

小特集

ソーシャルブックマークは進化し続ける!

～ソーシャルブックマーク研究会の議論から

5



私がチャレンジした SBM データマイニング

神林 亮 ▶ 筑波大学大学院システム情報工学研究科

「こんなサービスがあればもっと便利になるのに」と思うことは誰もが経験すると思うが、筆者の場合は、SBM サービスの利用の中でそのような経験があった。

本稿では筆者がそのような自身の要望に応えるために開発したサービスを紹介する。紹介するサービスはデータマイニングの手法を SBM サービス、具体的には、「(株)はてな」が提供する、はてなブックマーク¹⁾より得られるデータへ適用したものである。

データマイニングの題材としての SBM

SBM では、サービスの利用者がブックマークしたページ、設定したタグ、またユーザ間の関係などの情報が一部の例外を除き公開されており、その量も豊富である。データマイニングの題材として使えるような情報のうち、このように多様な情報からなり、かつ量が豊富で誰でもアクセスできるものは筆者の知る限りほかにない。したがって、SBM はデータマイニングの手法の適用に優れた題材であるといえる。

開発したサービスの概要

▶ Web ページの推薦

ブラウザの機能としての“ブックマーク”は、継続的にアクセスするサイトを保持しておくことを主たる目的として利用されるが、SBM の“ブックマーク”はニュースサイトやブログにおいて新しく発信される情報の記録を目的に利用されることも多い。このため SBM で新しくブックマークされたページを追いかけることで、Web 上で新しく発信される情報を効率よく収集できる可能性がある。しかしながら SBM では日々膨大な数のページがブックマークされるため、それらのページから、ユーザが自身で必要な情報を見つけることは困難である。

そこで、データマイニング技術を用いてフィルタリングを行い、ユーザの嗜好に合ったページを推薦するサー



図-1 Kikker の推薦ページの例

ビス Kikker²⁾を開発した。

Kikker で利用した手法はベクトル間の類似度計算であり、ユーザの嗜好を表現するベクタとページの内容を表現するベクタ間の類似度(コサイン距離)をページの評価値とし、推薦を行った。

ページの推薦にはユーザの嗜好情報が必要である。Kikker は嗜好情報を以下の手順で取得する。まずユーザが、はてなブックマークでブックマークしたページをクローリングし、ユーザの嗜好情報の初期情報とする。その後は Kikker が推薦したページをユーザが閲覧するたびに、閲覧したページの内容をユーザの嗜好として Kikker が保持する情報に反映し、推薦の質を向上させる。これによりユーザの Kikker 利用が進むにつれ、ユーザはより自身の嗜好にあったページの情報を得ることが可能となる。

Kikker が推薦を行った結果を図-1 に示す。図-1 ではユーザ ryo に対して推薦が行われており、推薦の程度を表すポイントや、はてなブックマークでのブックマーク数、また抽出されたキーワードなどとともにページへの

がある。したがって、個人であっても企業が提供するようなサービスと同等規模の計算量を必要とするサービスを提供できる可能性がある。

まとめ

本稿では SBM を題材とし筆者が開発した 3 つのサービスを紹介した。いずれも基本的な手法のみを利用しており、マイニング結果の質は高くないが、筆者や筆者と同様の要求を持つユーザにとっては、ある程度“使える”サービスとなったのではないかと考えている。

読者の皆様も、SBM をもっと便利にするサービスのアイデアがあれば、実際に開発を行い公開してみたいかがだろうか。きっと多くの人(筆者を含む)に感謝されるのではないと思う。

参考文献

- 1) <http://b.hatena.ne.jp/>
- 2) 現在は停止中である。
- 3) http://www.coins.tsukuba.ac.jp/~i041184/distribute/hatebu_mawarinohito/hatebu_mawarinohito.html
- 4) Clauset, A. : Finding Local Community Structure in Networks, Physical Review E, Vol.72, 026132 (2005).
- 5) <http://code.google.com/intl/ja/appengine/>
- 6) <http://aws.amazon.com/ec2/>

(平成 20 年 10 月 28 日受付)

▶▶ 神林 亮 (正会員)

kanbayashi@hpcs.cs.tsukuba.ac.jp

筑波大学大学院システム情報工学研究科博士前期課程 1 年次在学中。大規模並列分散システムに関する研究に興味を持つ。



図 -3 kookle のディレクトリの例

提供可能なサービス規模について

紹介したサービスは自宅の PC1 台を用いて運用が行われていたが、計算量の多いデータマイニング手法の利用や、利用者数の増加があれば、必要となる計算処理の提供は困難になると予想される。しかしながら、複数台のマシンを用いての対応も、調達のための資金や、設置するためのスペース、また消費電力などの問題により個人では難しい。

このような問題は Google App Engine⁵⁾ や Amazon EC2⁶⁾ のようなコンピューティングクラウドを提供するサービスの利用や、PS3 に搭載された Cell プロセッサの活用、また GPGPU の採用などにより解決できる可能性

