手によるタグ付けの問題点	18
5.2. タグの自動付与	19
5.3. 従来手法	19
5.4. 従来手法における課題	20
5.5. 実現方法	20
6. 提案システムの設計と実装	22
6.1. 提案システムの要件	22
6.2. 提案システムの概要	22
6.3. 追加機能	23

6.4. 改善機能	25
7. 評価実験	28
7.1. 提案システムの要件	28
7.2. 推薦リストの評価	28
7.3. 考察	29
8. まとめと今後の課題	34
謝辞	35
参考文献	36
付録1：第2回テキストマイニングシンポジウム 論文（2012.8）	38
付録2：修了発表会 発表資料	42