

Pinterest における Pin 推薦のための 推薦アルゴリズムの比較

足立 健太郎^{1,a)} 土方 嘉徳^{1,b)} Joseph A. Konstan^{2,c)}

概要: Pinterest は人々をつなぐソーシャルネットワーキングの機能と既存のコンテンツをまとめて新たなコンテンツを作成するキュレーションの機能を併せ持つサービスである。Pinterest には 2015 年 3 月末の時点で 500 億の画像が存在し、ユーザが自らの意図に合う画像を適切に見つけるためには画像推薦の機能が必要である。現在の主要な推薦アルゴリズムには、コンテンツベースフィルタリング、協調フィルタリング、社会ネットワークに基づく推薦、人気度に基づく推薦があるが、それぞれ異なる種類の情報を用いている。Pinterest は、上記のアルゴリズムで用いられる情報を全て有しているため、任意のアルゴリズムを適用できる。さらに、Pinterest のユーザは自分のためだけでなく、他人に見せるために特定のテーマに沿ったコレクションを作成することもあり、従来の推薦システムの想定とは異なるモチベーションを持っていると言える。上記のように新しいタイプのサービスである Pinterest を対象に、どの推薦アルゴリズムが有効に働くのかを知るため、主要な推薦アルゴリズムを正確性と利便性の観点から評価を行なった。

1. はじめに

近年、ソーシャルネットワーキングサービスの機能とキュレーションサービスの機能を併せ持つ、画像を対象としたサービスである Pinterest が人気を集めている。Pinterest では、ユーザは他のユーザが投稿した画像にコメント・共有・コメントすることができ、また他のユーザやコレクションをフォローすることが出来る。これらはソーシャルネットワーキングサービスの機能である。また、ユーザは自らの嗜好や意図、特定のテーマに基づいてコレクションを自由に作成し、そこに画像を分類して収集することが出来る。これらはキュレーションサービスの機能である。

Pinterest には 2015 年 3 月末の時点で 500 億もの画像が存在 ^{*1} するため、ユーザは自らのコレクションに合う画像を探すことが困難である。そのため、推薦システムによる画像の推薦が必要である。現在、推薦システムを実現するためのアルゴリズムはいくつか種類があるが、その種類により用いる情報の種類も異なっている。主要なアルゴリズムには、推薦するアイテムのコンテンツ情報を用いるコ

ンテンツベースフィルタリング、複数のユーザのアイテムに対する評価値を用いる協調フィルタリング、ユーザ同士のフォロー関係を用いる社会ネットワークに基づく推薦、アイテムの人気度を用いる人気度に基づく推薦がある。Pinterest は、これらのアルゴリズムが必要とする情報をすべて有している。従って、任意のアルゴリズムを適用可能である。そのため、どの推薦手法が Pinterest における画像推薦に有用なのかを知ることは重要である。

従来の推薦システムが想定しているユーザは、自らの趣味・嗜好に基づいてアイテムを消費する人たちである。しかし、Pinterest においては、ユーザは他人に見せるために、特定のテーマに合う画像のみからなるコレクションを作成することがある。また、従来の推薦システムが想定していたように、自らの趣味・嗜好に基づいてアイテムを集めるユーザも存在する。そのため、異なる目的を持つユーザが混在していると言える。このようなサービスにおいて、どの推薦アルゴリズムが有効であるかを知る必要がある。また、ユーザの利用目的が異なるので、正確性に関する指標だけでなく、多様性、新規性、セレンディピティといった利便性に関する指標でも評価を行い、考察を行う必要がある。

本研究は、ソーシャルネットワーキングサービスとキュレーションサービスの特徴を併せ持つ新しいサービスである Pinterest での画像推薦において、代表的な推薦アルゴリズムの性能を複数の観点から評価し、比較するものである。

¹ 大阪大学大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University

² ミネソタ大学 GroupLens Research
University of Minnesota, GroupLens Research

a) adachi@nishilab.sys.es.osaka-u.ac.jp

b) hijikata@sys.es.osaka-u.ac.jp

c) konstan@umn.edu

*1 <https://engineering.pinterest.com/blog/introducing-pinterest-developers-platform>

本研究の貢献は以下の2点である。

- 異なる目的を持つユーザが存在し、複数の種類の情報を推薦に利用可能なPinterestに対して、コンテンツベースフィルタリング、協調フィルタリング、社会ネットワークに基づく推薦、人気度に基づく推薦という、主要な推薦アルゴリズムの比較を初めて行った。
- 推薦結果を正確性、多様性、新規性、セレンディピティという主要な評価指標で評価し、各手法の利点・欠点を示した。

本稿の構成は以下の通りである。2章で関連研究を述べ、3章で作成したデータセットと比較するアルゴリズム、評価で用いる指標の説明を行う。4章でアルゴリズムの比較実験の結果を示し、5章で結果に対する考察を行う。最後に6章でまとめを述べる。

