

Title	潜在トピックを利用した協調フィルタリングにおける文書の内容が推薦性能に与える影響
Author(s)	西村, 章宏
Citation	大阪大学, 2023, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/92210
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

潜在トピックを利用した
協調フィルタリングにおける
文書の内容が推薦性能に与える影響

2023 年 3 月

西村 章宏

潜在トピックを利用した
協調フィルタリングにおける
文書の内容が推薦性能に与える影響

博士（工学）論文提出先
大阪大学大学院基礎工学研究科

2023年3月

西村 章宏

概要

大量の情報の中からユーザの情報獲得を支援する方法の一つとして推薦システムがある。推薦システムを実現する方法としては、内容ベースフィルタリングと協調フィルタリングに大きく分けることができ、これらを組み合わせた手法も数多く存在する。協調フィルタリングの中では、Matrix Factorization (MF) が評価値の欠損が多い現実のデータセットに対して優れた推薦を行うことができる。ただし、MF は評価値の欠損に強いものの、評価値が極端に少ない場合には適用することが困難である。これは一般にコールドスタート問題と呼ばれており、評価値だけでは不足している情報を補うために、評価値以外の情報（サイド情報と呼ばれる）が利用される。評価値とサイド情報を合わせて活用するアプローチとしては、協調フィルタリングと内容ベースフィルタリングの両方を組み合わせたハイブリッド法が主流である。近年のハイブリッド法の多くは、結合させる手法に前後関係がなく同等のレベルでモデル化を行う完全結合である。MF をベースとした完全結合に注目すると、サイド情報としてアイテムに関する文書を利用する手法が多く提案されている。その中で Collaborative Topic Regression (CTR) は文書から抽出したトピックを利用するアプローチの先駆けとなったモデルであり、その組み合わせ方法がシンプルであるため様々な派生手法を考える上で基準となる手法である。本論文では、CTR において3つの課題を提起する。課題1は“ユーザに関する文書を用いる場合の推薦性能についての詳細な分析”であり、従来のアイテムに関する文書を用いる場合に比べて文書から抽出されるトピック内容にどのような差異があるか、そしてトピックの差異により推薦性能にどのような差異が生じるか分析する必要がある。課題2は“評価値行列と文書が持つ特徴およびハイパーパラメータ間の関係性が推薦性能に与える影響の分析”であり、評価値行列と文書における特徴とは何であるか具体的に定義し、それらの特徴とハイパーパラメータ間の関係についての仮説を立てて検証する必要がある。課題3は“文書中の語彙を制御することによる推薦性能に与える影響の分析”であり、文書中の単語を何らかの基準によって削除、または他の情報源を利用して追加することによる推薦性能への影響を分析する必要がある。この3つの課題への取り組みとして、著者は4つの研

究を行った。課題 1 に対しては、ユーザに関する文書を用いた際のトピック内容と可視化結果についての分析を行った。その結果、ユーザに関する文書から抽出されたトピック内容はユーザが興味を持っている対象やユーザの属性であり、これらはユーザ層という観点を表していることが分かった。次に、CTR モデルにおいて、トピック情報源としてアイテムに関する文書情報を用いた場合と、ユーザに関する文書を用いた場合を実験により比較し、正確性と利便性という 2 つの観点から推薦性能がどれだけ異なるのかを分析した。実験の結果、アイテムに関する文書を用いた場合では利便性の高い推薦、ユーザに関する文書を用いた場合では正確性の高い推薦を行う傾向が分かった。課題 2 および課題 3 に対しては、スパース度と利用する単語の種類およびハイパーパラメータにおいて推薦性能に影響を与える要因について仮説を立て、複数のデータセットを用いた実験によりそれらの要因間の関係性についての分析を行った。実験の結果、スパース度と文書頻度の組み合わせに関しては、正確性と利便性のバランスが良い推薦を行うのに複数のデータセットで共通する傾向が存在することが分かった。課題 3 に対しては、アイテムに関する文書とは異なる情報源を用いて文書の内容に合った単語を追加することにより、推薦性能が向上するケースが存在するかの分析を行った。実験の結果、興味の幅が広いユーザに限定した場合には、語彙補完を行うことで推薦性能が向上することが分かった。最後に、本研究をふまえて筆者が考える今後の展望を述べる。

Content

第 1 章	序論	1
1.1	推薦システム	1
1.2	トピックを利用した協調フィルタリング	6
1.2.1	CTR の概要	6
1.2.2	CTR の課題	8
1.3	本研究の貢献	10
1.3.1	ユーザに関する文書を用いる場合の影響を分析	11
1.3.2	スパース度と利用する単語の種類およびハイパーパラメータの組み合わせが性能に与える影響を分析	12
1.3.3	推薦対象のユーザに適した語彙補完による推薦性能の向上	13
1.3.4	実用上の貢献	14
1.4	本論文の構成	16
第 2 章	ユーザに関する文書から抽出されるトピックの分析	17
2.1	背景と目的	19
2.2	関連研究	20
2.2.1	SNS を利用したアイテム情報の抽出	20
2.2.2	SNS を利用したユーザ情報の抽出	20

2.2.3	ユーザ間の関係の可視化	21
2.3	実験方法	21
2.3.1	情報源の種類	21
2.3.2	前処理	23
2.3.3	特徴量抽出	25
2.3.4	可視化	27
2.3.5	データセット・ハイパーパラメータ	29
2.4	結果と考察	30
2.4.1	可視化手法毎の結果	31
2.4.2	情報源毎の結果	35
2.5	結論	41
第3章	トピック抽出元となる文書の種類が推薦性能に与える影響の分析	43
3.1	背景と目的	44
3.2	関連研究	45
3.2.1	MFとサイド情報	45
3.2.2	MFとトピックモデルの融合	46
3.3	推薦モデル	47
3.3.1	iCTR	47
3.3.2	uCTR	52
3.4	実験設定	53
3.4.1	データセットとスパース度の設定	53
3.4.2	前処理	55
3.4.3	ハイパーパラメータの設定	56
3.4.4	評価方法	57
3.5	結果と考察	59

	3.5.1	Perplexity	59
	3.5.2	再現率	60
	3.5.3	被覆率	61
	3.5.4	多様性	62
	3.5.5	推薦例	63
3.6		結論	66
	3.6.1	結果の要約	66
	3.6.2	今後の課題	68
第 4 章		スパース度と利用する単語の種類が推薦性能に与える影響の分析	69
4.1		背景と目的	69
4.2		関連研究	72
	4.2.1	サイド情報を利用する MF における課題	72
4.3		単語選択手法	73
4.4		実験設定	74
	4.4.1	データセットとスパース度の設定	74
	4.4.2	前処理	76
	4.4.3	語彙サブセットの設定	76
	4.4.4	ハイパーパラメータの設定	77
	4.4.5	評価方法	79
4.5		結果と考察	80
	4.5.1	ハイパーパラメータ間の関係性	80
	4.5.2	スパース度と単語選択 (CiteULike)	82
	4.5.3	スパース度と単語選択による影響のデータセット間での比較	87
	4.5.4	文書情報による性能向上が見込めないケース	89
4.6		結論	90

第 5 章	語彙補完を活用した複数分野に関心を持つユーザを対象とする推薦	93
5.1	背景と目的	95
5.2	関連研究	97
5.2.1	特定のユーザにより適した推薦を行う手法	97
5.2.2	学習済み言語モデルを活用した推薦に関する研究	98
5.3	語彙補完手法	98
5.4	実験設定	100
5.4.1	データセットとスパース度の設定	100
5.4.2	語彙補完の設定と文書に対する前処理	101
5.4.3	ハイパーパラメータの設定	102
5.4.4	評価方法	103
5.5	結果と考察	103
5.5.1	スパース度が低いケース (original)	103
5.5.2	スパース度が非常に高いケース (max2u)	105
5.5.3	推薦例の比較	107
5.6	結論	108
第 6 章	結論	111
6.1	課題と結論	111
6.1.1	課題 1 に対する結論	111
6.1.2	課題 2 に対する結論	112
6.1.3	課題 3 に対する結論	113
6.2	今後の展望	114
謝辞		117
参考文献		119

第 1 章

序論

1.1 推薦システム

情報過多によりユーザが本来必要とする情報が他の情報に埋もれてしまうと、ユーザはその情報を容易に取得できなくなる。このような状況を解決するため、大量の情報の中からユーザの情報獲得を支援する方法の一つとして推薦システム (Recommender System) がある。推薦システムに関する研究は 1990 年代から現在まで広く行われている。推薦システムを実現する方法としては、ユーザの興味や嗜好に関する情報とアイテムの内容に関する情報を元に推薦を行う内容ベースフィルタリング (Content-Based Filtering)、対象ユーザの嗜好と似た嗜好を持つ他者が好むアイテムを推薦する協調フィルタリング (Collaborative Filtering) に大きく分けることができ、これらを組み合わせた手法も数多く存在する [1, 2, 3].

内容ベースフィルタリングは情報検索の分野から派生しており、図 1.1 に示すように、ユーザの嗜好を表現するプロフィールとアイテムの性質を表現するプロフィールを用いて、対象となるユーザのプロフィールに適合するアイテムを探し出す。基本的な手法としては、ユーザのプロフィールとアイテムのプロフィールをそれぞれ特徴ベクトルで表現し、距離を基準にユーザに適したアイテムを見つけることで推薦を行う。ユーザのプ

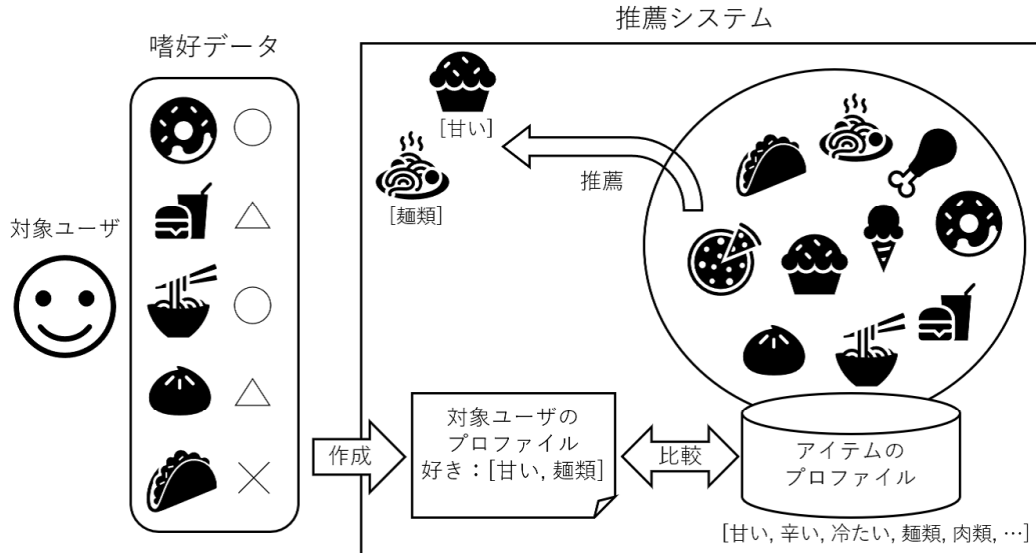
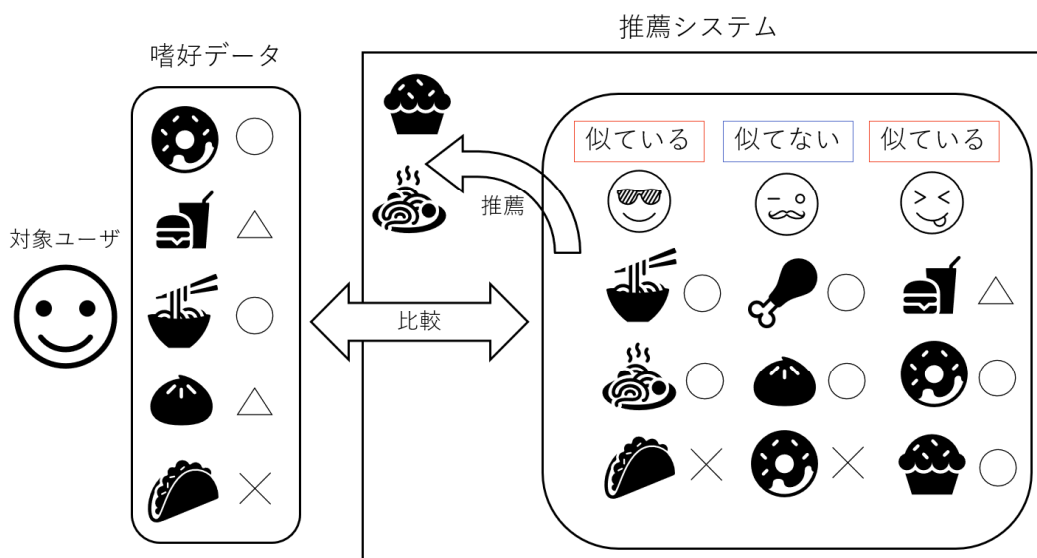


図 1.1 内容ベースフィルタリングの概要図

ロファイル構築には、ユーザから直接アイテムに関する嗜好を聞き出す方法や、購買履歴のような過去の情報からユーザのアイテムに対する嗜好を得る方法が用いられている。アイテムのプロファイル構築には、アイテムに関するドメイン知識やアイテムに関するテキスト情報、アイテム自体から抽出した情報を利用する。例えば、本であれば作者やジャンル、パソコンであればスペックや値段といった、アイテムに付与されたメタデータをベクトルで表現する方法がある [1]。メタデータ以外では、アイテムの説明文やレビュー文などのテキスト情報や、画像、音楽から特徴量を抽出してベクトルに変換する方法がある [4]。内容ベースフィルタリングは、プロファイルが適切に構築されていれば、新しく追加されたアイテムや新規ユーザであっても推薦が可能となる。一方で短所としては、対象ユーザの履歴だけに依存するため、そのユーザがこれまで興味を持ってきたアイテム群とは異なる分野のアイテムは推薦されにくいことが挙げられる [1]。加えて、プロファイル構築のために膨大な量のユーザとアイテムに関するデータを収集する手間があることと、それらのプロファイル構築処理に計算コストがかかることも短所である。

協調フィルタリングは、Resnick らが提案したニュース記事の推薦システムである



(a) 協調フィルタリングの概要図

Diagram (b) shows the pattern of the evaluation value matrix. At the top, eight food icons are listed: chicken drumstick, burger, bowl of ramen, bowl of ramen, dumpling, taco, donut, and cupcake. Below them is a grid with four rows representing users and eight columns representing food items. The rows are labeled with user icons and similarity labels: a smiley face (Target User), sunglasses (Similar), a mustache (Not similar), and a winking smiley face (Similar). The grid cells contain ratings: △, ○, or ×.

