

多様性と意外性を考慮したリコメンドエンジン

—ユーザ視線リコメンドモデルの開発—

1. 背景

近年, Web において個人が情報を気軽に発信できるようになり, また個々人は Web を経由して簡単につながれるようになった. これにより, Web の情報量はかつてない勢いで増加している. このいわゆるソーシャルな時代における主要な情報探索手段の一つにフィードという手法がある. これはお気に入りのユーザやサイトをフォローして, その更新を取得するという手法である. しかしこの単純な手法ではまかないきれないほど情報の量は増え, ユーザが情報を探すストレスは増加している. そこでフィードに変わる新たな情報探索手段が求められており, その一つにリコメンドエンジンによる情報フィルタリングがある.

クリエイターらは, 2011 年からユーザの SNS のデータを分析し, そのユーザに合った記事を推薦する Web サービス「Gunosy」を開発・運営していた. Gunosy のリコメンドエンジンには, ユーザの興味を当てることはできていたが, それゆえにユーザにとって既視感のある記事や, 新鮮味が無く偏った内容の記事が推薦されるという問題があった.

そもそも, 従来のリコメンドエンジンは「ユーザが欲しい情報」を「ユーザが興味を持つ情報」と定義しており, 推薦の評価もユーザの興味にいかにより一致するかという方向性で発展してきた. しかしリコメンドエンジンがフィードに変わり主要な情報探索手段となるには興味により一致しているだけでは足りず, 推薦結果全体として多様性や意外性をもつ必要がある. 興味の一一致にのみ主眼をおくと情報の偏りが起きてしまうためである.

2. 目的

本プロジェクトは, 多様性と意外性を考慮したリコメンドエンジンを開発することを目的とする. まず, プロジェクト開始以前に開発した, Gunosy で利用しているリコメンドエンジンにおけるユーザの行動を分析し, 解決すべき問題の特定をする. 問題を解決する仮説を立て, それに基づいたアルゴリズムを開発する. 新アルゴリズムを実サービスで運用し, 効果を検証する.

3. 開発の内容

まず, プロジェクト以前に開発したリコメンドエンジンの問題の特定を目的に, ユーザの行動を分析した. 推薦記事リストのランダムな位置の記事に普通の記事挿入し, クリック率を比較した. 結果が図 1 である. なお, ここで普通の記事とは, リコメンドエンジンとは無関係に選択した記事であることを意味する. リスト内上位の記事は, リコメンドエンジンによる記

事の方がクリック率が高いが、リスト内中位で同程度、リスト内下位の記事では普通の記事の方がクリック率が高いことが分かる。プロジェクト開始以前のリコメンドエンジンは、中位以下の記事の推薦には機能しておらず、ここを改善させる余地があると考える。

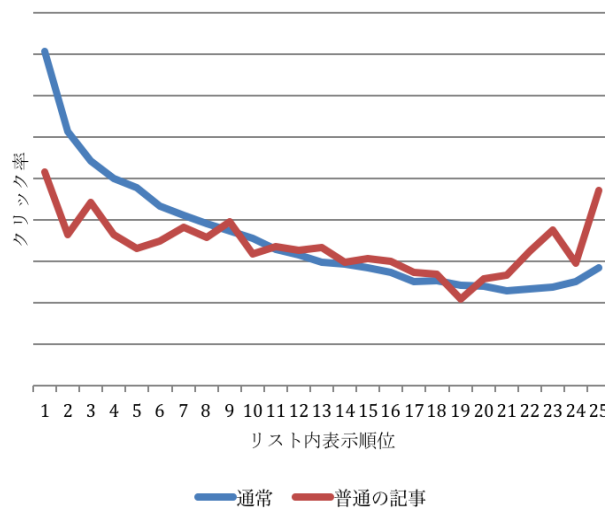


図 1 アルゴリズムにより推薦された記事と普通の記事のクリック率の比較

(「通常」はアルゴリズムにより推薦された記事、「普通の記事」はリコメンドエンジンとは無関係の記事を表す。なお、クリック率の数値は省略してある。)

次に、推薦位置の影響を分析するために、推薦リストをスコアとは逆順に表示する実験を行った。結果が図 2 である。上位ではアルゴリズムのスコア順の表示の方がクリック率が高く、下位ではアルゴリズムのスコア順とは逆順に表示した方がクリック率が高くなっている。このことから、推薦位置の影響を排除しても、プロジェクト以前のアルゴリズムは機能していることが分かる。また、下位にスコアの高い記事を表示させ、クリック率が向上したということは、下位の記事を推薦するアルゴリズムを変更することによって、クリック率を向上させることが可能であることを示していると考える。

通常推薦システムは、推薦記事リストを出力する前に、ユーザのプロフィールを定義し、そのプロフィールと記事のデータをもとに推薦リストを出力する。すなわち、固定されたユーザプロフィールから推薦リストを出力している。しかし、これは実際のユーザの推薦リストの利用にそぐわない。ユーザは通常、推薦リストにある全記事を平等な基準で見ない。PC であろうがスマートフォンであろうがユーザは推薦リストの上から順に記事を見る。そのため、推薦リストを閲覧していく中で、ユーザの興味は変化していくと考えた。この時、上位の記事で出現したトピックに対する興味は満たされ、下位の記事を閲覧する際には、そのトピックに対する興味は減少していると考えた。そこで、例えばリストの n 番目の記事は、ユーザがすでにその上にある $n-1$ 番目までの記事を読んだことを前提にし、ユーザプロフィールを再