2. 関連研究

ここでは、Pinterestについて調査を行っている研究と、推薦アルゴリズムの比較調査を行っている研究について紹介する。

2.1 Pinterestに関する研究

Pinterestは2010年に開始された比較的新しいサービスである。Pinterestに関する研究は、サービスの初期段階における利用状況に関する調査が多くを占めている。Gilbertらは、Pinterestのユーザの性別に注目した調査を行い、アメリカではPinterestユーザの殆どが女性であり、男性はより多くのフォロワを持つ傾向があると報告している[9]。また、Pinterestの利用のされ方についても調査されている[10][20][15]。主に、飲食や装飾、芸術、ファッションなどの日常的な使われ方がされている一方で[10][20]、警察による指名手配のために使われる事例も報告されている[15]。また、MillerらはPinterestでのユーザ登録時に他のサービスでのフォロー関係をインポートするソーシャルブートストラッピングに注目し、男性と女性でサービスへの満足度に差があるかどうかを調べている[17]。

近年では、Pinterestにおける推薦システムも開発されている。例えば、KamathはPinterestにおいて、カテゴリ・ピンの説明・リピン数・いいね数を特徴としたベクトル空間モデルを作成してボードの推薦を行っている[13]。また、ある画像が提示された時に、それをリピンするかどうかや、それをどのボードに入れるかを予測する研究[22]もなされている。

2.2 アルゴリズムの比較研究

Huangらはいくつかの推薦アルゴリズムについて、服・本・映画という3つのジャンルのデータセットを用いて精度、再現率、F値、ランクスコア[3]、ROC曲線の指標で評価を行った[12]。比較したアルゴリズムは3種類の協調

フィルタリング、潜在因子を用いたモデルによる推薦、活性化拡散モデルによる推薦、リンク解析による推薦、人気度に基づく推薦の7種類である。Cremonesiらは、評価平均が高いものを推薦する方式、評価数が多いものを推薦する方式、評価したアイテムが似ている近傍ユーザとの相関を用いた推薦方式、コサイン類似度を用いた推薦方式、SVDによる推薦とSVDから派生した方式の6種類を比較した[7]。評価にはMovieLens[11]とNetflix[2]の2種類のデータセットが用いられた。

CachedaらはMovieLensとNetflixを用いて10種類の協調フィルタリングを比較し、さらに精度に関する評価指標を2つ提案している[4]。BelloginらはMovieLensのデータセットを用いてユーザベース協調フィルタリング、アイテムベース協調フィルタリング、SVDを用いた協調フィルタリングの3つを、精度、再現率、nDCG、平均二乗誤差という正確性に関する指標を用いて比較した[1]。Kluverらは推薦システムを利用したばかりのユーザに推薦するためのアルゴリズムの比較を行った[14]。評価平均が高いものを推薦する方式、ユーザの評価平均で調整した評価平均が高いものを推薦する方式、アイテムベース協調フィルタリング、ユーザベース協調フィルタリング、SVDを比較した。評価指標には過去の研究から11の指標を用いており、幅広い考察が行われている。

我々の研究においてはPinterestに対するコンテンツベースフィルタリング、協調フィルタリング、社会ネットワークに基づく推薦、人気度に基づく推薦を比較している。評価指標には、正確性だけでなく、多様性、意外性、新規性も用いている。

3. 比較実験の方法

ここでは、Pinterestの紹介を行い、実験に用いるデータのクローリング方法と比較するアルゴリズム、比較に用いる評価指標について説明する。

3.1 Pinterestについて

Pinterestは画像に関するピンボード形式のソーシャルキュレーションサービスである。ユーザは、興味の似た人々とのコミュニケーションを通して、自分の趣味や嗜好に合う画像を発見することができる。具体的には、ユーザは、他のメディアで発見した画像や自分で撮影した画像を“ボード”と呼ばれる自分のコレクションに入れて公開できる。Pinterestにアップロードした画像は“ピン”と呼ばれる。また、Pinterestで見つけた他人のピンをボードに追加できる。この行動は“リピン”と呼ばれている。Pinterestにおけるユーザ行動のほとんどがリピンであるという報告が存在する[10]。また、自分の興味のあるものや好きなものに関する画像を多くピンするユーザをフォローすることで、自らのホームフィードを興味のあるものや好

きなものの画像で埋めることができる。ここで、ユーザのフォロー関係の説明をする際の用語としてフォロウィとフォローを定義する。本研究ではフォロウィとはアクティブユーザ（対象ユーザ）がフォローしているユーザのこととし、フォローとはアクティブユーザをフォローしているユーザのこととする。

3.2 データについて

Pinterest は我々の調査に必要なデータセットを作成するために有効な API を提供していない。そのため、我々はクロールングのプログラムを実装した。

3.2.1 クローリング手法

我々のクローリング手法は Chang らの手法 [6] を参考にし、API がなくても推薦に必要な各種の情報を取得できるようにしたものである。我々のクローリング手法は、まずシードユーザを決定し、そこからフォロウィかフォローをランダムに深さ優先探索でたどることでユーザを収集するものである。次に、取得したユーザ ID をもとに、各ユーザのボードを収集する。そしてユーザの各ボードに含まれる全てのピンを収集する。