対象ユーザ	チキン	ハンバーガー	ラーメン	ラーメン	餃子	タコ	ドーナツ	カップケーキ
対象ユーザ (スマiley)		△	○		△	×		○
似ている (サングラス)			○	○		×		
似てない (髭)	○				○			×
似ている (ウィンク)		△					○	○

(b) 評価値行列の模式図

図 1.2 協調フィルタリングと評価値行列の概要図

GroupLens [5] が基礎となっている。協調フィルタリングではユーザがアイテムを好む程度（評価値）が行列形式で表現され、この行列は評価値行列と呼ばれる。協調フィルタリングの概要と評価値行列の模式図を図 1.2 に示す。仮に評価値行列の行がユーザを表すなら、対象ユーザの行ベクトルと他のユーザの行ベクトルから類似度を求め、類似度が

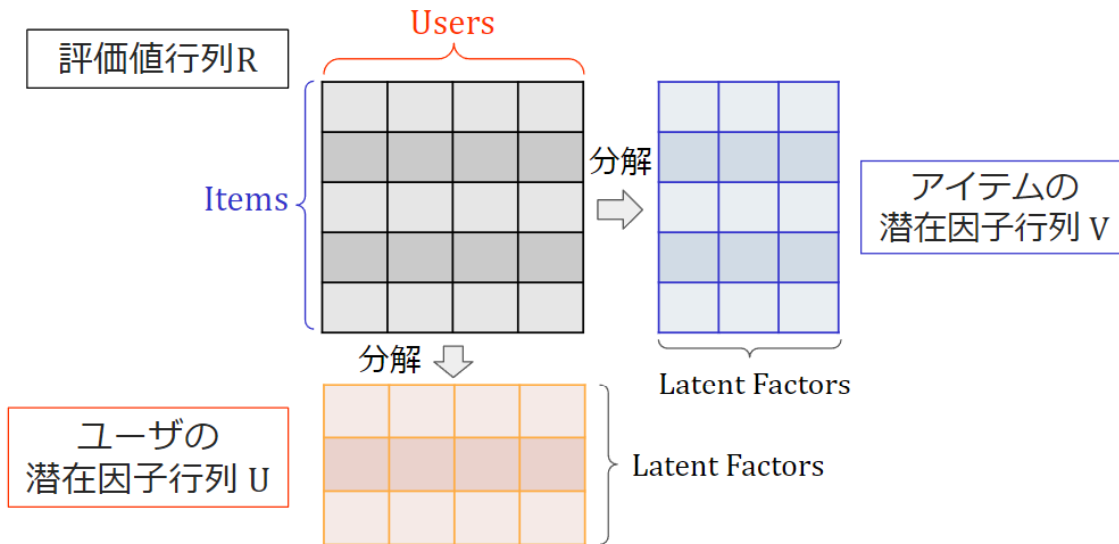


図 1.3 Matrix Factorization (MF) の模式図

高いユーザが好むアイテムを対象ユーザに推薦する。同様にアイテム間の類似度を求め、対象ユーザが好むアイテムに似たアイテムを推薦することもできる。協調フィルタリングの長所は、アイテムに関するプロフィール構築が不要、すなわち特徴ベクトルを作成する必要がなく手軽に利用できる点である。

一方、ユーザや評価値を十分な量だけ用意しないといけない点が短所である。特に実システムやサービスで得られる評価値行列は、欠損値の割合（以降、スパース度と呼ぶ）が高く、評価値の不足により推薦が正常に機能せず性能が大きく低下することがある。そのため、より少ない評価値から効率的に未知の評価値を推定する手法として、Matrix Factorization (MF) [9, 10] が提案されており、商用の実データに対しても優れた推薦を行うことができる [15]。

MF はユーザとアイテムをより低次元な潜在特徴空間へマッピングする手法であり、与えられた評価値行列はユーザの潜在因子行列とアイテムの潜在因子行列に分解される。図 1.3 に示す MF の模式図を参照すると、ユーザの潜在因子行列 U とアイテムの潜在因子行列 V は、 U の行および V の列が共通の潜在因子 (Latent Factor) を表現している。つまり、ユーザの潜在因子行列は各ユーザの特徴量である潜在因子ベクトルを並べたもの

であり，アイテムの潜在因子行列は各アイテムの特徴量である潜在因子ベクトルを並べたものである．そのため，あるユーザとあるアイテムの潜在因子ベクトルの積をとることで，潜在因子の類似性を考慮した評価値が求まる．これにより，評価値行列 \mathbf{R} における欠損値に対して予測評価値を推定することができ，欠損値の存在しない予測評価値行列を求めることができる．ただし，MF による欠損値への対処にも限度があり，評価値が極端に少ない新規アイテムや新規ユーザに対してはうまく適用することができない．これは一般にコールドスタート問題（Cold-Start Problem）と呼ばれる [1, 11, 14]．

コールドスタート問題への対処方法として，ユーザに直接質問を行って評価値を取得する方法 [12, 13] は確実にユーザの嗜好を得られるが，ユーザに手間をかけさせることになり利用場面が限定される．別のコールドスタート問題への対処方法として，評価値だけでは不足している情報を補うために，サイド情報（Side Information）をユーザに直接質問することなく暗黙的に取得して活用するアプローチがある．サイド情報の例としては，アイテムに付与されたメタデータ，アイテムの説明文やレビュー文などのテキスト情報，ソーシャルメディアから得られるユーザ間のリンク情報やインタラクション情報 [14] などが挙げられる．評価値とサイド情報を合わせて活用するアプローチとしては，協調フィルタリングと内容ベースフィルタリングの両方を組み合わせたハイブリッド手法 [2, 3] が主流である．ハイブリッド手法は，協調フィルタリングと内容ベースフィルタリングの結合の度合いが疎であるものから順に，混合（Mixed），切替え（Switching），メタ推薦（Meta-Recommendation），継続（Cascade），特徴拡張（Feature Augmentation），抽象情報（Abstraction），完全結合（Total Integration）の 7 分類が提唱されている [2]．最も結合が疎である混合の例としては，内容ベースフィルタリングにより出力された推薦結果と協調フィルタリングにより出力された推薦結果を混合して提示する手法 [6] がある．結合が密である抽象情報の例としては，内容ベースフィルタリングにおいて，ユーザの特徴ベクトルは全ユーザのプロファイルに依存して生成されると定義する方法 [7] がある．これは，ユーザの特徴ベクトル生成時にユーザ間の類似度を利用する点が協調フィルタリングの役割を担っており，その特徴ベクトルを入力として内容ベースフィルタリング

を行うということで、結合に前後関係があるのが抽象情報である。そして近年のハイブリッド手法の多くは、結合させる手法に前後関係がなく同等のレベルでモデル化を行う完全結合である。完全結合に分類されるモデルの多くは、評価値とサイド情報の生成過程を記述した確率モデルとして定式化されている。MF をベースとした完全結合に注目すると、サイド情報としてアイテムに関する文書を利用する手法が多く提案されている [16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]。これらの手法の特徴は、アイテムに関する文書から潜在トピック (Latent Topic) を抽出し、ユーザとアイテムを表現する潜在因子ベクトルにこの潜在トピックを反映させていることである (以降、潜在トピックは単にトピックと呼ぶ)。例として、代表的な手法の模式図を図 1.4 に示す。図中のモデルの詳細に関しては、1.2.1 節で説明する。これにより、評価値が全く存在しない新規アイテムに対しても、抽出されたトピックに誘起して潜在因子ベクトルが生成されるため、評価値を推定することができる。加えて、入力として与えるアイテムに関する文書は単語単位に分離できていればよく、アイテムに関するプロフィール構築が不要であるため協調フィルタリングと内容ベースフィルタリングの利点が共存している。

1.2 トピックを利用した協調フィルタリング

1.2.1 CTR の概要

1.1 節で述べたように、サイド情報としてアイテムに関する文書を利用する MF は様々な手法が提案されている。それらの中で、図 1.4 に示す Collaborative Topic Regression (CTR) [16] は文書から抽出したトピックを利用するアプローチの先駆けとなったモデルであり、その組み合わせ方法がシンプルであるため様々な派生手法を考える上で基準となる手法である [39]。CTR はアイテムに関する文書からトピックを抽出し、それをアイテムの潜在因子ベクトルに反映させることで、評価値行列 R に欠損値が多い場合においても高い推薦精度を維持することに成功している。文書からのトピック抽出は、自然言語処理やテキストマイニングの分野で発展したトピックモデル (Topic Model) [40, 42] の

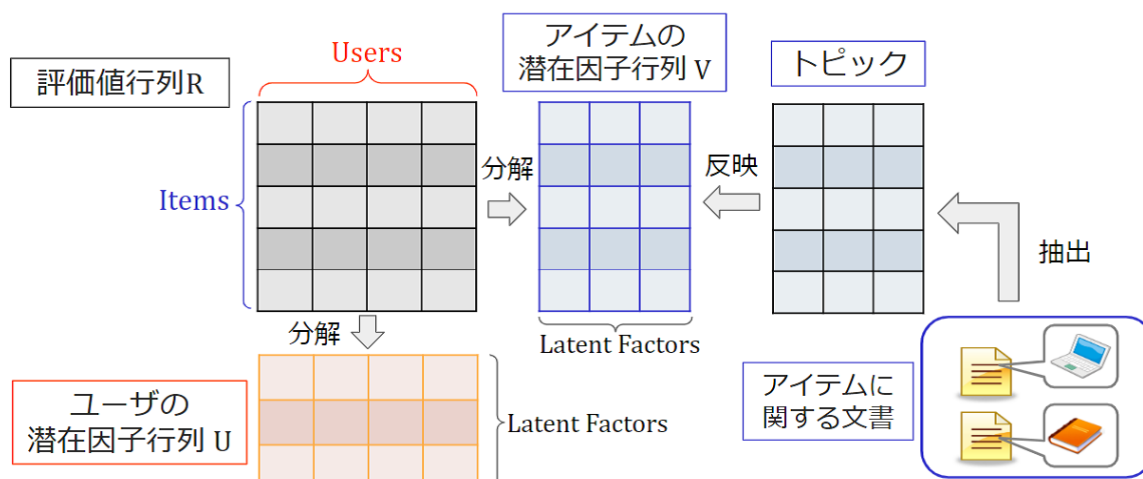


図 1.4 アイテムに関する文書を利用する Collaborative Topic Regression (CTR) の模式図

技術を利用している。トピックモデルは、特異値分解に端を発する潜在意味解析 (Latent Semantic Analysis) を行う手法の発展形であり、特異値分解を確率モデルとして定式化した pLSI[43] を階層ベイズモデルに拡張したモデルの総称である。

トピックモデルの基礎となるのは Latent Dirichlet Allocation (LDA) [40] という手法である。LDA は文書の生成過程を表現する階層ベイズモデルであり、観測変数である文書中の単語をもとに、各文書が有するトピックを表す確率分布 (トピック比率) と各トピックを構成する単語の確率分布 (単語比率) を推定する。模式図で表すと、図 1.5 となる。トピックには明確なラベルは付与されておらず、そのトピックの確率上位の単語から類推して何を表しているか解釈する。階層ベイズモデルの枠組みにおいて、MF も確率モデルとして定式化することができれば、トピック比率とアイテムの潜在因子ベクトルとの間に依存関係を定義することができる。あるアイテムの潜在因子ベクトルは、そのアイテムに関する文書が有するトピックに基づき生成されるという依存関係を定義することで、評価値が全く存在しない新規アイテムに対しても評価値を推定することが可能になる。また、トピックモデルは自然言語で書かれた文書を簡易な前処理を行うだけで適用できることも利点である。

CTR は MF を確率モデルとして定義した Probabilistic Matrix Factorization (PMF) [44, 45] と LDA[40] を融合したモデルである。より具体的には, CTR はトピックからアイテムの潜在因子へ依存関係を定義することで融合を行ったモデルである。PMF とトピックモデルをどのように融合させるかにより, すなわちモデルを構成する変数間の依存関係の定義による違いで派生手法がいくつか提案されている。例としては, Hidden factors as topics (HFT) [17] は CTR に似たモデル構成だが, トピックの抽出元である文書がレビュー文である点と, アイテムの潜在因子ベクトルからトピックが生成されると仮定しており CTR と依存関係が逆である点が異なる。また別の例として, TopicMF[18] はアイテムとユーザの両方の潜在因子ベクトルからトピックへ依存関係を定義している。PMF とトピックモデルだけでなく, ソーシャルメディアから得られるリンク情報やインタラクション情報を利用できるようにモデルを拡張するアプローチも提案されている [20, 21, 22, 23, 24, 26]。近年では, トピックモデルによる特徴量の抽出ではなく, 深層学習を利用して特徴量を抽出するアプローチも見られる [27, 28]。特に, 文書中の単語を Bug-of-Words の形式で扱うトピックモデルでは失われてしまう語順の情報を活用するアプローチが注目されている。語順の情報を活用するため, Long Short-Term Memory (LSTM) とトピックモデルを組み合わせた手法 [29] や, Attention を導入した畳み込みニューラルネットワーク (CNN) で特徴量を抽出する手法 [30] が提案されている。

1.2.2 CTR の課題

CTR は評価値行列だけでなくアイテムに関する文書を利用するタイプの MF において基準となる手法である。利用する文書の種類という観点では, ユーザのプロフィール文のようなユーザに関する文書を用いた場合の分析はほとんど行われてこなかった。また, 1.2.1 節で述べたように, ソーシャルメディアから得られるリンク情報やインタラクション情報といった文書以外の追加の情報を利用できるようにする拡張 [20, 21, 22, 23, 24] や文書から特徴量を抽出する方法の改良 [27, 28] についての研究は盛んに行われているが,

各トピックの単語比率

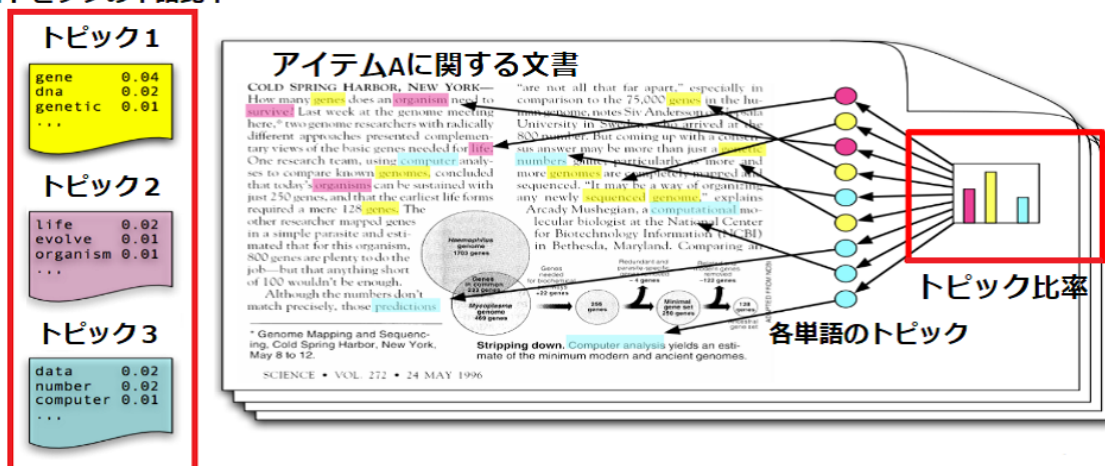


図 1.5 Latent Dirichlet Allocation (LDA) の模式図. ([41] より引用および加工)

CTR をベースとしたモデルにおいて、評価値行列が持つ特徴と文書自体が持つ特徴が推薦性能に与える影響については深く研究されていない。以上の分析や研究があまり行われていない内容をもとに、CTR に関して以下の課題を挙げる。

- 課題 1. ユーザに関する文書を用いる場合の推薦性能についての詳細な分析
- 課題 2. 評価値行列が持つ特徴と文書自体が持つ特徴およびハイパーパラメータ間の関係性が推薦性能に与える影響の分析
- 課題 3. 文書中の語彙を制御することによる推薦性能に与える影響の分析

課題 1. に関して、Wang らが最初に提唱した CTR[16] では、アイテムに関する文書からトピックを抽出した際の結果を分析しているが、ユーザに関する文書を用いることに関しては深く言及されていない。文書から抽出されるトピック内容にどのような差異があるか、そしてトピックの差異により推薦性能を評価するいくつかの指標にどのような差異が生じるかは明らかになっていない。課題 2. に関して、1.2.1 節で述べた CTR をベースとした派生手法 [20, 21, 22, 23, 24, 25, 27] は、モデルの拡張により特定のドメインにおける推薦性能の向上に焦点を当てているが、CTR がどのような特徴を持つデータセッ

