

ユーザ興味の推定において概念辞書を利用した新着 RSS 記事の推薦

瀬尾 崇一郎

ブログやニュースサイトといった日々更新される Web サイトの多くは、RDF Site Summary (以下、RSS) というフォーマットで、タイトルや記事の要約といったサイトの更新情報をまとめて配信している。このようなサービスを RSS フィード (以下、フィード) といい、このフィードを RSS リーダというソフトウェアに登録することにより、複数のサイトの記事をまとめて閲覧できる。しかし、登録フィード数が増えると取得する記事が多くなり、ユーザが自分にとって興味のある記事を探し出し、閲覧することの手間が大きくなってしまふ。昨年度阪口研究室の浅野は、多量の新着 RSS 記事の内容から、記事に対するユーザの興味の度合いを推定し、それが高い記事から優先して表示するシステムを開発したが、推定能力の不足から十分な推薦ができていなかった。

本研究は、このユーザ興味の推定において概念辞書を利用することで、より有効な新着 RSS 記事の推薦を行うことを目的とする。新着 RSS 記事のタイトルと内容から単語を抽出することに加え、概念辞書を利用することでその単語の概念と関係を持つ概念を探索し取得する。それら語と概念の一つ一つに重みを付け、記事と関連付けて保存する。記事中の語と概念の重みの算出には RTF-GIDF を用い、概念探索により取得した概念には、探索元となった概念の重みから、概念関係の遠さに応じた係数をかけて重みとする。ユーザの興味を得る手がかりとしてユーザの操作から、記事に対して未読、閲覧、無視、有用もしくは不要、といった評価を取得し保存する。この評価から記事に対するユーザの興味の有無を判別し、興味がある記事の語と概念の重みは正の値、興味がない記事の重みは負の値とし、語や概念ごとに重みを積算した値を求める。この積算値と、その語や概念が出現する記事に対する閲覧記事の割合との積を取り、これをその語や概念に対するユーザの興味度とする。そして各記事について、ユーザの興味度が高い 5 つの語または概念の総和を取り、これが高いものからユーザに優先して提示する。

新着 RSS 記事推薦の有効性を確認するために、被験者に本研究で開発した記事推薦システムを利用してもらい、ユーザにとって興味のある記事がどれだけ優先して表示されるかを評価した。実験では大学生 4 名に対し、推薦システムを用いて 341 件の記事の閲覧を行ってもらった。その結果、ユーザが興味を持つ記事を、4 割~6 割の精度で推薦することができた。しかし本研究のシステムでは、概念辞書を使わずに推定した場合でも同等の推薦ができていたため、推定能力は概念辞書によっては必ずしも向上しなかったことがわかった。この原因としては、記事に対する注目度の計算手法において、記事中の概念から取得した関係概念を、記事中に出現する語や概念と同じ様に扱っている点に問題があると考えられる。概念辞書によって取得した概念を、より正確に記事への注目度に反映する手法の開発が、残された課題と考えられる。

(指導教員 阪口 哲男)