シードユーザは、Pinterest の最新のピンというページの上位にあるピンを投稿したユーザ 10 人とする。シードユーザの中からランダムに 1 人を選択する。そのユーザのフォロウィかフォローの中からランダムに 1 人のユーザを選択し、ユーザ ID を取得する。この手順を繰り返すことで深さ優先探索によるユーザ探索が実現される。ここでユーザ ID とは、Pinterest のユーザに与えられる、文字列で表現される ID である。Pinterest の URL にユーザ ID を追加することで ([https://www.pinterest.com/\[ユーザ ID\]/](https://www.pinterest.com/[ユーザ ID]/))、そのユーザのボード一覧が取得できる。さらに、ボード名を追加することで ([https://www.pinterest.com/\[ユーザ ID\]/\[ボード名\]/](https://www.pinterest.com/[ユーザ ID]/[ボード名]/))、そのボードに含まれる画像一覧が取得できる。

なお、上記のデータの取得は API により実現できるものではないため、Web ブラウザを通してアクセスすることにより実現する。具体的には、クローリングのプログラムから、Google Chrome と Mozilla Firefox のスクロール操作を制御し、順に Web ブラウザに表示される Web ページを拡張する。

全てのデータが Web ブラウザに表示されれば、その時にブラウザに表示されているコンテンツを、DOM (Document Object Model) の API にて取得する。ただしピンに関しては、スクロールに従ってダウンロードされると同時に上部のデータが削除されるため、ルールベースで決定したスクロール回数ごとに HTML を保存した。

必要なユーザ ID やボード ID、ピン ID、またこれらに関連する情報（ボード名やカテゴリ名など）は保存した HTML より情報抽出の手法により抽出した。なお、この抽

出には簡単なルールベースのプログラムを独自に開発し、各ユーザに対して 1 回ずつクロールングを行った。

データセットの収集は、2014 年 11 月 1 日から 2015 年 4 月 30 日の 6 ヶ月の期間に、13 台の計算機を用い、20 個のノードの並列処理により行った。

3.2.2 データセット

取得したデータの種類の表 1 に示すとおりである。最終的に取得したユーザは 200 人、そのうち 50 ユーザについては 2 ステップ先のフォロウィ 7,585 人とフォロー 6,909 人のボード・ピンも取得した。200 人が持つボード数は 7,674 で、それぞれのボード名とカテゴリを取得した。ボードのカテゴリはユーザがボードを作成する際に任意に設定するものであるため、46.82% のボードにはカテゴリが付いていなかった。取得したユーザ 200 人からピンを 1,161,373 個取得し、それぞれのリピン数・いいね数・ピンの説明も取得した。上記 200 人のユーザによるリピンは 2,549,036 個取得し、ユーザは平均して 33.8 個のボード、6,111.4 個のピンを所持していた。

表 1 取得したデータ

データ種類	関連するデータ
ユーザ	ユーザ ID, フォロウ, フォロウィ, ボード, ピン
ボード	ボード名, カテゴリ, ボードの説明文
ピン	ピン ID, ピンの説明文, リピン数

3.3 比較するアルゴリズム

ここでは、比較する 4 つのアルゴリズムの説明を行う。近年ではこれらの改良版やハイブリッドのものも発表されているが、本研究では各アルゴリズムの基本的な性質を確かめるため、それぞれ単体で実装したものの比較を行う。

3.3.1 コンテンツベースフィルタリング

コンテンツベースフィルタリングはアイテムのコンテンツ情報を用いてアイテムとユーザを特徴ベクトルで表現し、それらの類似度を計算することでユーザに合うアイテムを推薦するという方式である。この方式を Pinterest に適用するために、まず、ピンの説明文と属するボード名を用いて bag-of-words によりピンベクトルを作成する。次に、各ユーザが持つピンのピンベクトルの平均をとることで各ユーザのユーザベクトルを作成する。作成したユーザベクトルとピンのベクトルのコサイン類似度を計算し、類似度の高いベクトルで表されたピンを推薦する。

3.3.2 協調フィルタリング

協調フィルタリングは、ユーザベース [18] のものとアイテムベース [19] のものが存在する。どちらも最初にユーザーアイテム行列を作成する。行がユーザ、列がアイテムに

相当し、要素はユーザによるアイテムへの評価値である。Pinterest では、ユーザがピンをコレクションに入れたかどうかという 2 値である。各行がそれぞれユーザを表現するベクトルになる。このベクトルの距離を計算しアクティブユーザと近いベクトルとなるユーザのピンから推薦を行うのがユーザベース協調フィルタリングである。同じように、列で分割した場合は各列がそれぞれアイテムを表現するベクトルになる。このベクトルが類似しているアイテムを推薦するのがアイテムベース協調フィルタリングである。