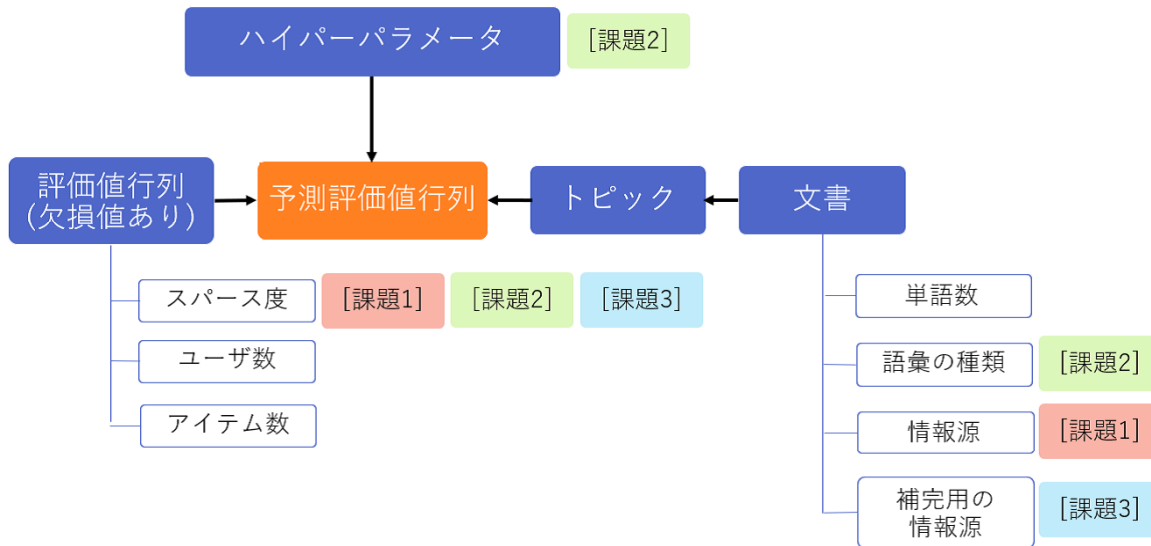


図 1.6 本論文において着目する CTR の構成要素と課題の関係

トに対して、どのようにハイパーパラメータを設定すれば性能が向上するかといった詳細な分析は行われていない。評価値行列が持つ特徴と文書自体が持つ特徴を具体的に定義し、それらの特徴とハイパーパラメータ間に関する仮説を立てて検証する必要がある。課題 3. に関して、文書中の単語を何らかの基準によって削除、または他の情報源を利用して追加することによる推薦性能への影響について詳細な分析は行われていない。課題 2. でそれぞれの特徴間に関する関係性が明らかになることでその知見を活用することができ、例えば評価値行列の特徴に応じて文書中の単語を追加・削除して語彙を操作することで推薦性能の向上に繋がる可能性がある。最後に、本論文において着目する CTR の構成要素と課題の関係を図 1.6 に示す。

1.3 本研究の貢献

CTR に関する 1.2.2 節の 3 つの課題に対する、本論文の貢献を 1.3.1 節から 1.3.3 節で述べる。本論文では、課題への取り組みとして 4 つの研究を行う。それぞれの課題と研究の関係を図 1.7 に示す。

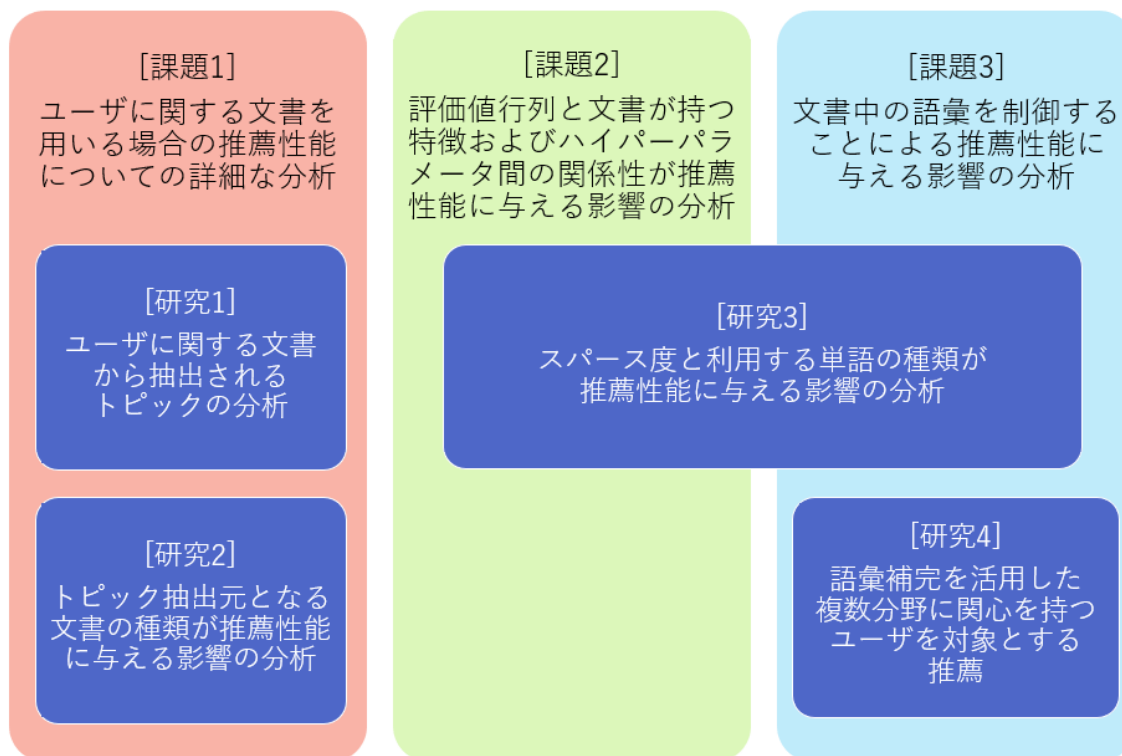


図 1.7 本論文が取り扱う課題と研究項目との関係

1.3.1 ユーザに関する文書を用いる場合の影響を分析

1.2.2 節の課題 1 に対して、ユーザに関する文書を用いた際のトピック内容と推薦結果について分析を行った。まずは CTR を利用した推薦結果の分析を行う前に、ユーザに関する文書から抽出されるトピックはどのような内容であるかを分析した。ユーザに関する文書から抽出されるトピックに関しての知見を得ることは、ユーザに関する文書を利用した際の CTR による推薦結果を理解するうえで意義がある。そこで、ユーザに関する文書を大量に取得することができるソーシャル・ネットワーキング・サービス (SNS) の一つである Twitter を利用し、抽出したトピック内容の分析とトピックを特徴量としてアイテム間の関係を可視化した結果の分析を行うことで、アイテムとユーザそれぞれで

得られるトピックの差異を明らかにした。その結果、ユーザに関する文書から抽出されたトピック内容はユーザが興味を持っている対象やユーザの属性であり、これらはユーザ層という観点を表していることが分かった。この詳細は2章で説明する。

次に、ユーザに関する文書から得られるユーザ層というトピックを潜在因子ベクトルに反映させることを考える。CTRはアイテムから抽出したトピックとユーザから抽出したトピックを同時に扱うことを想定しておらず、潜在因子ベクトルに反映させることができるトピックはアイテムもしくはユーザから抽出したどちらか一方だけとなる。本研究では、CTRモデルにおいて、トピック情報源としてアイテムに関する文書情報を用いた場合と、ユーザに関する文書を用いた場合を実験により比較し、ユーザの嗜好に適した推薦であるか（正確性）と偏った推薦でないか（利便性）という2つの観点から推薦性能がどれだけ異なるのかを調べる。また、極端に評価値が欠損している場合や評価値が十分に存在する場合とで比較するため、評価値行列のスパース度を変化させた複数の設定で実験を行う。実験の結果、アイテムに関する文書を用いた場合では利便性の高い推薦、ユーザに関する文書を用いた場合では正確性の高い推薦を行う傾向が分かった。この詳細は3章で説明する。

1.3.2 スパース度と利用する単語の種類およびハイパーパラメータの組み合わせが性能に与える影響を分析

1.2.2節の課題2および課題3に対して、推薦性能に影響を与える要因について仮説を立て、複数のデータセットを用いた実験によりそれらの要因間の関係性について分析を行った。まず推薦性能に影響を与える要因として、トピックすなわち文書中の単語の構成、評価値のスパース度、トピックと評価値の影響を制御するハイパーパラメータ、この3つが主要な要因ではないかと考える。CTRは評価値から抽出された潜在因子に文書から抽出されたトピックを反映させることで、不足した情報を補うモデルである。評価値行列のスパース度が高い状況では、評価値からアイテム間またはユーザ間の局所的な

関係は抽出できるが、アイテムのカテゴリのような大域的な関係は抽出が困難だと考えられる。このような場合は、文書から抽出されたトピックにより大域的な関係を表す情報を補完できればよく、文書中でそのような大域的な情報を持つのは様々な文書に出現する単語（文書頻度の高い単語）だと考える。逆にスパース度が低い状況では、評価値から大域的な関係は十分に抽出可能であり、大域的な関係を表すトピックはノイズになってしまう可能性がある。この場合は、文書頻度の高い単語は選択しない方がよい。よって、CTR を効率的に利用するためには、スパース度に応じて有益なトピックを抽出する必要があると考え、文書頻度に基づいて利用する単語を選択することが重要だと仮定する。また、CTR にはいくつかのハイパーパラメータが存在しており、その中でもトピックをどれくらい強く反映するかを調整するものは生成される潜在因子ベクトルを通して推薦結果に大きく影響を与える。これらの要素は互いに密接に関連し合っている可能性が高く、その関係性を明らかにし、CTR の性能を十分に引き出すための知見を整理することは意義がある。本研究では、異なる評価値行列のスパース度の下で、文書頻度を基準に選択する単語を変え、また潜在因子ベクトルに影響するハイパーパラメータの組み合わせを変えることで、CTR の性能がどのように変化するか複数のデータセットで比較し、共通する推薦結果の傾向と文書情報の利用による性能向上が期待できない条件をまとめる。実験の結果、スパース度と文書頻度の組み合わせに関しては、複数のデータセットで共通の傾向が存在することが分かった。この詳細は 4 章で説明する。

1.3.3 推薦対象のユーザに適した語彙補完による推薦性能の向上

1.2.2 節の課題 3 に対して、アイテムに関する文書とは異なる情報源を用いて文書の内容に合った単語を追加することにより、推薦性能が向上するケースが存在するか分析を行った。CTR において、各アイテムの特徴を表現する潜在因子ベクトル（アイテムの潜在因子行列の行ベクトル）は、評価値情報から抽出される暗黙的な関係性に加えて文書情報から抽出されるトピックが反映されている。トピック抽出の段階では、文書を構成す

る単語から、その文書が持つトピック比率を推定する。トピック比率は確率分布であり、同じ分野の文書同士はトピック比率の類似度は高くなる。しかし、分野が近くてもそれぞれの文書を構成する語彙が非常に専門的だと文書間で共通する語彙は少なくなるため、文書間のトピック比率の類似度は小さくなりやすく、結果として潜在因子ベクトルへも影響して推薦対象として選択されにくくなる。その様な文書に対して、文書の内容に適合した専門度のより低い単語を追加することができれば、狭い同分野の文書だけでなく近い分野の文書との類似度が大きくなる。

ここで論文・科学記事といった専門性の高いアイテムを扱うドメインにおいて、推薦対象として複数の異なる分野に関心を持っている興味の幅が広いユーザを想定する。興味の幅が広いユーザは、そのユーザが興味を持つ分野における人気が高いアイテムより、それに近い分野における人気が高いアイテムの方を好む傾向があると筆者は考えた。本研究では、トピック抽出元となる文書へ単語を追加すること（以降、語彙補完と呼ぶ）で、上記で述べたようにユーザの興味に近い分野における人気が高いアイテムが推薦対象に加わるようにし、興味の幅が広いユーザへの推薦性能の向上を目指す。実験の結果、興味の幅が広いユーザに限定した場合には、語彙補完を行うことで推薦性能が向上することが分かった。この詳細は5章で説明する。

1.3.4 実用上の貢献

1.3.1 節から 1.3.3 節では、CTR の各課題に対する本研究における貢献の概要を説明を行った。本節では、それらの貢献を実用上どう活用できるかに焦点を当てて説明を行う。推薦サービスを提供する企業や組織が、協調フィルタリングによる推薦システムを構築しようとする場合を想定する。仮に既存の購買履歴やブックマーク履歴といった評価情報が十分に集まってない状況で、アイテムの説明文やユーザの自己紹介といったプロフィール情報が利用できるなら、CTR をベースとしたモデルが採用候補となるだろう。しかし、CTR をベースとしたモデルを採用することを決める前に、文書情報を利用

することで本当に性能向上が期待できるか検証を行うべきである。1.3.2 節の分析より、評価値行列の特徴から性能向上が期待できないケースを判断することができる。CTR をベースとしたモデルを採用することを決めた後、アイテムに関する文書かユーザに関する文書のどちらを主に利用するか検討が必要になる。判断基準としては、利用できる文書の量や、1.3.1 節の分析より、正確性と利便性のどちらの評価指標を優先するかが挙げられる。ハイパーパラメータの設定に関しては、一般にグリッドサーチで決めることになるが、ハイパーパラメータの数とグリッドの数が増えるにつれ指数的に探索時間が増加してしまう。1.3.2 節の分析より、極端な性能低下を引き起こす可能性がある組み合わせが存在することが分かったため、それをグリッドサーチから除外することで値を決めるのに費やす時間を削減することができる。評価値行列のスパース度や推薦対象に応じて、文書に単語の追加・削減の処理を行うことで、CTR を効率的に活用できるか検討すべきである。1.3.2 節の分析より、スパース度に応じて、文書頻度を基準に語彙を選択して文書中の単語を削減することで性能低下を抑制できる。また、1.3.3 節の分析より、興味の幅が広いユーザを推薦対象とするなら、元の文書とは異なる情報源を利用して文書へ単語を追加することで性能向上が期待できる。これまで述べた実用上で期待される貢献をまとめると以下となる。

- 評価値行列の特徴を調べることにより、文書の利用による性能向上の可能性を判断できる
- 評価値行列のスパース度および重視する性能指標に応じて、利用するのはアイテムに関する文書が良いか、ユーザに関する文書が良いか判断できる
- 評価値行列のスパース度に応じて、文書頻度を基準に語彙を選択して文書中の単語を削減することで性能低下を抑制できる
- 推薦対象として興味の幅が広いユーザを想定するなら、異なる情報源を利用して単語を追加することで性能向上が期待できる
- 極端な性能低下を引き起こす可能性があるハイパーパラメータの組み合わせを未

然に回避できる

1.4 本論文の構成

本論文の構成は以下となる。1章は、推薦システムに関する研究分野の概要とその中で
の本研究の立ち位置の説明、および本研究の貢献についての説明。2章は、1.3.1節の前
半で述べた、ユーザに関する文書から抽出されるトピックはどのような内容であるかにつ
いての分析を行う。3章は、1.3.1節の後半で述べた、ユーザに関する文書から得られる
トピックを潜在因子ベクトルに反映した場合と従来のアイテムに関する文書の場合、お
よびMFとの推薦性能を比較し、情報源の違いによる影響を分析を行う。4章は、1.3.2
節で述べた、スパース度と利用する単語の種類およびハイパーパラメータの組み合わせ
が性能に与える影響についての分析を行う。5章は、1.3.3節で述べた、推薦対象のユー
ザに適した語彙補完による推薦性能の向上を検証する。最後に6章では、これらの研究
についての結論と今後の発展に関して述べる。

第 2 章

ユーザに関する文書から抽出される トピックの分析

本章では、1.2.2 節で述べた課題 1 に対する取り組みとして、ユーザに関する文書から抽出されるトピックがどのような内容であるか分析を行う。分析に際して、ユーザに関する文書から抽出されるトピック内容と従来のアイテムに関する文書から抽出されるトピックの内容を比較することで、ユーザに関する文書から抽出されるトピックへの理解は深まると思われる。一般に、ユーザに関する情報の取得はサービス提供者でないと困難であるが、一部のソーシャル・ネットワーキング・サービス（SNS）ではユーザに関する情報が一般公開されているので、それを利用することで対処可能である。本研究では、様々な種類のユーザおよびアイテムに関する十分な量の文書を API（Application Programming Interface）を利用して容易に取得可能であることを理由に、実験対象のドメインとして SNS の一種である Twitter を選択した。Twitter では、ユーザの短文投稿（tweet）に加え、ユーザ本人が記述したプロフィール文など情報源の異なるテキスト情報が得られる。ここでは容易に取得可能な Twitter のユーザに関する情報を対象に、記述されているトピックがどのようなものであるかを示し、アイテムに関する情報（文書）とどれだけ異なるかを示す。抽出されたトピックは、推薦対象のアイテムに全く関連してい