定義することが自然であると考えた。そこで、新しいアルゴリズムでは、 n 番目記事を推薦する際、 $n-1$ 番目までに登場したトピックの興味スコアを減衰させるという方法で推薦システムを構築した。

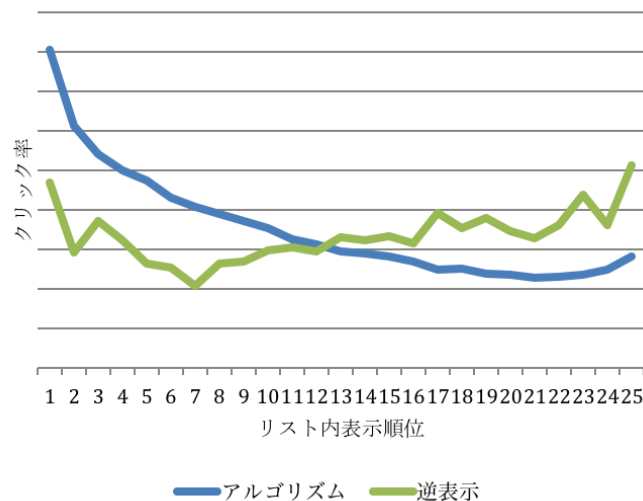


図 2 アルゴリズムをスコア順に並べた推薦リストとスコアと逆順に並べた推薦リストのクリック率の比較(クリック率の数値は省略してある.)

開発したアルゴリズムを Gunosy のユーザに適用し、その効果を検証した。適用前後で、ユーザのアクティブ率の傾向に変化が見られる。新規登録 1 ヶ月後で Gunosy を利用し続けているユーザの割合が、プロジェクト開始時点で、72%前後だったのが、アルゴリズム適用の直前では、68%前後まで下降していた。それが、新アルゴリズム適用後は、約 1 ヶ月半でユーザ数が約 2 倍に成長したにもかかわらず、ユーザの 1 ヶ月アクティブ率は当初の 72%前後まで上昇していた。新アルゴリズム適用以前は、約 3 倍に成長する過程で、ユーザの 1 ヶ月アクティブ率が約 4%減少していたことを考慮すると、この結果は新アルゴリズムの有効性を示すものであると考える。また、1 週間当たりの使用日数も、3.06 日から 3.31 日に 0.25 日、8.2%増加しており、新アルゴリズムによりユーザのサービスへのエンゲージメントを高めることも出来た。登録後 3 週間目のユーザでは、12 位以下の記事で、アルゴリズム変更後のクリック率が大きくなっていることが分かる。長期的にサービスを利用しているユーザに、新アルゴリズムが有効に働いていることが分かる。

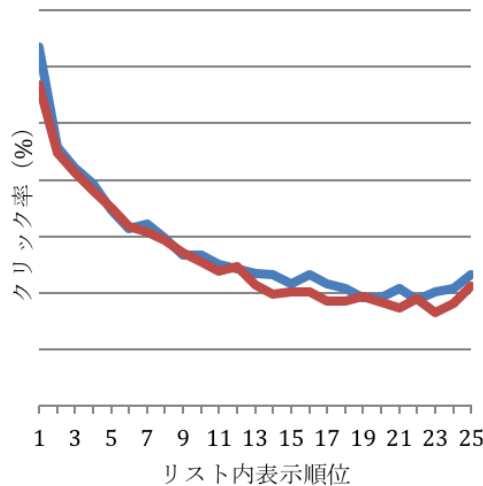


図 3 登録後 3 週間目のユーザのクリック率
 (赤がプロジェクト以前のシステム, 青がプロジェクトで開発したシステム.
 クリック率の数値は省略してある.)

4. 従来の技術(または機能)との相違

従来のリコメンドエンジンでは、固定されたユーザプロフィールから推薦リストを作成しているが、本プロジェクトで開発したリコメンドエンジンでは、 n 番目記事を推薦する際、 $n-1$ 番目までに登場したトピックの興味スコアを減衰させ、推薦リストを作成している。本プロジェクトで開発したリコメンドエンジンは、一度に推薦記事リストを作成するのと違い、興味スコアの低い記事でも推薦リストに含めることが出来る。このことによって、推薦記事リストの多様性が向上され、ユーザにとって意外と感じる記事が含まれる確率も上昇する。

5. 期待される効果

本プロジェクトで開発したリコメンドエンジンにより、ユーザは楽に情報収集を行えるだけでなく、新しい発見も得ることが出来るようになる。Web 上の情報を用いたデータマイニングは、学術的にだけでなく、多くの企業が取り組むなど、産業的にも重要視されている。本プロジェクトの成果が広まることは、データマイニングを実サービスに利用した例としても重要であると考える。

6. 普及(または活用)の見通し

ユーザ数は 2012 年 4 月時点では約 5000 人、2012 年 7 月時点で約 1 万 50000 人、2013 年 1 月時点で約 5 万人と成長した。その後も、ユーザ数は順調に増加し、本報告書執筆時の 2 月 3 日では 7 万人を超えている。2013 年内に 100 万ユーザを目標としている。

7. クリエータ名(所属)

チーフクリエイター: 福島良典(東京大学大学院工学研究科)

コクリエイター: 吉田宏司(東京大学大学院工学研究科)

Web サイト: Gunosy <http://gunosy.com>