一般的に、行列が疎である場合には次元削減アルゴリズムとして SVD や Matrix Factorization が使われる。本研究における行列もスパース率が 1.4% と、ある程度疎な行列と言えるため、SVD を用いて次元削減したものを用いる。さらに、評価値行列以外の情報として、ピンのカテゴリとボードに使われる単語をそれぞれ用いて次元削減したものを用いる。ピンのカテゴリを用いた場合は、アイテムの次元は 31 次元となった。ボードに使われる単語を用いた場合は、5,239 次元となった（ステミング処理後）。ただし、アイテムの次元を圧縮すると、行列の各列（カテゴリや単語）は多数のピンの情報を一般化したものになる。そのため、推薦すべき列が決定しただけではピンの推薦を行うことができない。そこで、以下の式のように人気度を用いて各列に適合する画像にスコア付けを行い、上位の画像を推薦する。

$$Score_i = CFscore_{factor(i)} \cdot popularity_i$$

$Score_i$ はあるアイテム i の推薦スコア、 $CFscore_{factor(i)}$ は協調フィルタリングによってアイテム i の特徴を表す要素についた値、 $popularity_i$ はあるアイテム i の人気度である。

3.3.3 社会ネットワークに基づく推薦

社会ネットワークに基づく推薦は、フォロー関係に基づくネットワークにおいて、アクティブユーザの近傍ユーザを用いてアイテムを推薦する方式である。近傍ユーザはアクティブユーザと近い嗜好を持っていると期待するものである。我々は、アクティブユーザから 2 人目以内で到達できるユーザのうち、アクティブユーザのホームフィードにはピンが表示されないユーザのみを推薦に用いる。また、そのようなユーザまでの探索パスを 3 つのパターンに分けて、それぞれを用いた方式を比較する。以下の説明では、Pinterest の用語を用いているが、これらの方式は多くの SNS でも利用可能なものである。

2 ステップ先のフォロウィのピンから推薦

2 ステップ先のフォロウィとは、アクティブユーザのフォロウィのフォロウィのことである。ユーザは通常、興味のあるユーザをフォローしていると思われ、2 ステップ先のユーザはアクティブユーザと似た嗜好を持つ可能性が高い。そこで彼らのピンを推薦する。この方式では、2 ス

テップ先のフォロフィはアクティブユーザのフォロウィからのフォロ数により重み付けされる。さらに、彼らのピンはリピン回数により重み付けされる。この 2 種類の重みを掛け合わせたものを基準として、上位のピンを推薦する。

フォロワのピンから推薦

ユーザは通常、興味のあるユーザをフォローしていると思われ、アクティブユーザのフォロワはアクティブユーザと似た嗜好を持つ可能性が高い。そこで彼らのピンを推薦する。この方式では、フォロワのピンがリピン回数により重み付けされる。この重みを基準として、上位のピンを推薦する。

同じユーザをフォローしているユーザのピンから推薦

前述の 2 つのパターン同様、アクティブユーザがフォローしているユーザをフォローしているユーザは、アクティブユーザと似た嗜好を持つ可能性が高いと考えられる。そこで彼らのピンを推薦する。この方式では、上記ユーザ群のピンがリピン回数で重み付けされる。この重みを基準として、上位のピンを推薦する。

3.3.4 人気度に基づく推薦

これは最も単純な推薦手法で、個別化されない推薦である。推薦するピンは、データセット内で人気なピンである。データセット内のリピン数の多い画像のうち、アクティブユーザがピンした画像以外が、リピン数の多い順に推薦される。

3.4 評価について

我々は 5 分割交差検定法により、作成したデータセットを学習用データセットとテスト用データセットに分割して評価を行う。また、Pinterest のデータはユーザが画像をコレクションに含めているかそうでないかを評価値として持つ “unary” なデータである。そこで、これに適用可能な指標を使用する。正解データとしてはリピンと Like が考えられるが、データセットにおいてリピンの数が圧倒的に多いため、リピンを正解データとして採用した。分類なしに保存する Like とは異なり、リピンはユーザが画像を対応するボードに追加する行動であるため、Pinterest 特有の行動であるとも考えている。以下の説明における “正解ピン” とは、データセット内でユーザがボードに含めていたピンのことである。

3.4.1 正確性

正確性の評価指標には以下の 4 つを使用する。

Hit-rate [8]