なければ推薦に利用することは難しく、アイテムに関する文書と全く同じであればユーザに関する文書を利用する必要性がなくなる。

本章では、Twitter 上で話題になりやすい有名人をアイテムとする。その有名人本人が発信した tweet は、その有名人に関する情報やファンを意識した有名人のキャラクター性を表現する発言を含んでいるため、アイテムに関する文書とする。ユーザのプロフィール文は、そのユーザに関する社会的情報や趣味・興味を含んでいることが多いため、ユーザに関する文書とする。また、ユーザが有名人に言及した tweet をアイテムおよびユーザに関する文書の比較対象とする。ユーザが有名人に言及した tweet は、推薦システムの文脈に当てはめると、ユーザがアイテムに対する意見や感想を記述したレビュー文に相当する。この環境でユーザに関する情報がアイテムの推薦性能の向上に貢献する可能性があるかどうかを明らかにする。以下に、Twitter における情報源と文書の種類との対応関係を示す。

- ユーザが有名人に言及した tweet：アイテムに対する擬似的なレビュー文書
- ユーザのプロフィール文：ユーザに関する文書
- 有名人本人が発信した tweet：アイテムに関する文書

本研究では、Twitter における上記の情報源を対象に、トピックを構成する語彙の分析および可視化による比較分析を行うことで、ユーザに関する文書から抽出したトピックへの理解を深める。

本章の構成は以下となる。本章の研究における背景と目的を 2.1 節で述べる。関連研究として、SNS を利用したアイテムやユーザの情報抽出に関する研究と、有名人間の関係の可視化を行う研究を 2.2 節で示す。実験方法の説明を 2.3 節で行う。実験結果の提示と考察を 2.4 節で行う。実験結果から導かれる結論を 2.5 節で述べる。

2.1 背景と目的

短文投稿 SNS の一つである Twitter は、全世界における月間アクティブユーザ数が約 3 億 3000 万人 (2019 年時点)、日本においては約 4500 万人 (2017 年時点) で幅広いユーザー層に利用されている [77]. Twitter における投稿の単位となる 1 つの短文発言は tweet ^{*1} と呼ばれている. tweet はそのユーザの日常に関する内容であることが一般的だが、その時々で話題になっているニュースや有名人に関する内容の tweet も多く見られる. 本章の冒頭で述べた通り、ここではアイテムとして Twitter 上で話題になりやすい有名人を選択する. 本章における分析で必要となる、有名人本人が発信した tweet, 一般ユーザのプロフィール文, 一般ユーザが有名人に言及した tweet は、いずれも Twitter の公式 API を利用して容易に取得することができる. これらのテキスト情報に対して、抽出されるトピック内容の分析をそれぞれ行う.

文書から抽出されるトピックには明確なラベルは付与されないため、トピック内容の解釈を行う方法を検討する必要がある. 個々のトピック内容を把握するための方法としては、トピックを構成する上位単語を並べる方法 [40] や、トピックを構成する単語と一致度が高い Wikipedia 記事を探す手法 [47] などが知られている. 別のアプローチとして、トピックがどのような観点を表しているかを把握することでトピック内容の解釈を行う方法が考えられる. 例えば、本の説明文を文書としてトピック抽出を行うと「ファンタジー」、「ビジネス」、「プログラミング」といった“本のジャンル”という観点に基づくトピックが得られる. 個々のトピックの内容からトピックがどのような観点を表しているかを把握することもできるが、より直感的に把握できる方法として、トピックに基づく類似度で平面上に文書を配置して可視化することを提案する. 可視化結果において、どういった内容のクラスタができていないか分析することで、トピックがどのような観点を表しているか推測する.

^{*1} 全角 140 文字以内、半角 280 文字以内の投稿文.

本章では、トピックを構成する上位単語を並べる方法と可視化の2通りの方法でトピック内容の分析を行う。可視化の分析については、有名人間の関係の可視化という実験設定で、テキスト情報から可視化を行う方法を提案する。以上、それぞれの情報源におけるトピックを構成する上位単語および可視化結果を比較することでトピックの差異を分析し、ユーザに関する文書から抽出されるトピックを明確にすることを旨とする。

2.2 関連研究

2.2.1 SNS を利用したアイテム情報の抽出

Twitter や Facebook といった SNS 上で発信された情報を利用して、商品や店舗、人物に対する評判情報の抽出や意見・感想の要約を行う研究には様々なものがある。飲食店やホテルに対する評判情報の分析 [48, 49] では、ソーシャルメディアのマーケティング効果が明らかになり、企業のマーケティング戦略に影響を与えている [50]。山本ら [51] は、テレビ番組を視聴しながら投稿される意見や感想の要約を行った。Meng ら [52] は、商品や人物に関する Twitter 上での意見要約を行うため、ハッシュタグで商品や人物に言及した tweet 群からトピックを抽出し、ポジティブとネガティブそれぞれの代表となる tweet を選択した。Park ら [53] は、実名公開型の SNS である Cyworld における政治家への一般ユーザのコメントから、セマンティックネットワーク解析および感情分析を用いてその政治家への集団感情を分析した。

2.2.2 SNS を利用したユーザ情報の抽出

意見や感想を投稿したユーザに焦点を当て、どういった人物であるかの推定やユーザ間の類似度を求める研究が行われている。年代、性別、社会的地位などのデモグラフィック情報を推定するために、日常的な投稿内容を解析する方法 [54] や、プロフィール画像と自己紹介文から推定した後で地域毎の人口統計データで補正をかける方法 [55] が提案

されている。Guy ら [56] は、社内 SNS における他者への評価を元にユーザ間の類似度を求めた。古賀ら [57] は Twitter において、ある対象ユーザのリツイートやフォロー関係などを文書として LDA を行い、ユーザの興味や嗜好に関するトピックに注目してユーザ間の類似度を求めた。

2.2.3 ユーザ間の関係の可視化

藤本ら [58] は、ユーザの Web 閲覧行動には潜在的なトピックが存在すると仮定し、サーバに残ったアクセスログの URL を階層型 URL 辞書で集約した後 LDA を行い、抽出したユーザ層を可視化した。人物関係の情報を提示するための可視化技術に注目すると、松尾ら [59] の研究がある。彼らは学会における人間関係を自動的に抽出し、その関係を表すネットワークを可視化することで複雑な関係や研究者のクラスタを直観的に分かるように提示した。

2.3 実験方法

本実験で使用する情報源の種類ごとの特徴とトピックモデルの入力とするための処理を説明する。また、テキスト情報から可視化を行う方法を説明する。手順としては、以下に示す 3 つの処理を順に行う。

1. 前処理：tweet 中に含まれるノイズに対する処理および形態素解析を行う
2. 特徴量抽出：有名人毎のトピックに基づく特徴量を求める
3. 可視化：有名人間の類似関係を 2 次元平面上に可視化する

2.3.1 情報源の種類

本実験で比較を行う情報源をそれぞれ説明する。

- 情報源 A：有名人に関して言及したユーザの tweet

対象有名人に対するユーザの率直な発言が見られ、ユーザが持つ対象有名人に対する印象・感想という観点のトピックが抽出されると仮定する。トピックモデルの入力とするための処理は以下となる。

1. 対象有名人 i の名称を含む tweet を集めて一つのテキストにマージする
2. このテキストに対して前処理を行い、得られた単語集合を文書 d_i とし、文書 d_i の集合を D とする

- 情報源 B：有名人に関して言及したユーザのプロフィール

tweet を行ったユーザの社会的情報や趣味・興味が見られ、対象有名人に関心を持っているユーザ層という観点のトピックが抽出されると仮定する。トピックモデルの入力とするための処理は以下となる。

1. 対象有名人 i の名称を含む tweet (情報源 A) を発信したユーザの集合を作成する
2. この集合中の各ユーザのプロフィール文^{*2}に対して前処理を行い、得られた単語集合を文書 d_i とし、文書 d_i の集合を D とする

- 情報源 C：有名人本人が発信した tweet

有名人はファンを意識した発言を行うと考えられるため、対象有名人がどう見られたいかという観点のトピックが抽出されると仮定する。トピックモデルの入力とするための処理は以下となる。

1. 対象有名人 i が発信した tweet を集めて1つのテキストにする
2. このテキストに対して前処理を行い、得られた単語集合を文書 d_i とし、文書 d_i の集合を D とする。

*2 ユーザが自由に記入できる 160 文字以内の自己紹介文。

2.3.2 前処理

収集した tweet 群の中には、テキスト中の一部に外部のサイト名や blog の記事名といった自動で付与された定型文であるものが多数存在する。こういったノイズとなるテキスト（以降、ノイズテキストと呼ぶ）が多く含まれてしまうと、ユーザの tweet 中に含まれる印象・感想といった情報の抽出が妨げられるため除去する必要がある。情報源 A, C に対しては、2.3.2.1 で説明するノイズテキスト除去処理を適用する。

次の特徴量抽出で用いるトピックモデルの入力である文書は、テキストを単語単位に分解して語順を無視した bag-of-words 形式である。この文書を作成する処理を 2.3.2.2 で説明する。

2.3.2.1 ノイズテキスト除去処理

ノイズテキストの特徴として、URL を伴い、同じ内容の tweet が多数発信されていることが多い。そこで、URL を伴う tweet において、2 つ以上の tweet に共通して出現するテキスト部分を特定し、その重複箇所だけを除去することにする。以下に具体的な除去方法を示す。

- (1) URL を含む tweet の URL より前のテキスト全てと後ろのテキスト全てを抽出し、ノイズテキスト候補集合（集合 A）を作成する。また、空の集合 R を用意する。
- (2) 集合 A から文字列（要素 e ）を一つ選択し、その文字列の中心 N 文字を抽出する（文字列の中心位置から、 $N/2$ 文字分手前の文字列と、 $N/2$ 文字分後ろの文字列を抽出する）。ただし、文字数が N 文字未満の場合は、その要素 e を集合 A から取り除いて次の要素の判定 (2) へ移る。なお、 N の値は小さ過ぎるとあらゆるテキストをノイズテキストと誤判定してしまい、大き過ぎるとノイズテキストを特定できないため、ノイズテキストの特定結果を見て調整を行う。2.4 節の実験では $N = 6$ に設定した。

- (3) 集合 A において、要素 e 以外で (2) で抽出した文字列を文中に含む要素を探す。存在する場合、そのマッチした要素の集合 (集合 B) を作りノイズテキストの同定 (4) へ進む。存在しない場合、この要素 e を集合 A から取り除いて次の要素の判定 (2) へ移る。
- (4) ノイズテキストの同定を行う。まずはノイズテキストの左端を見つける。切り取る範囲を要素 e の中心 N 文字から左側に広げ、集合 B 中の各要素と部分的にマッチするか調べる。この操作を集合 B 中のどの要素ともマッチしなくなるまで続け、ノイズテキストの左端を特定する。同様に、右端も探索する。
- (5) 特定したノイズテキストを集合 R に追加し、集合 A の次の要素の判定 (2) へ移る。すべての要素に対して判定を終えた場合、(6) へ進む。
- (6) 集合 R の要素を文字数の大きい順番にソートし、オリジナルの tweet 集合中の tweet において各要素とマッチする文字列を除去する。この除去処理を行った tweet 集合を以降用いる。

上記の処理とは別に、全ての tweet から URL とハッシュタグ^{*3}の除去も行う。ハッシュタグは tweet の内容を捉えるのに有益であるが、スパムの中には tweet 内容とは関係の無いハッシュタグを複数記述したものが多いため除去の対象とした。また、URL を含んでいない場合でも、ある程度の文字数がある tweet で完全一致するものが複数ある場合はスパム投稿である可能性が高く、tweet の文字数が M 文字以上^{*4}で文字列が完全一致するものは除去する。

2.3.2.2 文書の作成

各テキストに対して、形態素解析ツール MeCab[78] を利用して、単語単位への分解、および単語を原形に揃える処理を行う。この際、Mecab で固有名詞や専門用語を扱えるようにするため、MeCab で利用する辞書に以下を登録する処理を行った。

^{*3} #から始まる文字列で、話題や主張を表すキーワード。投稿者が自分で付与する。

^{*4} 本研究では、 $M = 42$ (tweet の文字数上限の 3 割) に設定した

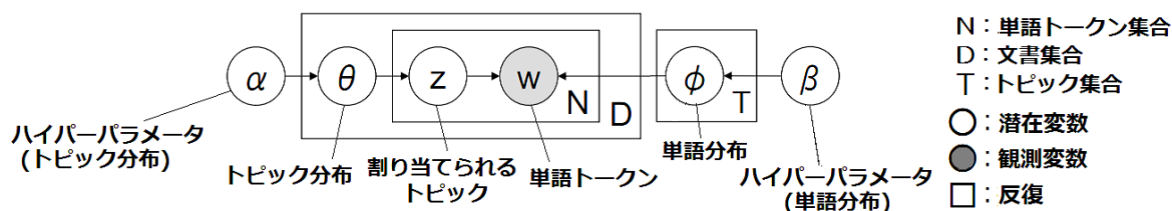


図 2.1 Latent Dirichlet Allocation (LDA) のグラフィカルモデル

- Wikipedia に存在する項目名
- 専門用語 (キーワード) 自動抽出モジュール TermExtract[79] で抽出した, tweet によく出現する専門用語やフレーズ

有名人 i に関するテキストを単語単位に分解した後, 品詞が名詞・形容詞・動詞と判定された単語のみを原形で抽出し, bag-of-words 形式の文書 d_i を作成する. 全ての有名人に対してこの処理を行って得られた d_i の集合を文書集合 D とする.

2.3.3 特徴量抽出

文書集合 D を入力とし, 各有名人 (文書 d_i と対応している. 以降, 添字の i を省略して d と表す) のトピックに基づく特徴量を求める. 文書毎の特徴量を求める手法としては, 単語の出現頻度と稀少度を考慮した TF-IDF を算出する方法が一般的によく用いられている. ただし, この方法では文書中の完全一致する単語しか考慮されておらず, 類似した意味の語の影響を反映させることができない. そこで, 文書の背後に存在するトピックを考慮したソフトクラスタリングを行う手法であるトピックモデルを用いる. 本研究では, トピックモデルの一種である Latent Dirichlet Allocation (LDA) を用いて, 文書毎のトピック比率を推定する.

2.3.3.1 LDA

LDA は, Blei らにより提案された確率的トピックモデルである [40]. トピックモデルとは, ある文書が単一または複数のトピックに属する単語から構成されているという仮

定をおき、その文書を構成するトピックの比率（トピック比率） θ と、それらのトピック毎の単語生成確率（単語比率） ϕ に基づき確率的に文書を生成するモデルである。文書 $d \in D$ は N_d 個の単語（トークン）列 w_d で構成される。 w_d の n 番目に現れたトークンを w_{dn} と表す。全文書の w_{dn} を集め、その中でユニークな単語の集合を語彙 W とする。また、各文書のトークン数を N_d 、全トークン数（ N_d の和）を N 、トピック数を K とする。文書 d のトピック比率は、 $\theta_d = (\theta_{d1}, \theta_{d2}, \dots, \theta_{d|K|})$ の確率分布であり、 $0 \leq \theta_{dk} \leq 1$ 、 $\sum_{k=1}^K \theta_{dk} = 1$ である。トピック k の単語比率は、 $\phi_k = (\phi_{k1}, \phi_{k2}, \dots, \phi_{k|W|})$ の確率分布であり、 $0 \leq \phi_{kw} \leq 1$ 、 $\sum_{w=1}^W \phi_{kw} = 1$ である。以下に LDA の生成過程を示す。

1. トピック毎に、ディリクレ分布 $Dir(\beta)$ から ϕ_k を生成。
2. 文書毎に，
 - a. ディリクレ分布 $Dir(\alpha)$ から θ_d を生成。
 - b. 文書中の各トークン毎に，
 - i. 多項分布 $Mult(\theta_d)$ から、トークン w_{dn} に割り当てられるトピック z_{dn} を生成。
 - ii. 多項分布 $Mult(\phi_{z_{dn}, w})$ から、トークン w_{dn} を生成。

上記の内容をグラフィカルモデルで表すと、図 2.1 のようになる。入力として与える必要があるのは、文書集合 D 、トピック数 K 、反復回数、 θ と ϕ の事前分布であるディリクレ分布のハイパーパラメータ α と β である。ここで、ハイパーパラメータ α と β は一般にベクトルであるが、Griffiths に従いすべてのベクトル要素を同じ値に設定する [61]。

次に、与えられた文書集合 D からパラメータ θ 、 ϕ の推定を行う方法を説明する。推定手法としては、Blei らが用いた変分ベイズ法による手法 [40] の他に、Griffiths らが用いた Gibbs Sampling による手法 [61] がよく利用される。この Gibbs Sampling は、現在の状態から次の状態を決定する際に 1 変数ずつサンプリングして更新していく手法であり、LDA においては現在のパラメータに基づいてトークンに割り当てられるトピック z_{dn} を

一つずつ順番に更新していく処理を行う。全トークンの更新を1反復とみなし、この反復回数が十分であれば真の定常分布からサンプリングを行うのと同じ結果をもたらすため、反復を重ねれば重ねるほど変分ベイズ法よりも良い結果を示すことが分かっている [43, 61]。ここで、文書集合をすべて直列化して一つのトークン列にした際のトークンの添字を j 、 j 番目のトークンのトピックを z_j 、 j 番目のトークンである単語を w_j 、 j 番目のトークンが含まれる文書を d_j と表す。Gibbs Sampling による j 番目のトークンのトピックの更新は、式 (2.3.1) に示す分布からのサンプリングで行う。

$$P(z_j = k \mid \mathbf{z}_{\setminus j}, \mathbf{w}) \propto \frac{N_{\setminus j,k}^{w_j} + \beta}{N_{\setminus j,k}^{(\cdot)} + W\beta} \cdot \frac{N_{\setminus j,k}^{d_j} + \alpha}{N_d - 1 + K\alpha} \dots\dots\dots (2.3.1)$$

ただし、 $\setminus j$ は集合から j だけを除いた差集合を、 (\cdot) 記号はその添字について総和をとることを意味している。詳しくは、 $N_{\setminus j,k}^{w_j}$ は単語 w_j において j 番目以外でトピック k が割り当てられたトークン数、 $N_{\setminus j,k}^{(\cdot)}$ は j 番目以外でトピック k が割り当てられたトークン数、 $N_{\setminus j,k}^{d_j}$ は文書 d_j において j 番目以外でトピック k が割り当てられたトークン数を表している。

全トークンのトピックを更新して得られた新たなトピック $\hat{\mathbf{z}}$ から、パラメータの推定量 $\hat{\theta}_{dk}$ と $\hat{\phi}_{kw}$ はそれぞれ式 (2.3.2) と式 (2.3.3) で求められる。

$$\hat{\theta}_{dk} = \frac{N_k^d + \alpha}{N_d + K\alpha} \dots\dots\dots (2.3.2)$$

$$\hat{\phi}_{kw} = \frac{N_k^w + \beta}{N_k^{(\cdot)} + W\beta} \dots\dots\dots (2.3.3)$$

2.3.4 可視化

有名人毎の特徴量を用いて、有名人間の類似関係を2次元平面上の配置により表現する。可視化手法としては、主成分分析 (PCA) で求めた情報量が最大になる2軸に点を射影する手法、距離尺度を表現する一般的な手法である多次元尺度構成法 (MDS)、ニュー

ラルネットワークに基づき平面へ要素の割り当てを決める自己組織化写像 (SOM) の3つをそれぞれ用いて結果の比較を行う。ただし、MDS に関しては特徴量をそのまま用いることはできないので、距離行列に変換してから実行する。

2.3.4.1 主成分分析

主成分分析 (PCA) により、トピックに基づく有名人間の関係を最も表現できる2軸を求める。2.3.3 で求めたトピックを元の軸とすると、PCA は複数のトピックの影響を反映した有名人間の違いを表すのに効果的な新たな軸を発見する。また、PCA は各軸に対して意味的な解釈を行うことが可能であり、何らかの観点に基づく要約が行われることが期待できる。これはユーザへ提示する際に、ユーザの提示内容に対する理解を促進する効果があると予想される。

2.3.4.2 多次元尺度構成法

多次元尺度構成法 (MDS) は、2者間の距離尺度を維持した配置を求める方法であり、PCA とは異なり配置した空間における軸には意味的な解釈が存在しない。MDS は入力に要素間の距離行列が必要であり、まず 2.3.3 で求めたトピック比率から距離行列を算出する。今回用いるのは、古典的多次元尺度構成法 (CMDS) と呼ばれる最も基本的な MDS 手法である。距離行列を算出する方法には、次に説明する JS 情報量を用いる。

JS 情報量

確率分布の類似度を測る手法としては、比較対象の一方の確率分布の値に0が存在しないことを制約条件に持つ Kullback-Leibler divergence (KL 情報量) や、2つの平均的な確率分布までの KL 情報量の平均を求める Jensen-Shannon divergence (JS 情報量) などがある。本研究では、比較の対称性を満たす JS 情報量を用いて類似度の比較を行う。ここで、対象 A と B のトピック比率 θ_A, θ_B が与えられたとき、JS 情報量は式 (2.3.4) で求める。

$$D_{JS}(A, B) = \frac{1}{2} \left(\sum_{k=1}^{|K|} \theta_{Ak} \log \frac{\theta_{Ak}}{\theta_{Rk}} + \sum_{k=1}^{|K|} \theta_{Bk} \log \frac{\theta_{Bk}}{\theta_{Rk}} \right) \dots\dots\dots (2.3.4)$$

ただし、 $\theta_{Rk} = \frac{1}{2} (\theta_{Ak} + \theta_{Bk})$ である。この D_{JS} は、値が小さいほど対象間の類似度が高いことを意味する。

2.3.4.3 自己組織化写像

自己組織化写像 (SOM) は、ニューラルネットワークによる教師なし学習の一種であり、類似した特徴を持つ入力同士を近くに配置するような出力を行う。入力は 2.3.3 で求めたトピック比率をそのままベクトルとして用いる。ニューラルネットワークである SOM は、細かな手法のバラエティや多数のパラメータが存在し、これらの設定により結果が変化する。

2.3.5 データセット・ハイパーパラメータ

2.3.5.1 データセット

表 2.1 に示す有名人に対して、Twitter 公式 API を利用して情報源 A, B, C のデータを収集した。実験対象となる有名人は、有名人ツイッターランキング^{*5}を参考に、政治・芸能・経済・スポーツ・文化人など様々な分野から満遍なく選択した。収集するデータの詳細は以下である。

- 情報源 A：有名人の名称・呼称^{*6}をクエリとして検索して得られた 2013 年 9 月 2 日から 8 日までの 1 週間分の tweet
- 情報源 B：情報源 A で収集した tweet を行ったユーザーの 2013 年 9 月時点でのプロフィール
- 情報源 C：有名人の 2013 年 9 月時点から最大 3200 まで遡った tweet

^{*5} www.talentswit.com

^{*6} Wikipedia に登録された名称に加え、Wikipedia の本人記事中に記載された愛称・略称も用いる

表 2.1 実験対象となる有名人の一覧（順不同）

(a) 一覧 (1/3)	(b) 一覧 (2/3)	(c) 一覧 (3/3)
春名風花	鬼龍院翔	橋下徹
宮迫博之	喜矢武豊	孫正義
松本人志	宇治原史規	香川真司
スギちゃん	伊集院光	ダルビッシュ有
山本太郎	篠田麻里子	東浩紀
ローラ	平野綾	乙武洋匡
剛力彩芽	中川翔子	茂木健一郎
西川貴教	堀江貴文	田原総一朗
GACKT	安倍晋三	前山田健一（ヒャダイン）
きゃりーぱみゅぱみゅ	東国原英夫	太田順也（ZUN）

2.3.5.2 ハイパーパラメータ

LDA のトピック数 K は、[60] の手法により形成されたクリーク数である 17 に設定した。LDA の他のハイパーパラメータは、Blei ら [40] と同様に $\alpha = \frac{50}{K}$, $\beta = 0.1$ とし、Gibbs Sampling の反復回数は 1000 に設定した。SOM に関しては、 30×30 の正方形格子、逐次型学習、反復回数は 1 万回、近傍関数はガウス関数、ベクトル間の比較はコサイン類似度、参照ベクトルの初期値は乱数で設定した。

2.4 結果と考察

3 種類の可視化手法に対する可視化結果の比較を行った後、最も良かった可視化手法を用いて、3 種類の情報源に対する可視化結果の特徴の分析を行う。分析を行う際の基準として、可視化結果において、ある程度事実と則しており一般的に納得できる集団が存在するか（妥当性）、有名人間の意外な関係を表す集団を見つけられるか（発見性）という 2 点に着目する。CTR による推薦にどう反映されるかという観点で考えると、可視化対象

となる情報源に対して次のように解釈できる。トピックの内容は予測評価値を求めるための特徴量である潜在因子ベクトルに反映されるため、トピックの内容が妥当で無ければ予測評価値は信用できないものとなる。そのため、妥当性を満たす集団は少なくとも1つは存在することが望まれ、集団の数が多いほど有用なトピックが豊富な情報源であると考えられる。トピックの内容に発見性がある場合、その情報源を用いることで発見性を持たない情報源では行うことができない真新しい推薦を行える可能性がある。他の情報源では得られない集団が存在するなら、その集団に関連するトピックが特徴量に反映されることで、他の情報源を用いた場合では行えない推薦が可能になると考えられる。

本実験において、可視化結果に対しての評価は筆者が定性的に行う。そのため主観的な判定が含まれてしまうが、妥当性と発見性の判定基準は明確にしておく。妥当性については、Wikipedia やブログポータルサイトなどで行われている芸能人や有名人に関する分類（たとえば、「お笑いタレント」、「俳優」、「声優」、「サッカー選手」など）を反映している密な集団であるかを判定の基準とした。発見性については、上記の有名人に関する分類以外の理由で説明付することができる密な集団であるかを判定の基準とした。なお、本実験では可視化結果のスクリーンショットを示し、その中で発見されたグループ（密に集まった有名人集合）に注目するが、そのグループはスクリーンショット上で四角で囲んで、また (a)~(e) の記号を振って示すことにする。各グループを本文中で参照するときには、(a) のように表記する。

2.4.1 可視化手法毎の結果

情報源を B に固定し、可視化手法毎の可視化結果における特徴を調べる。なお、この可視化結果は情報源 B を用いているので、その有名人に関心を持っているユーザー層に基づいて類似度を求めていることになる。

2.4.1.1 PCA

PCAによる可視化結果を図2.2に示す。ただし、PCAの可視化結果は離れた位置に大きく3つの集団が現れる結果となっており、紙面の都合上その中間部分を省略（白抜きで表示）している。PCAによる可視化結果の主な特徴として、縦横の軸に意味が存在しているため集団の解釈がしやすいという利点がある。今回の結果では、縦軸がカジュアル・フォーマルを表していることが分かる一方、横軸の解釈は困難であった。妥当性に関しては、やや外れ値が見られるが、ユーザの興味は上部(a)の政治家と下部(c)の芸能人に分けることができている。発見性に関しては、中央(b)に知識系芸能人、下部(d)にネットをよく利用するユーザや(e)にヴィジュアル系バンド（V系バンド）といった特定のユーザ層に人気がある有名人が見られる。

2.4.1.2 MDS

MDSによる可視化結果を図2.3に示す。基本的にPCAと似たような集団が見られる。有名人が集中している部分はより密集している傾向が見られ、PCAよりも集団が明確に表れている。特に、図2.2のPCAの結果では、芸能人の塊がややぼらけているが、MDSではこれらが密になっている。妥当性で言うと、外れ値が少ない点でPCAよりも優れていると言える。発見性に関しては、中央やや右上(b)に知識系芸能人の集団、中央やや下部(d)にネット上での有名人の集団、左下(e)にV系バンドが見られる。

2.4.1.3 SOM

PCAによる可視化結果を図2.4に示す。SOMによる可視化結果の主な特徴として、格子状の座標へ飛び石のように有名人が配置される（図2.4に格子は表示していない）。妥当性に関しては、上部(a)の政治家の集団が見られる。中央やや右下(b)の密集した部分には芸能人が多く集まっているが、田原総一郎と孫正義といった評論家・実業家、他にもネット上の有名人が含まれており、上手く分離できていない。下部(c)のV系バンド

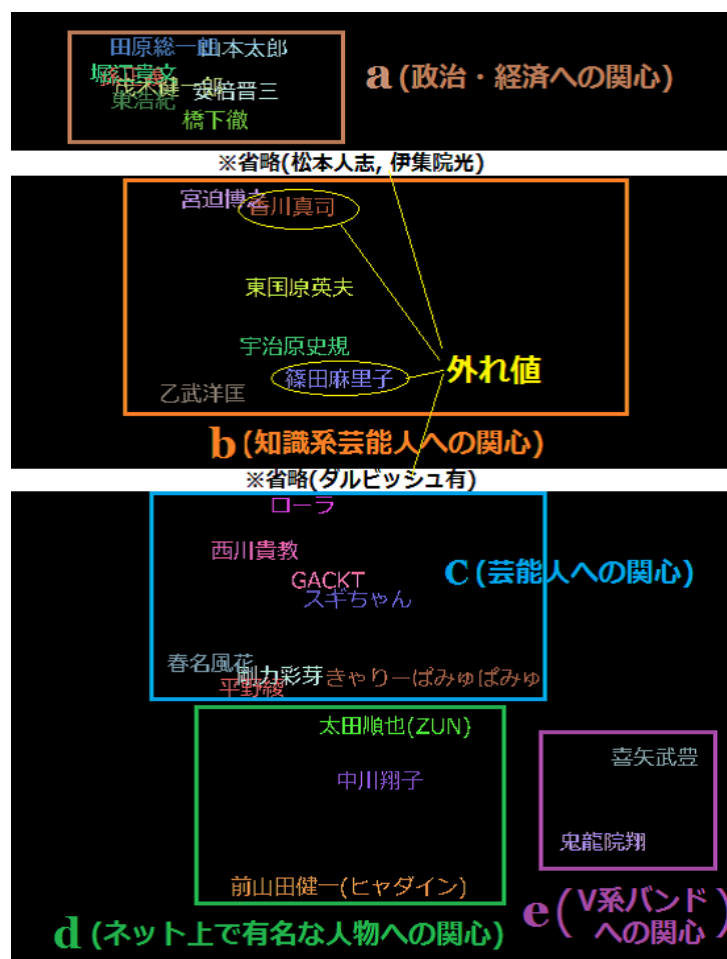


図 2.2 情報源 B（有名人に関して言及したユーザのプロフィール） - PCA

の集団の中に乙武洋匡が現れているのは妥当ではない。加えて、下部 (b) の集団に含まれていた方がよい芸能人として、PCA・MDS と比較してきゃりーぱみゅぱみゅ、MDS と比較して篠田麻里子が (b) の集団から孤立している点も妥当性に疑問を残す要因となっている。発見性に関しては、PCA や MDS と比べて、下部 (c) に V 系バンドの有名人がより多く現れているが、知識系芸能人やネット上の有名人といったグループは見られなかった。