全ユーザのうち推薦リストに正解ピンが 1 つでも入っていたユーザの割合で、以下の式で計算される。

$$hit-rate = \frac{num_{hits}}{n}$$

num_{hits} は推薦リストに正解ピンが 1 つでも入っていたユーザの数、 n は全ユーザ数である。

Precision

上位 N 位の推薦リストにおいて、正解ピンの数の割合で、精度（適合率）とも呼ばれる。以下の式で計算される。

$$Precision = \frac{|T_i \cap L_i|}{|L_i|}$$

T_i は正解ピンの集合、 L_i は推薦リストに含まれるピンの集合である。

Mean Average Precision @ N

上位 N 位の推薦リストを 1 位から順に見て、正解ピンが出た順位における精度をすべて計算し、その平均を取ったもので、以下の式で計算される。

$$MAP = \frac{1}{m} \sum_n AP_n$$

AP_n は n 位までで正解ピンが出た順位における精度の平均、 m は推薦リスト内の正解ピン数である。

Recall @ N

上位 N 位の推薦リストにおいて、正解ピンのうち推薦されたピンの割合で、再現率とも呼ばれる。以下の式で計算される。

$$Recall = \frac{|T \cap L|}{|T|}$$

T は正解ピンの集合、 L は推薦リストに含まれるピンの集合である。Pinterest のような大規模なデータセットを持つサービスでは、知っていればコレクションに追加したであろうアイテムを正確性の評価に参入できないため、適合率（精度）よりも使われることが多い評価指標である。

3.4.2 多様性

リスト内類似度 [23]

推薦リスト内のピンの多様性についての指標である。ピンのベクトルを用いて推薦リスト内の各ピン間のコサイン距離を計算し、その平均を取ったものがリスト内類似度である。以下の式で計算される。リスト内類似度の値が低いほど多様性が高い推薦リストである。

$$ILS(L_i) = \frac{\sum_{b_k \in L_i} \sum_{b_e \in L_i, b_k \neq b_e} sim(b_k, b_e)}{NC_2}$$

N は推薦リストに含まれるピンの数、 b_k, b_e はピン、 $sim(b_k, b_e)$ は bag-of-words によるピンベクトルのコサイン類似度である。

3.4.3 新規性

ロングテールに基づく指標 [5]

これは単一の式で表すことが出来ないが以下の手法で定量化できるものである。横軸をリピン数に対する対数軸とし、縦軸にリピン数の累積度数を取ったグラフを描くと、それはロングテールの形になる。このグラフからピンをロングテールのヘッド、ミッド、テールに従って 3 等分し、推薦リスト内のアイテムペアが 3 等分したどこの部分にどれだけ分布しているかを調べる。ヘッド由来のアイテムペ

アが多ければ、そのシステムは人気のアイテムばかりを推薦しており、新規性のあるアイテムを推薦する能力が低いことになる。

3.4.4 意外性

Unserendipity [21]

ユーザにとって、その推薦がどれほど意外なものであったかを測る指標である。ピンのベクトルを用いて、ユーザがコレクションに含むピンのベクトルと、推薦リスト内のピンのベクトルの類似度を計算し、その平均が Unserendipity となる。この値が低いほど意外性が高い推薦リストである。

$$Unseren = \sum_{u \in U} \frac{1}{|U||H_u|} \sum_{h \in H_u} \sum_{i \in L_u} \frac{sim(i, h)}{|L_u|}$$

U はユーザ集合、 H_u はユーザ u が持つしたピン集合、 L_u はユーザ u への推薦リストで、 $sim(i, h)$ は bag-of-words によるピンベクトルのコサイン類似度である。

4. 各アルゴリズムにおける評価結果

ここでは、各アルゴリズムについて評価した結果を順に示す。正確性の指標をはじめ、全体的にかなり小さな値になっている。そのため、結果についてはこのスケールにおける数値の大小について述べる。以降の図表では、コンテンツベースフィルタリングを“content”，アイテムベース協調フィルタリングを“item CF”，ユーザベース協調フィルタリングを“user CF”，SVD を用いた協調フィルタリングを“SVD”，カテゴリを用いた協調フィルタリングを“category”，ボード名に含まれる単語を用いた協調フィルタリングを“words”，2 ステップ先のフォロウィからの推薦を“social 2step”，フォロワのピンからの推薦を“social follower”，同じユーザをフォローしているユーザのピンからの推薦を“social co-follow”，人気度に基づく推薦を“popularity”，データセットから無作為にアイテムを抽出したものを“random”と表記する。

4.1 正確性

図 1 に示すように、ヒット率はコンテンツベースフィルタリングが最も高い値であった。また、全ての手法がランダムに作成した結果よりも良いことがわかる。次に、図 2 に示す Precision@10, MAP@10, Recall@10 について述べる。これらに関しては、次元削減を行わなかった協調フィルタリングがうまく働かなかったことが顕著に出ており、ランダムに作成した結果と同等かそれ以下の結果になっている。全ての結果において、コンテンツベースフィルタリングが最も高い値で、精度と MAP に関しては次いで社会ネットワークに基づく推薦が高い値になっている。特に MAP に関しては、社会ネットワークに基づく推薦の中でも、アクティブユーザからのパスが 1 本であるフォロワのピンを推薦する手法が高い値を示している。再現率

関しては、SVDにより次元削減を行った協調フィルタリングが2番めに高い値になっている。さらに、図3にMAP@100をMAP@10とともに示す。値は少し小さくなるものの、各アルゴリズムの結果はMAP@10と同じような傾向である。多くのアイテムを推薦した場合でも、アルゴリズムの正確さに関する順番は変わらないことがわかる。