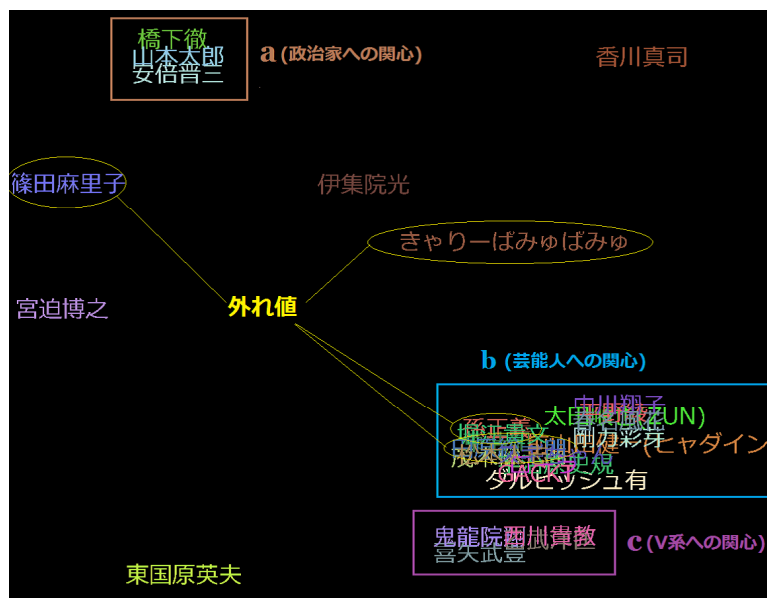


図 2.4 情報源 B（有名人に関して言及したユーザのプロフィール） - SOM

表 2.2 可視化手法毎の結果まとめ

可視化手法	PCA	MDS	SOM
妥当性	○	◎	△
発見性	◎	◎	△

2.4.2 情報源毎の結果

2.4.1 節の結果より，可視化手法を最も優れていた MDS に固定し，情報源毎の可視化結果における特徴を調べる。

2.4.2.1 情報源 A：有名人に関して言及したユーザの tweet

情報源 A はユーザが持つ対象有名人に対する印象・感想という観点のトピックに注目しており，このため時事ニュースや有名人に対する現在の世論の影響を強く受ける傾向がある。表 2.3 に得られたトピックの上位単語を示す。トピックの種類としては，好意的・批判的といったポジティブ・ネガティブを表すものや，アイドル・俳優・お笑い芸人・政

表 2.3 主なトピック（情報源 A：有名人に関して言及したユーザの tweet）

トピック解釈	好意的	批判的	男性有名人	AKB	ゆるキャラ
上位単語	てる	はるかぜちゃん	山田孝之	卒業	熊本
	する	菜々緒	城田優	似る	可愛い
	好き	大人	真島ヒロ	板野友美	(笑)
	見る	叩く	映画	笑	くまモン
	歌う	匿名	かっこいい	AKB	ふる
	可愛い	批判	イケメン	篠田麻里子	グッズ
トピック解釈	政治（政府）	政治（大阪維新）	映画「ガッチャマン」		
上位単語	安倍総理	橋本	誕生日		
	いる	蓮舫	剛力		
	日本	市長	可愛い		
	総理	維新	綾野剛		
	麻生太郎	批判	ゴーリキー		
	首相	大阪	実写化		

治・ゆるキャラなど話題に関連するグループ、時事ニュースがトピックとして見られた。次に、可視化結果を図 2.5 に示す。妥当性に関しては、図 2.5 の上部 (a) に俳優・アイドル・芸人を中心とする芸能人、中央やや上部 (b) にスポーツ選手、下部 (d) に政治家・評論家・実業家が集まっておりが集まっている。(a) の部分に関して、一部に芸能人ではないネット上の有名人が含まれているものの、この配置は多くのユーザにも理解されるのではないかと思われる。発見性に関しては、中央やや下部 (c) にバラエティ番組にも出演するが世間からは知識人と捉えられている知識系芸能人が集まっている。

2.4.2.2 情報源 B：有名人に関して言及したユーザのプロフィール

情報源 B は有名人に言及したユーザのプロフィールであり、同じユーザ層に関心を持たれている有名人同士が近くに配置されることになる。表 2.4 に得られたトピックの上位単語を示す。トピックの種類としては、アイドルやスポーツ、学問などユーザが興味を持っている対象が主に得られ、他には学生や BOT のようなユーザ自身が何者であるかを

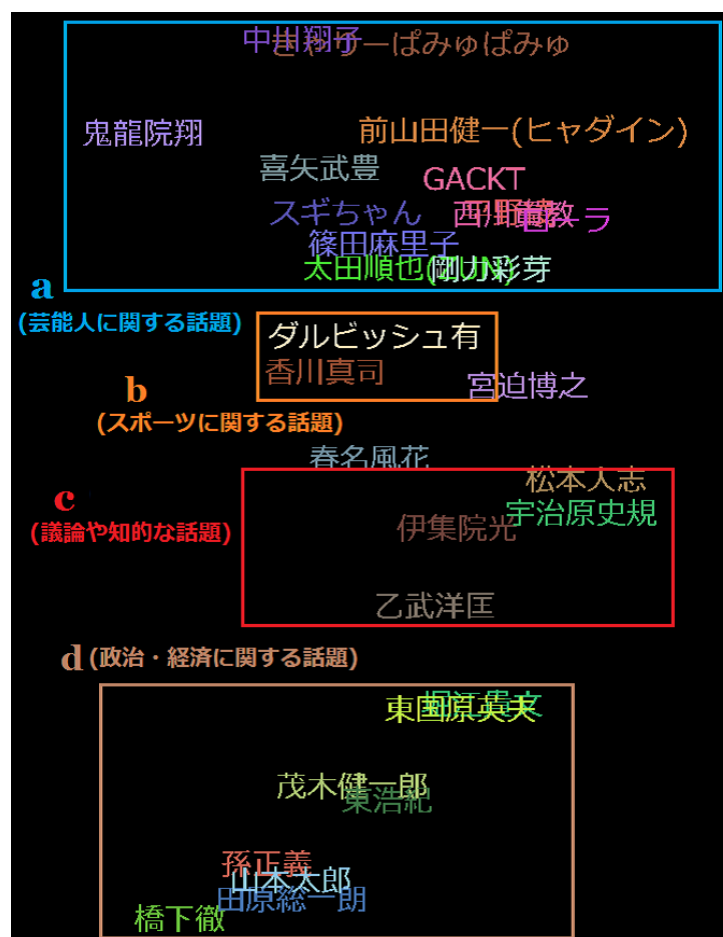


図 2.5 情報源 A（有名人に関して言及したユーザの tweet） - MDS

表すトピックが得られた。次に、可視化結果を図 2.3 に示す。妥当性に関しては、図 2.3 の上部 (a) に政治家・評論家・実業家が集まっており、中央 (c) に芸能人が集まる結果となっている。情報源 A では周囲から分離していたスポーツ選手の集団は (c) の付近に見られ、はっきりと分離していないものの互いに近い位置に配置されている。発見性に関しては、情報源 A と同様に中央やや右上 (b) に知識系芸能人が集まっている。中央やや下部 (d) には、ネット上での有名人の集団が表れている点が興味深い。下部 (e) には V 系バンドであるゴールデンボンバーの 2 人が他の芸能人や音楽関係者とは大きく隔離した位置にあり、彼らに関心を持つユーザは特異であることが分かる。また、情報源 A に比

表 2.4 主なトピック（情報源 B：有名人に関して言及したユーザのプロフィール）

トピック解釈	ジャニ系	アイドル(ももクロ)	アイドル (AKB)	V系バンド	スポーツ
上位単語	嵐	推す	AKB	金爆	サッカー
	NEWS	ももクロ	推す	ゴールデンボンバー	野球
	組	モノノフ	前田敦子	V系	香川真司
	二宮	箱	篠田麻里子	ギルド	選手
	相葉	あーりん	SKE48	シド	MLB
	潤	しおりん	NMB48	己龍	SAMURAI
トピック解釈	アニメ・ゲーム	学問	政治・経済	学生	BOT
上位単語	アニメ	物理	日本	好き	紹介
	好き	伊織	政治	フォロー	名言
	ゲーム	ぼるぼる	反対	音楽	面白い
	ボカロ	村上春樹	脱原発	気軽	よろしく申し上げます
	漫画	文学	TPP	野球	ニュース
	声優	ゲンロン	日本人	高校	2ch

表 2.5 主なトピック（情報源 C：有名人本人の tweet）

トピック解釈	イベント情報	ブログ情報	政治	モデル撮影
トピック 上位単語	今日	今日	Blog	今日
	ありがとう	ありがとうございます	Official	明日
	ライブ	ブログ	自民党	ありがとう
	今夜	更新	応援	撮影
	頑張る	QT	県	みんな
	皆様	頑張る	議員	終わる

べて、四角で囲んだ部分の有名人がより密集している傾向が見られる。例えば、(a) の政治家・評論家・実業家の塊や、(b) の知識系芸能人の塊は、情報源 A よりも密集していることが分かる。(d) のネット上での有名人の集合も密な配置となっている。このように、情報源 B は情報源 A よりも集団がより密になる傾向があり、グループの発見が容易になる利点がある。加えて、得られた集団の種類は情報源 A に比べて豊富であり、より多くの発見が可能であることを示している。



図 2.6 情報源 C (有名人本人の tweet) - MDS

2.4.2.3 情報源 C：有名人本人が発信した tweet

情報源 C は有名人本人の tweet 内容を反映している。表 2.5 に得られたトピックの上位単語を示す。トピックの種類としては、政治・モデル撮影といった有名人本人の職業に関するトピックの他、ライブイベントの告知やブログ更新の告知といったトピックが見られたが、それ以外のトピックは解釈できないものが多い結果となった。次に、可視化結果を図 2.6 に示す。妥当性に関しては、図 2.6 の上部 (a) に政治家・実業家が集まっており、下部 (c) に芸能人が集まる結果となっている。スポーツ選手や V 系バンドの有名人は (c) に含まれているものの、それらの有名人同士は近くに配置されている。ただし、外れ値と思われるものもいくつかあり、剛力彩芽 (tweet 数が極端に少ないため) や中川翔子 (アニメ、ゲームの話題が多いため) などが目立って現れた。加えて、(c) の部分を詳しく見

表 2.6 情報源毎の結果まとめ

情報源	A (有名人に関して言及したユーザの tweet)	B (有名人に関して言及したユーザのプロフィール)	C (有名人本人の tweet)
主なトピック	話題内容, 感情極性	ユーザの趣味・興味	有名人の仕事内容, 広報
妥当性	◎	◎	○
発見性	△	◎	△

ると、伊集院光・松本人志といった知識系芸能人の集団である (b) に含まれるべき有名人が見られ、情報源 A,B と比べて妥当性はやや低い結果といえる。発見性に関しては、中央部 (b) に作家・評論家と知識系芸能人の集団が見られ、情報源 A と同等といえる。

2.4.2.4 情報源毎の結果まとめ

表 2.6 に情報源毎の結果の要約を示す。有名人の集団が既知の分類通りであるかを表す妥当性に関して、外れ値が存在せず加点要素があるものは◎、いくつか外れ値が存在するものは○で優劣を表記する。加点要素としては、情報源 A では他の情報源では見られなかったスポーツに関する集団が見られたこと、情報源 B では集団がより密になる傾向がありグループがより明確に分離できることを挙げる。よって妥当性に関しては、情報源 A と情報源 B がどちらも同等に優れている結果となった。情報源 A “有名人に関して言及したユーザの tweet” では言及する有名人に関連する内容からかけ離れることはないため、例えばスポーツ選手に対してはスポーツの話題がなされており、妥当性が高くなるのは納得のいく結果である。情報源 B “有名人に関して言及したユーザのプロフィール” では集団がより密になる傾向があることから、ユーザ層によって好まれる有名人の傾向が明確にあるのだと考えられる。

興味深い集団であるかを表す発見性に関して、該当する集団が3つ見られたものは◎、1つだけ見られたものは△で優劣を表記する。発見性に関しては、情報源 B が最も優れており、情報源 A と情報源 C は劣る結果となった。情報源 B “有名人に関して言及したユーザのプロフィール” の結果から、ネット上での有名人やヴィジュアル系を好むユーザ

層は明確に区別できることが分かる。

2.5 結論

本章では、Twitter から取得したデータを分析することで、ユーザに関する文書から抽出したトピックへの理解を深めるための分析を行った。個々のトピック内容およびトピックがどのような観点を表しているかを把握するためにトピックを構成する上位単語の分析、更にトピックがどのような観点を表しているかを直感的に把握するために可視化結果の分析を採用した。トピックの抽出元である情報源として、アイテムに関する文書、ユーザに関する文書、アイテムに対する擬似的なレビュー文書を用意し、トピックを構成する上位単語および可視化結果を比較することで情報源毎のトピックの差異を分析した。

トピックを構成する上位単語の分析においては、ユーザに関する文書からは、ユーザが興味を持っている対象やユーザの属性を表すトピックが明確に得られることが分かった。ユーザに関する文書から抽出された個々のトピック内容は、ユーザが興味を持っている対象やユーザの属性ということから、これらはユーザ層という観点を表していると考えられる。可視化のタスクにおいては、多くのノイズが含まれたテキストに対して前処理を行い、各情報源から得られるトピックに基づいた有名人毎の特徴量を算出し、これを元に2次元平面上での配置を決定した。この可視化結果に対して、妥当性と発見性という2つの観点で、可視化手法毎および情報源毎に特徴の分析を行った。可視化手法の中では、2者間の距離尺度を維持した配置を求める方法である MDS が最も妥当性を保ち、多くの発見を行えることが分かった。そして MDS による可視化を用いて情報源の比較を行うことで、ユーザに関する文書を利用した場合は妥当性を保ちつつ、知識人として捉えられている芸能人やネット上で人気の有名人といった発見性のある集団を抽出できることが分かった。また、ユーザに関する文書を利用した場合の可視化結果で見られた集団は、どういう興味を持ったユーザに関心を持たれているか、すなわちユーザ層という観点を表していると考えられる。

以上より、ユーザに関する文書からは、ユーザ層という観点を表すトピックが明確に得られることを示した。また、ユーザに関する文書から抽出したトピックはアイテムに関する文書から抽出したトピックに比べて妥当性だけでなく発見性も優れていたことから、3章におけるCTRのトピック抽出のための新たな情報源として有益である可能性が高いと考えられる。

第 3 章

トピック抽出元となる文書の種類が 推薦性能に与える影響の分析

本章では、1.2.2 節で述べた課題 1 に対する取り組みとして、ユーザに関する文書を用いる場合の推薦性能は従来のアイテムに関する文書を用いる場合と比べてどのような差異が生じるか分析する。2 章より、ユーザに関する文書からは、主にユーザが興味を持っている対象を表すトピックが得られることが判明した。これはアイテムに関する文書からは得られないトピックであるため、トピックの影響を受ける CTR の推薦結果も両者で異なることが予測される。

本章の構成は以下となる。研究の背景と動機となる考えや仮定および本章の目的を 3.1 節で述べる。関連研究として、アイテムやユーザに関するサイド情報を利用する研究と、CTR に関連のある研究を 3.2 節で示す。アイテムとユーザそれぞれの文書を利用する CTR モデルの説明を 3.3 節で行う。実験で使用するデータセットの説明やハイパーパラメータの設定、評価指標など実験設定を 3.4 節で説明する。実験結果の提示と考察を 3.5 節で行う。実験結果から導かれる結論と課題を 3.6 節で述べる。

3.1 背景と目的

Collaborative Topic Regression (CTR) [16] は、文書から抽出したトピックを利用するアプローチの先駆けとなったモデルである。CTR を提案した Wang ら [16] は、科学記事や論文をアイテムとした推薦において、トピックを抽出する情報源としてアイテムのタイトルやアブストラクトといったアイテムに関する文書を用いている。これらの文書からは、例えば“Bayesian Statistics”や“Computational Neuroscience”といったアイテム（科学記事や論文）の分野を表すトピックが得られる。CTR では、トピックをアイテムの潜在因子ベクトルに反映させるのでアイテムの潜在因子はアイテムの分野を表すことになり、またアイテムの潜在因子ベクトルとユーザの潜在因子ベクトルは共通の潜在因子を持っている。すなわち、アイテムに関する文書を用いる場合、アイテムの潜在因子ベクトル v_j はアイテム j がどの分野に属しているか、ユーザの潜在因子ベクトル u_i はユーザ i がどのアイテムの分野に興味があるかを表現していると解釈できる。一方で、ユーザに関するテキスト情報を用いることに関して、Wang らの論文 [16] では深く言及されていない。2章の分析より、ユーザに関する文書からはユーザ層という観点を表すトピックが明確に得られることが判明した。すなわち、ユーザに関する文書を用いる場合、アイテムの潜在因子ベクトル v_j はアイテム j がどのユーザ層から好まれるか、ユーザの潜在因子ベクトル u_i はユーザ i がどのユーザ層に属しているかを表現していると解釈できる。よって、アイテムに関する文書を用いた場合の CTR（以降、iCTR と呼ぶ）と、ユーザに関する文書を用いた場合の CTR（以降、uCTR と呼ぶ）は、潜在因子が表現する内容が大きく異なるので、推薦結果の違いが明確に表れると仮定する。また、評価値行列のスパース度が高い、すなわち評価値行列から得られる情報が少ないほど推薦性能に与えるトピックの影響は強くなると仮定する。