4.2 多様性

図4に示す多様性について述べる。使用した指標がリスト内類似度のため、値が高ければ多様性は低いことを示すことに注意する。結果は、ピンの属するボードに使用された単語を用いて次元削減した協調フィルタリングが最も高い値になり、ピンの属するボードについたカテゴリを用いて次元削減した協調フィルタリングが2番目に高い値となった。ボード名やピンのカテゴリは、多様性の計算に用いた類似度計算でも用いているため、これらのアルゴリズムとコンテンツベースフィルタリングが高い値になることは理解できるものである。ここで、次元削減を行っていない協調フィルタリングのリスト内類似度が高くなっている要因として、人気のある特定のカテゴリから人気なピンが推薦されたことが考えられる。これは、我々のデータにおいて13.3%のリピンが全体の1%のピンからなっており、リピン数の分布がかなり偏っていることから示唆されることである。それに対して、社会ネットワークに基づく推薦や人気度に基づく推薦の値は低く、多様性の高い推薦ができていていることがわかる。

4.3 新規性

次に、新規性について述べる。ロングテールに基づく指標では、推薦リスト内のアイテムペアがロングテールのグラフにおけるどの部分にどれくらい属しているかを測る。特に、共にヘッドにあるアイテムペアが多ければその推薦は新規性が低いと評価でき、反対にミッドやテールにあるアイテムが多く推薦できていれば新規性が高いという評価になる。我々のデータセットでは、文献[21]の方法に従って3分割したところ、ヘッドとミッドの境界はリピン数の多い順の105番目のアイテム、ミッドとテールの境界は11040番目のアイテムとなった。結果は表2に示すとおりである。スペースの都合上、傾向が似ていた協調フィルタリングはSVDのみ、社会ネットワークに基づく推薦は2ステップ先のフォロウィのみを示す。協調フィルタリングはどれも共にヘッドに存在するアイテムペアが多く、新規性は低い推薦であるといえる。コンテンツベースフィルタリングは共にヘッドに存在するアイテムペアは少なく、テールにあるアイテムも推薦できていることがわかる。社会ネットワークに基づく推薦もコンテンツベースフィルタリング同様にミッドとテールにあるアイテムを推薦できており、新規性が高い推薦である。

表2 ロングテールに基づく指標

手法	アイテム	ヘッド	ミッド	テール
content	ヘッド	6.34%	60.43%	33.23%
	ミッド	5.36%	53.06%	41.58%
	テール	3.65%	50.63%	45.72%
SVD	ヘッド	60.34%	29.63%	10.03%
	ミッド	16.93%	74.20%	8.87%
	テール	3.92%	39.63%	56.45%
2 step	ヘッド	10.34%	38.65%	51.01%
	ミッド	20.15%	58.32%	21.53%
	テール	2.90%	38.58%	58.52%
popularity	ヘッド	100%	0%	0%
	ミッド	0%	0%	0%
	テール	0%	0%	0%
random	ヘッド	4.67%	22.89%	72.44%
	ミッド	4.12%	25.19%	70.69%
	テール	5.01%	23.72%	71.27%

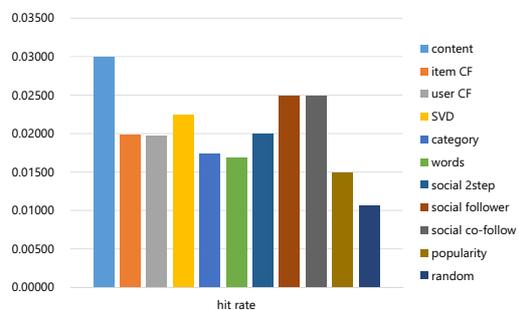


図1 ヒット率

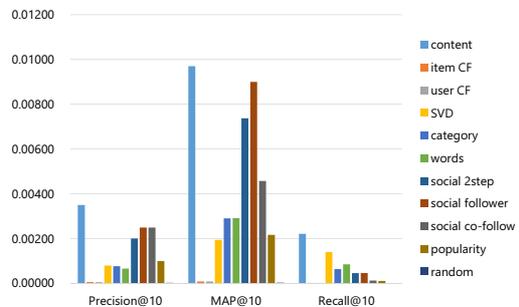


図2 Precision-MAP-Recall

4.4 セレンディピティ

最後に、図5に示すセレンディピティについて述べる。計算式からも分かるように、Unserendipityの値が高ければユーザのピンと推薦リストのピンの類似度が高く、セレンディピティが低いことに注意する。結果から、協調フィルタリングと人気度に基づく推薦がセレンディピティの高い推薦ができて、コンテンツベースフィルタリングと社会ネットワークに基づく推薦はセレンディピティの低い推薦になることがわかる。

5. 考察

正確性の指標をはじめ、用いた評価指標の全てにおいて

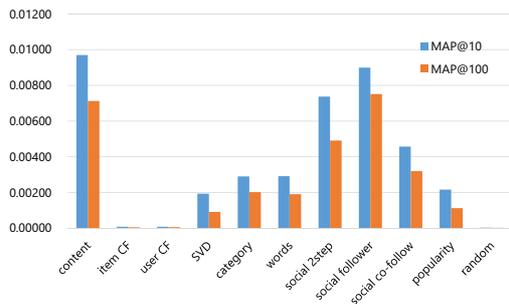


図 3 MAP@10 と MAP@100

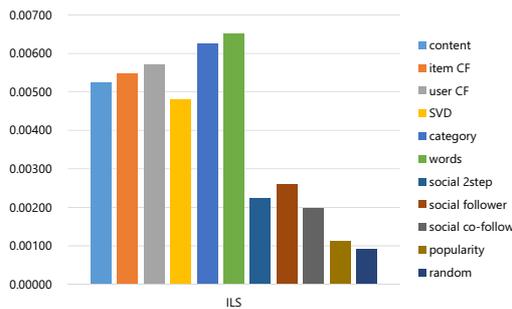


図 4 リスト内類似度

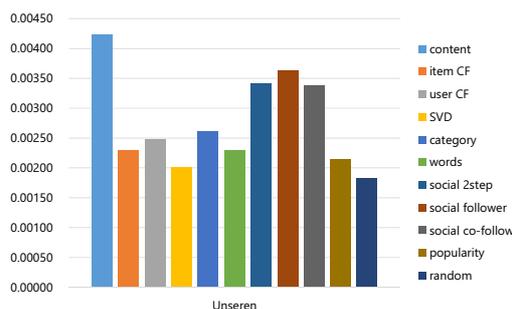


図 5 セレンディピティ

値の絶対値は小さい。つまり、少数のアイテムのみから構成される推薦リストを1回だけユーザに提示するだけではユーザは各推薦手法による結果の違いを知覚するのは難しいと言える。ただし、我々は実験をオフラインで行なったため、評価で正解とみなされる画像は、ユーザがすでにピンした画像のみである。しかし実際には、もしユーザが知ればボードに入れたいと思うが、知らないがためにピンできていない画像について全く考慮していない。従って、オンラインによる実用では少なくとも本研究よりも大きな値が保証されるものと考えてよい。これらを踏まえて、我々は推薦の提示方法を工夫することで現実的な適用が可能であると考えている。

具体的な提示方法として、ユーザがPinterestを利用する際に、メイン画面の横に小さい領域を作りそこで推薦画像を数多く表示する、という方法が考えられる。ページの切り替わりや、ある程度時間が経過するごとに推薦リストを更新することでユーザは推薦リストの中から良い画像を見つけることができるかもしれない。ユーザはこの推薦の

中に気になるものがあればチェックし、良い画像がなければ無視することができるため、従来の行動を阻害することなく繰り返し推薦を行うことができる。以降の議論では、この提示方法を前提とする。

結果における精度は最も高いコンテンツベースフィルタリングで0.4%弱であり、これは例えば1回で20個画像を推薦する場合、12-13回に1回良いものが見つかるというレベルである。これがSVDで次元削減した協調フィルタリングでは0.1%弱であり4倍の差がある。この場合は約50回に1回良いものが見つかるというレベルである。この差により、ユーザはアルゴリズムの推薦能力の差に気づくと思われる。正確性を重視した推薦を行う際には、コンテンツベースフィルタリングが最も良い結果を出しており、有用であると言える。

それに対して、多様性、新規性、セレンディピティに関しては熟慮が必要である。従来の推薦システムでは、ユーザは自らの嗜好に合うアイテムを消費することを想定してきた。しかし、特にアイテムベースの協調フィルタリングでは、推薦が狭い範囲のジャンルにとどまったり、多くのユーザが知っているアイテムに偏ったりすることが指摘されてきた[16]。そのため、多様性や新規性が高いことが望まれてきた。

一方、Pinterestにおいては、ユーザは自らのコレクションのテーマに沿った画像を収集する。ここでは、ユーザは自分の嗜好に合う画像だけでなく、自分の嗜好とは別に、特定のテーマに合う画像を集める。このような使い方では、多様性や新規性の高い推薦は有用でないかもしれない。推薦の多様性は低い方が、ユーザが現在注目しているボードのテーマに沿った画像を得られるかもしれない。また、新規性が低い方が、ボードのテーマを代表するような人気の画像を得られるかもしれない。これらにより、有用なデータベースの働きをするボードを作成できる可能性がある。

また、セレンディピティが高い推薦には、ユーザが今までピンしてこなかったジャンルやカテゴリの画像が含まれている可能性が高い。そのため、ユーザがあるテーマに沿って多くの画像をピンしてきたボードに、画像を追加する目的では使うことができない。しかし、これまで作成してきたものとは異なる新たなテーマのボードを作成するきっかけになる可能性がある。ユーザが新しいボードを作成したいというニーズが確認できれば、セレンディピティが高い推薦を提供しても良いと思われる。