本章では、iCTR, uCTR, および MF を用いて科学記事・論文を推薦する実験を行い、評価指標を比較することで推薦性能への影響を分析する。評価指標としては、推薦結果

がユーザの嗜好に適しているかを評価する正確性に加えて、全体として特定の人気アイテムだけが推薦されるような偏った推薦結果になっていないかを評価する利便性を採用し、この2つの観点で評価を行う。また、評価値行列中の評価値が極端に欠損している場合や評価値が十分に存在する場合とで比較するため、評価値行列のスパース度を変化させた複数の設定で実験を行う。本章の貢献をまとめると以下のとおりである。

1. CTR モデルにおいて、アイテムに関する文書を用いた場合とユーザに関する文書を用いた場合を比較し、情報源の違いが推薦性能に与える影響について分析した。
2. 性能評価には正確性に関する評価指標に加えて、推薦の利便性を表す被覆率と多様性を表す評価指標も用いた。
3. 上記モデルの性能評価において、スパース度との組み合わせで傾向が変化するか分析した。

3.2 関連研究

3.2.1 MF とサイド情報

協調フィルタリングの一種である Matrix Factorization (MF) [9, 10] は、評価値行列をユーザの潜在因子行列とアイテムの潜在因子行列に分解し、この2つの行列の積をとることで評価値行列を近似する手法である。MF が採用される主な理由は、評価値行列に欠損値が含まれていても適用可能であり、ユーザとアイテムの潜在因子ベクトルの積をとることだけで欠損値を推定することができるからである。ただし、MF は少ない評価値からでも推定を行うことが可能であるが、極端に評価値が不足するような状況では性能が大きく低下し、これはコールドスタート問題 (Cold-Start Problem) と呼ばれている [1, 11, 14]。このコールドスタート問題を緩和するには、評価値情報に加えてサイド情報 (アイテムやユーザに関する補助的情報) を利用することが有効である [14, 31]。サイド情報の例としては、ユーザのプロフィールやアイテムに対する説明文、アイテムに対す

るタグ情報やレビュー文，ソーシャルネットワークから得られるリンク情報などがある。Zhen ら [32] はユーザがアイテムに対して付与したタグを元にユーザ間の類似度を求め，それをユーザの潜在因子ベクトルに対する制約としてモデルを作成する手法を提案した。Bauman ら [33] はレビュー文からユーザが着目している要素とその極性を抽出して特徴量ベクトルとし，潜在因子ベクトルと同時に用いる手法を提案した。ユーザのアイテムに対する評価という点で見ると，評価値が極端に少ないようなアイテムがレビュー文を持つ可能性は低いという問題は残る。

3.2.2 MF とトピックモデルの融合

サイド情報から潜在因子を抽出する方法として，MF とトピックモデル [40] を組み合わせるアプローチが多く提案されている [16, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 29]。Collaborative Topic Regression (CTR) [16] はこのアプローチの先駆けとなったモデルである。CTR では，アイテムやユーザに関する文書がサイド情報として与えられ，トピックモデルによりそれらの文書から抽出したトピックを潜在因子に反映する。抽出したトピックに誘起して潜在因子ベクトルが生成されるため，評価値が全く存在しない新規アイテムに対しても，評価値が存在する他のアイテムとのトピックの類似関係を元に評価値を推定することができる。また，CTR [16] に似たモデルとして，Hidden factors as topics (HFT) [17] が提案されている。HFT は評価値とレビュー文はどちらもアイテムに対する評価であるため共通の潜在因子があると仮定することで，アイテムの潜在因子ベクトルとトピックとの関係式を定義した。HFT はトピックからアイテムの潜在因子ベクトルが生成されるという有向関係がない点が CTR と異なる。

サイド情報を取り入れるのにトピックモデルだけでなく，ソーシャルネットワークサービス (SNS) から得られるリンク情報やユーザ間の相互評価，SNS に関する研究によって得られた知見を元に CTR モデルを拡張するアプローチが提案されている。Ding ら [20] は，一般ユーザへの有名人の推薦において，アイテム間の関係すなわち有名人間の

フォロー関係という情報を取り入れるようモデルを拡張することで正確性を向上させた。Purushotham ら [21] は、ユーザはソーシャルネットワークを通じてコンテンツや意見を共有して影響し合っているという背景に注目し、音楽やブックマークの推薦において、ユーザ間のフォロー関係を取り入れるようモデルを拡張することで正確性を向上させた。Kang ら [22] は、ユーザの選択行動は嗜好だけでなく関係を持つ他のユーザの影響を受けることをモデル構造に組み込み、フォロー関係の情報を利用することなく、ソーシャルニュースサイトにおけるニュース記事の推薦の正確性を向上させた。Wang ら [23] は、アイテムに付与するタグの推薦において、ユーザのブックマーク履歴中のアイテムにエッジを設定することでアイテム間のグラフを作成し、アイテムに関する文書とこのグラフからアイテムの潜在因子ベクトルが生成されるようモデルを拡張することで正確性を向上させた。Chen ら [24] は、製品のレビューサイトにおける評価値の推定という実験において、レビューを行ったユーザが他のユーザから付与されるユーザの信頼度を表すスコアを取り入れるようモデルを拡張することで正確性を向上させた。

3.3 推薦モデル

本節では、iCTR および uCTR モデルの詳細を説明する。それぞれのモデルの模式図は、iCTR は図 1.4, uCTR は図 3.1 となる。以降はモデルの構造に関して、確率変数間の依存関係を視覚的に理解しやすく表記する方法であるグラフィカルモデルを参照しつつ、個々の変数が持つ役割について解説する。

3.3.1 iCTR

iCTR のグラフィカルモデルを図 3.2 に示す。図の上半分（赤枠で囲んだ部分）は 2.3.3.1 で説明した LDA の生成過程とおおよそ同じであり、単語比率に事前分布を設定していない点だけ異なる。図の下半分（青枠で囲んだ部分）は MF に相当する処理を表しており、MF を確率モデルとして定義した Probabilistic Matrix Factorization (PMF) [44, 45]

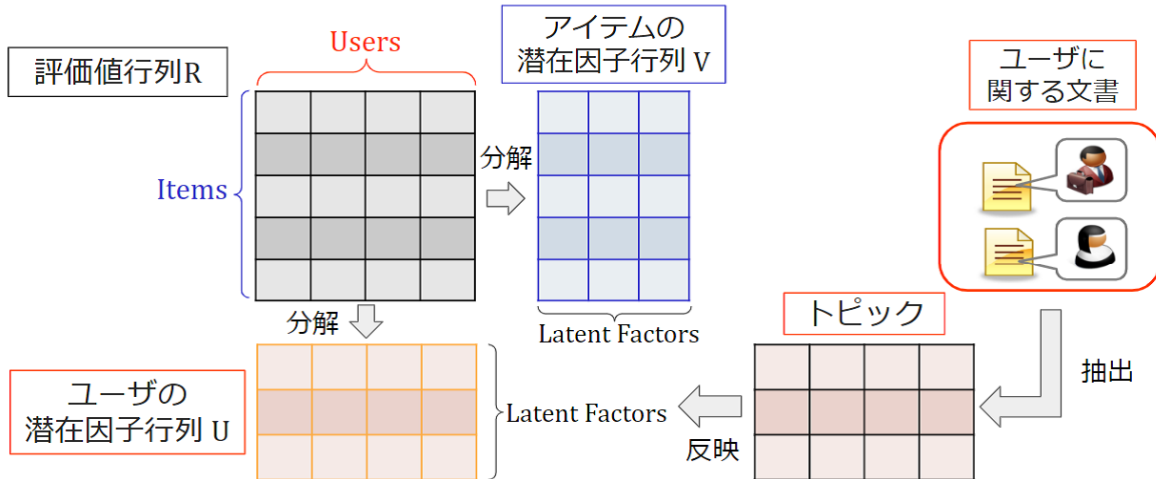


図 3.1 ユーザに関する文書を利用する Collaborative Topic Regression (uCTR) の模式図

がベースとなっている。図中の各変数については以降の節で詳しく説明するが、LDA 部分の θ と PMF 部分の ν との間に依存関係を定義したのが CTR モデルだと分かる。この依存関係により、アイテムに関する文書から抽出したアイテムを特徴づけるトピック比率 θ をアイテムの潜在因子ベクトル ν へ反映させている。

3.3.1.1 モデル説明

始めに、このモデルに登場する変数について説明を行う。添え字として、各ユーザを i (総数 I)、各アイテムを j (総数 J)、各トピックを k (総数 K) と表す。実際にデータとして得られる観測変数は、アイテム j に関する文書中の単語 (トークン) 列 w_j 、ユーザ i のアイテム j に対する評価値 r_{ij} である。 w_j の n 番目に現れたトークンを w_{jn} と表す。本章の評価実験において、評価値は $r_{ij} \in \{0, 1\}$ の二値である。 $r_{ij} = 0$ は、ユーザ i がアイテム j に興味を持っていない、もしくは存在を知らないことを意味する。 z_{jn} は潜在変数であり、トークン w_{jn} に対して割り当てられるトピックを表す。また、全アイテムの w_{jn} を集め、その中でユニークな単語の集合を語彙 W とする。

これら観測変数と潜在変数の他に、モデルを構成する要素としてパラメータが存在する。トピック比率 θ_j は K 次元の確率分布で、アイテム j がどのような分野に属するかという

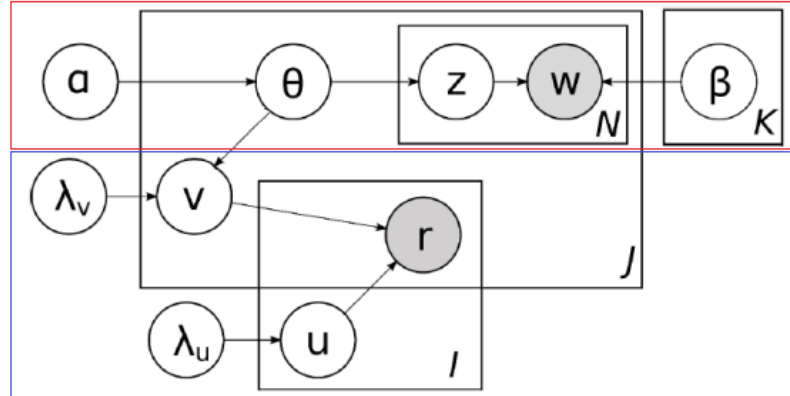


図 3.2 アイテムに関する文書を利用する Collaborative Topic Regression (iCTR) のグラフィカルモデル ([16] より引用および加工)

傾向を表している．単語比率 β_k は $|W|$ 次元の確率分布で，トピック k における語彙の出現確率を表している． u_i と v_j は，それぞれユーザ i とアイテム j の特徴を表現する K 次元の潜在因子ベクトルである． v_j はアイテム j がどの分野に属しているか， u_i はユーザ i がどのアイテムの分野に興味があるかを表現していると解釈できる．

iCTR にはパラメータを制御する変数として，ハイパーパラメータが存在する． K はトピック数であり潜在因子ベクトルの次元でもあり，モデルの複雑さを調整する働きがある． λ_u, λ_v は，それぞれ u_i, v_j を生成する正規分布の分散を制御するハイパーパラメータである． u_i は $N(0, \lambda_u^{-1} I_K)$ より生成され， λ_u は u_i に対して正則化（要素値が極端な値を取ることへのペナルティ）を行う効果がある．ただし， I_K は K 次元の単位行列， $N(\mu, \sigma)$ は正規分布を表す．一方， v_j は変数 ϵ_j を用いて $v_j = \epsilon_j + \theta_j$ ， $\epsilon_j \sim N(0, \lambda_v^{-1} I_K)$ より生成されるため， λ_v は v_j に対して正則化を行う効果がある．これは $\epsilon_j = v_j - \theta_j$ より， v_j と θ の乖離に対するペナルティとなる． α は， θ_j を生成するディリクレ分布のハイパーパラメータであり，アイテム全体における生成されるトピックの偏り具合を制御する．なお，本研究の実験では，Wang らの論文 [16] に従い，ハイパーパラメータ $\alpha, \lambda_u, \lambda_v$ はいずれもスカラーとする．最後に，評価値 r_{ij} は $N(u_i^T v_j, c_{ij}^{-1})$ より生成される．ただし， c_{ij} は既知の評価値 r_{ij} の影響を制御するハイパーパラメータであり，CTR では次の様に定義

する.

$$c_{ij} = \begin{cases} a (r_{ij} = 1 \text{ のとき}) & \dots\dots\dots (3.3.1) \\ b (r_{ij} = 0 \text{ のとき}) \end{cases}$$

$r_{ij} = 0$ は欠損値である要素を表し, $0 < b < a$ を満たす必要がある. CTR は以上のハイパーパラメータにより制御されるが, 特に K, a, λ_v はそれぞれモデルの複雑さ, 評価値の影響力, トピックの影響力を調整するため重要である. 最後に, CTR の生成過程を整理すると以下となる.

1. 各ユーザ i において, 正規分布 $N(0, \lambda_u^{-1} I_K)$ から \mathbf{u}_i を生成.
2. 各アイテム j において,
 - a. ディリクレ分布 $Dir(\alpha)$ から $\boldsymbol{\theta}_j$ を生成.
 - b. 正規分布 $N(0, \lambda_v^{-1} I_K)$ から K 次元ベクトル $\boldsymbol{\epsilon}_j$ を生成し, $\mathbf{v}_j = \boldsymbol{\epsilon}_j + \boldsymbol{\theta}_j$ とする.
 - c. 各トークン w_{jn} において,
 - i. 多項分布 $Mult(\boldsymbol{\theta}_j)$ から, トピック z_{jn} を生成.
 - ii. 多項分布 $Mult(\boldsymbol{\beta}_{z_{jn}})$ から, トークン w_{jn} を生成.
3. 各ユーザ・アイテムのペア (i, j) において, 正規分布 $N(\mathbf{u}_i^T \mathbf{v}_j, c_{ij}^{-1})$ から r_{ij} を生成.