6. おわりに

本稿ではPinterestの画像の推薦において、複数の主要なアルゴリズムによる結果を比較し、それを主要な複数の指標で評価した。評価結果の値は全てかなり小さな値であったが、提示法を工夫することでユーザがアルゴリズムの性能の差を知覚することが出来ると我々は考えている。正確

性に重点をおいた推薦を行うためには、コンテンツベースフィルタリングが有用であることが分かった。多様性や新規性が高い推薦を行うためには、社会ネットワークに基づく推薦が良く、セレンディピティが高い推薦を行うためには、協調フィルタリングが良いことが分かった。サービスの目的やユーザのニーズに合わせて、アルゴリズムを選択する必要があると言える。

今後は協調フィルタリングのSVDとピンのカテゴリを用いたものを組み合わせるなどハイブリッドの手法での評価を行いたいと考えている。さらに、オンラインでの実験や独自インターフェースによる被験者実験により、今回得られた知見のさらなる考察や実際の適用について具体化する予定である。

謝辞

本研究は科研費(15K12150)の助成を受けたものである。Pinterestからのデータ取得については、ミネソタ大学GroupLens ResearchのShuo Chang氏から助言をいただいた。ここに感謝の意を記す。

参考文献

- [1] A. Bellogin, Pablo Castells, and Ivan Cantador. Precision-oriented evaluation of recommender systems: an algorithmic comparison. In *Proceedings of the fifth ACM conference on Recommender systems*, pp. 333–336. ACM, 2011.
- [2] J. Bennett and Stan Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, Vol. 2007, p. 35, 2007.
- [3] J. S Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pp. 43–52. Morgan Kaufmann Publishers Inc., 1998.
- [4] F. Cacheda, Víctor Carneiro, Diego Fernández, et al. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB)*, Vol. 5, No. 1, p. 2, 2011.
- [5] Ò. Celma and Perfecto Herrera. A new approach to evaluating novel recommendations. In *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 179–186. ACM, 2008.
- [6] S. Chang, Vikas Kumar, Eric Gilbert, et al. Specialization, homophily, and gender in a social curation site: Findings from pinterest. pp. 674–686, 2014.
- [7] P. Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*, p. 39, 2010.
- [8] M. Deshpande and George Karypis. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 143–177, 2004.
- [9] E. Gilbert, Saeideh Bakhshi, Shuo Chang, et al. I need to try this?: a statistical overview of pinterest. pp. 2427–2436, 2013.
- [10] C. Hall and Michael Zarro. Social curation on the website Pinterest.com. *Proceedings of the ASIST Annual Meeting*, Vol. 49, No. 1, 2012.
- [11] J. L Herlocker, Joseph A. Konstan, Al Borchers, et al. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 230–237. ACM, 1999.
- [12] Z. Huang, Daniel Zeng, and Hsinchun Chen. A comparative study of recommendation algorithms in e-commerce applications. *IEEE Intelligent Systems*, Vol. 22, No. 5, pp. 68–78, 2007.
- [13] K. Yeswanth Kamath, Ana-Maria Popescu, and James Caverlee. Board Recommendation in Pinterest. *UMAP Workshops*, 2013.
- [14] D. Kluver and Joseph A. Konstan. Evaluating recommender behavior for new users. *Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14*, pp. 121–128, 2014.
- [15] N. Knell. Catching criminals on pinterest? *SOLUTIONS FOR STATE AND LOCAL GOVERNMENT* <http://www.govtech.com/public-safety/Catching-Criminals-on-Pinterest.html>.
- [16] S. M McNee, John Riedl, and Joseph A. Konstan. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*, pp. 1097–1101. ACM, 2006.
- [17] H. J Miller, Shuo Chang, and Loren G Terveen. I love this site! vs. it's a little girly: Perceptions of and initial user experience with pinterest. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1728–1740. ACM, 2015.
- [18] P. Resnick, Neophytos Iacovou, Mitesh Suchak, et al. Grouplens: an open architecture for collaborative filtering of netnews. pp. 175–186, 1994.
- [19] B. Sarwar, George Karypis, Joseph A. Konstan, et al. Item-based collaborative filtering recommendation algorithms. pp. 285–295, 2001.
- [20] M. Zarro, Catherine Hall, and Andrea Forte. Wedding Dresses and Wanted Criminals: Pinterest.com as an Infrastructure for Repository Building. *Seventh International AAAI Conference on Weblogs and Social Media*, pp. 650–658, 2013.
- [21] Y. C. Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, et al. Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 13–22. ACM, 2012.
- [22] C. Zhong, Dmytro Karamshuk, and Nishanth Sastry. Predicting pinterest: Automating a distributed human computation. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1417–1426. International World Wide Web Conferences Steering Committee, 2015.
- [23] C. Ziegler, Sean M McNee, Joseph A. Konstan, et al. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pp. 22–32. ACM, 2005.