3.3.1.2 モデルの学習

CTR モデルのパラメータ $(\mathbf{u}, \mathbf{v}, \boldsymbol{\theta}, \boldsymbol{\beta})$ の学習は, Wang らの論文 [16] に従い, 式 (3.3.2) に示すモデルの事後確率 L を最大化するよう反復的に最適化を行う. 各パラメータの更新は, 更新するパラメータ以外を固定して事後確率 L を最大化し, 更新したいパラメータを最適化する coordinate ascent 法 [62] を用いる.

$$L = - \sum_i \sum_j c_{ij} (r_{ij} - \mathbf{u}_i^T \mathbf{v}_j)^2 + \sum_j \sum_n \log \left(\sum_k \theta_{jk} \beta_{k, w_{jn}} \right)$$

$$-\lambda_u \sum_i \|\mathbf{u}_i\|^2 - \lambda_v \sum_j \|\mathbf{v}_j - \boldsymbol{\theta}_j\|^2 \dots \dots \dots (3.3.2)$$

\mathbf{u}, \mathbf{v} の更新は, それぞれ式 (3.3.3), 式 (3.3.4) により行う. 式中の U, V はそれぞれ列ベクトル $\mathbf{u}_i, \mathbf{v}_j$ を並べた行列, C^i, C^j は対角行列でそれぞれ $C_{dd}^i = c_{id}, C_{dd}^j = c_{dj}$, $\mathbf{r}_i, \mathbf{r}_j$ はベクトルでそれぞれ $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iJ}), \mathbf{r}_j = (r_{1j}, r_{2j}, \dots, r_{Ij})$ である. 各 \mathbf{u}_i に対して事後確率 L を大きくする勾配方向へ \mathbf{u}_i を微小量ずつ変化させていく. 各 \mathbf{v}_j に対しても同様に変化させていく.

$$\mathbf{u}_i = (VC^iV^T + \lambda_u I_K)^{-1} VC^i \mathbf{r}_i \dots \dots \dots (3.3.3)$$

$$\mathbf{v}_j = (UC^jU^T + \lambda_v I_K)^{-1} (UC^j \mathbf{r}_j + \lambda_v \boldsymbol{\theta}_j) \dots \dots \dots (3.3.4)$$

$\boldsymbol{\theta}$ の更新において, 式 (3.3.2) に示す事後確率 L の第 2 項目を直接求めることが非常に困難である. そこで, 式 (3.3.2) から $\boldsymbol{\theta}_j$ を含む項を取り出し, アイテム毎に分離して $L(\boldsymbol{\theta}_j)$ とする. 次に, $\phi_{jnk} = q(z_{jn} = k)$ を定義し, Jensen の不等式を利用して式 (3.3.5) の様に変形する.

$$\begin{aligned} L(\boldsymbol{\theta}_j) &= -\lambda_v \|\mathbf{v}_j - \boldsymbol{\theta}_j\|^2 + \sum_n \log\left(\sum_k \theta_{jk} \beta_{k, w_{jn}}\right) \\ &= -\lambda_v \|\mathbf{v}_j - \boldsymbol{\theta}_j\|^2 + \sum_n \log\left(\sum_k \phi_{jnk} \frac{\theta_{jk} \beta_{k, w_{jn}}}{\phi_{jnk}}\right) \\ &\geq -\lambda_v \|\mathbf{v}_j - \boldsymbol{\theta}_j\|^2 + \sum_n \sum_k \phi_{jnk} \{\log(\theta_{jk} \beta_{k, w_{jn}}) - \log(\phi_{jnk})\} \\ &= L(\boldsymbol{\theta}_j, \boldsymbol{\phi}_j) \dots \dots \dots (3.3.5) \end{aligned}$$

この $L(\boldsymbol{\theta}_j, \boldsymbol{\phi}_j)$ は, 本来求めたい $L(\boldsymbol{\theta}_j)$ の下限となっている. 最適な ϕ_{jnk} は, $L(\boldsymbol{\theta}_j, \boldsymbol{\phi}_j)$ より更新式が得られ, $\phi_{jnk} \propto \theta_{jk} \beta_{k, w_{jn}}$ となる. $\boldsymbol{\theta}_j$ の最適化に関しては, 解析的に行うことができないため, $\boldsymbol{\theta}_j$ が simplex であることを制約条件に projection gradient 法 [62] を利用して探索的に最適化を行う. 最後に, $\boldsymbol{\beta}$ の更新は式 (3.3.6) により行う.

$$\boldsymbol{\beta}_{kw} \propto \sum_j \sum_n \phi_{jnk} \delta(j, n, w) \dots \dots \dots (3.3.6)$$

$$\delta(i, n, w) = \begin{cases} 1 & (w_{jn} = w \text{ のとき}) \\ 0 & (w_{jn} \neq w \text{ のとき}) \end{cases}$$

以上のパラメータ (\mathbf{u} , \mathbf{v} , $\boldsymbol{\theta}$, $\boldsymbol{\beta}$) 更新処理を, 事後確率の変化が微小になり収束したと見なせるまで繰り返し実行する.

3.3.1.3 評価値の予測と推薦

学習が完了したモデルにおいて, 推定したパラメータ $\hat{\mathbf{u}}, \hat{\mathbf{v}}$ から式 (3.3.7) により予測評価値 \hat{r} を計算することができる. この予測評価値は実数値となるため, 値の大きい順にランキングを作成することができる. 推薦を行うには, 対象となるユーザ i に関連する評価値 r_{i*} の値が 0 の要素 (i, j) を抽出し, それらの予測評価値 \hat{r} を求め, 予測評価値 \hat{r}_{ij} の値が大きい順にアイテム j を列挙したものを, 推薦リストとしてユーザへ提示する.

$$r_{ij} \approx \hat{r}_{ij} = \mathbf{u}_i^T \mathbf{v}_j \dots \dots \dots (3.3.7)$$

3.3.2 uCTR

ユーザに関する文書を用いる場合のモデルである uCTR を定義する. uCTR のグラフィカルモデルを図 3.3 に示す. これは, 図 3.2 に示した iCTR のグラフィカルモデルにおいて, \mathbf{u} と \mathbf{v} および関連する λ_u と λ_v , I と J に対して, これらを入れ替えたモデルとなっている. uCTR では, トピック分布 $\boldsymbol{\theta}_i$ はユーザ i がどのようなユーザ層に属するかという傾向を表している. このトピックが反映される潜在因子ベクトル \mathbf{u} , \mathbf{v} も, 潜在因子の内容はユーザ層を表すものとなる. \mathbf{u}_i はユーザ i がどのユーザ層に属しているか, \mathbf{v}_j はアイテム j がどのユーザ層から好まれるかを表現していると解釈できる. uCTR のパラメータ学習に関しては, 3.3.1.2 節と同様に最大事後確率 (MAP) 推定により行う.

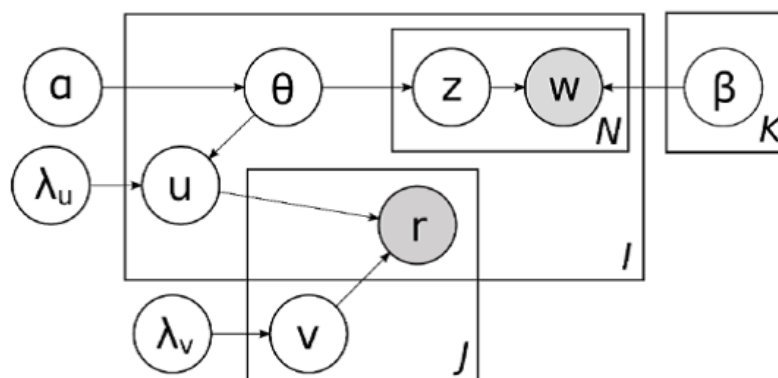


図 3.3 ユーザに関する文書を利用する Collaborative Topic Regression (uCTR) のグラフィカルモデル

3.4 実験設定

3.4.1 データセットとスパース度の設定

本実験では、Wang ら論文 [16] に倣い、ソーシャルブックマークサービス CiteULike^{*1} から取得したブックマーク情報と英文テキストデータを用いる。CiteULike において、2014 年 9 月から 11 月の間にブックマークを行ったユーザ、およびユーザが登録したすべての論文・科学記事を直接クロージングすることで収集した。上記に加えて、ユーザのプロフィール情報、論文・科学記事のタイトルと要約情報を収集した。収集したデータにおいて、評価値にはブックマーク情報（ある論文・記事をブックマークしたかどうか）を用いる。評価値の値は、ユーザがあるアイテム（論文・記事）に関心がある場合は 1、関心が無いもしくは存在を知らない場合には 0 となる。iCTR では、トピック抽出元となる文書として、アイテムの “article title” と “abstract” を用いる。uCTR では、トピック抽出元となる文書として、ユーザの “profile” と “interest”（ユーザが興味のある分野をキーワードで列挙したもの）を用いる。“profile” と “interest” の例（3 人分）を表 3.1 に載せ

^{*1} Web 上に存在する論文や Web ページのブックマークを行い、それらを整理・共有できるサービス

表 3.1 ユーザに関するテキスト情報の例

<p>“Profile” の例 (3 人分)</p> <ul style="list-style-type: none"> - PhD, ATR specialization in art therapy. - I am a member of the Software Engineering - Productivity tools department at *** - India. - I am generally interested in studies of learning and the design of instruction, especially in the areas of reading comprehension, science, and mathematics. In particular, I am interested in how objects external to the person (physical artifacts, symbolic and graphical representations) affect learning and reasoning.
<p>“Interest” の例 (3 人分)</p> <ul style="list-style-type: none"> - Cheminformatics, chemometrics, statistics, chemistry, metabolomics, systems biology, semantic web. - Complex networks, Spatial distribution of networks, Telecommunications, Graph theory, Network design - I work on Geno-transcriptomic study of solid tumours, diffusely infiltrating and thus with a complex architecture. My interest are therefore inter and intra tumour heterogeneity and tumour evolution. My biological background is based on Genomics and few Transcriptomics and Proteomics and Protein Networks. I have a technical biological background in : SNP and Expression Microarrays. FISH, PCR, Western Blot. And in Informatics : R, Python, SQL, PHP, Perl.

る。“profile”と“interest”の単語数の平均（単語数の標準偏差）は、それぞれ、39.2 (6.25) と 15.1 (3.88) であった。

評価値行列において、評価値存在率は次のように求める。また、これを用いてスパース度も次のように定義する。

表 3.2 データセットの統計量

	アイテム数	ユーザ数	ユニーク 単語数 (item)	ユニーク 単語数 (user)	総単語数 (item)	総単語数 (user)	スパース度 (%)	評価値 存在率 (%)
nlt=2	51,435	6,671	39,202	1,121	4,779,689	15,160	99.926	0.074
nlt=4	26,575	6,416	25,027	1,091	2,474,700	14,930	99.894	0.106
nlt=6	10,925	5,969	14,354	1,034	1,012,098	14,027	99.827	0.173
nlt=8	5,796	5,597	9,657	1,012	526,410	13,135	99.754	0.246
nlt=10	3,534	5,312	7,364	926	321,581	12,164	99.678	0.324

$$\text{評価値存在率} = \frac{\text{評価値の個数}}{\text{行数} \times \text{列数}} \dots\dots\dots (3.4.1)$$

$$\text{スパース度} = 1 - \text{評価値存在率} \dots\dots\dots (3.4.2)$$

取得したデータから作成した評価値行列のスパース度は 99.926 % であり非常に高い水準であった。スパース度による結果の変化を調べるため、被ブックマーク数が一定の閾値未満であるマイナーなアイテムとその評価値を除外することで、スパース度を低下させる処理を行った。被ブックマーク数の閾値は 2, 4, 6, 8, 10 に設定し、この閾値未満のアイテムを除外した評価値行列を順に $nlt = 2, 4, 6, 8, 10$ と表記する。これにより、スパース度の増減によるモデル毎の性能の変化についても分析する。最後に、データセットの主な統計量（スパース度と評価値存在率を含む）を表 3.2 に示す。

3.4.2 前処理

CiteULike から収集した論文・科学記事の大半は英語で書かれている。表記ゆれへの対処やノイズとなる単語を除去するため、英語のテキストに対して一般に用いられる以下の前処理を行う。

- 小文字に統一

- ステミング（単語の語形変化を取り除き，単語表現を揃える処理）
- ストップワード（he, the, this など）の除去

本実験では，ステミングには TreeTagger [80] を利用し，ストップワードの除去には SlothLib ^{*2} のストップワード一覧を利用した。

3.4.3 ハイパーパラメータの設定

3.3 節で説明した各ハイパーパラメータに設定する値は，グリッドサーチを用いて決定した。潜在因子ベクトルの次元に相当するトピック数 K は，特徴の表現力のある程度維持しつつ，特徴量の次元数が可能な限り小さい値となるように，段階的に設定して評価指標の値を確認した。具体的には，トピック数 $K = 5, 10, 30, 50, 100, 200$ において，*Recall* の向上が飽和した $K = 50$ を採用した。潜在因子ベクトルの正則化を調整する λ_v, λ_u は，Wang らの論文 [16] を参考に，iCTR では $\lambda_v = 1, 10, 100, \lambda_u = 0.1, 1, 10$ ，uCTR では λ_v と λ_u の値を入れ替えた設定で実行し，それぞれの組み合わせを調べた際の最も良い結果（iCTR では $(\lambda_v = 100, \lambda_u = 0.1)$ ，uCTR では $(\lambda_v = 0.1, \lambda_u = 100)$ ）を採用した。残りのハイパーパラメータは，Wang ら論文 [16] と同じ値を採用した ($a = 1, b = 0.01, \alpha = 1$)。パラメータ θ と β の初期値は，事前にデータセットに用いる文書集合を使って LDA モデルの学習を行い，その結果得られる θ と β の値を採用した。なお，LDA の学習は Collapsed Gibbs Sampling [61] で 500 回の反復を行った。その他のパラメータは，乱数により初期化を行う。また，MF に関しては，Wang らの論文 [16] と同様に，CTR の θ の値を全て 0 に固定することで実装を流用した。MF の他のハイパーパラメータは， $\lambda_v = 0.1, \lambda_u = 0.1, a = 1, b = 0.01, \alpha = 1$ を採用した。

^{*2} <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word>

3.4.4 評価方法

推薦システムの性能を調べるため、推薦の正確性を表す指標として再現率、推薦の利便性を表す指標として被覆率と多様性の評価を行う。不特定多数のユーザによる評価データを利用するため、交差検定によるオフライン評価 [11] を行う。交差検定は、まず評価値集合を 5 分割して訓練データ 4 つとテストデータ 1 つに分ける。テストデータは、擬似的にユーザが好む未知のアイテム（正解データ）とみなす。訓練データを利用してモデルの学習を行った後、上位 N 個の推薦アイテムを推薦リストとして獲得する。そして、推薦リストと正解データから評価指標の値を算出する。

正確性の評価指標には、Wang らの論文 [16] や類似の研究 [20, 21, 22, 23, 24, 25, 45] に倣い、再現率 (*Recall*) を採用する。適合率 (*Precision*) も評価指標の候補として考えられるが、オフライン評価ではユーザがまだ閲覧していないが閲覧すれば好むと予測されるアイテムは「好きでない」と扱われるため、アイテム数が非常に多い場合は再現率が適している。利便性の評価指標には、推薦システムが偏りなく様々なアイテムを実際に推薦できているかを表す被覆率 (*Coverage*) と、ユーザ毎に個別化された推薦ができているかを表す多様性 (*Diversity*) を採用する。本実験では、予測評価値 \hat{r}_{ij} を対象ユーザ i について降順に並べた際の上位 50 個のアイテムを推薦リスト L_i とする。

再現率 (*Recall*) は次式で求められる。

$$Recall = \frac{|T_i \cap \mathfrak{J}L_i|}{|T_i|} \dots\dots\dots (3.4.3)$$

ただし、 T_i はテストセット中のユーザ i が好むアイテム集合、 \mathfrak{J} はリストからリスト内のアイテム集合への写像を表す。最終的なシステムの再現率は、全ユーザの *Recall* の平均を取ることで求める。

被覆率は、全アイテムの中で実際に推薦可能なアイテムがどの程度存在するかを表す。本研究では、被覆率の指標の一つである Catalogue Coverage (*CC*) [63] を採用する。*CC*

は次式で求められる。

$$CC = \frac{|\cup_{i=1, \dots, I} \mathfrak{L}_i|}{|B|} \dots \dots \dots (3.4.4)$$

ただし、 B はテストセット中のアイテム集合である。 CC が低いほど、推薦システムは全アイテム中の一部のアイテムしか推薦できないことを示す。

多様性を表す指標は、凝集多様性 (Aggregate Diversity) [64], ユーザ間相違度 (Inter-User Diversity) [65], 時間的多様性 (Temporal Diversity) [66] などいくつかの指標が存在する。凝集多様性は、計算方法が簡単なものの極端な偏りを持つ場合 (例えば、半数のユーザには重複なくアイテムを推薦しているが、残りの半数のユーザには同じアイテムばかり推薦する場合) でも高い多様性を示してしまう。時間的多様性は、ユーザの利用時間の経過ごとに異なるアイテムを推薦できているかを計測する指標であるが、今回のデータセットは静的なものである (ユーザがブックマークした時間は記録されていない) ため、用いることができない。そこで、個々のユーザごとに多様な推薦結果が得られているかを直接的に示すユーザ間相違度 (IUD) を用いる。 IUD は、任意の2人のユーザに注目した時に、互いの推薦リストに含まれるアイテムがどれほど違っているかを表す。 IUD は次式で求められる。

$$IUD = \frac{1}{|U|C_2} \sum_{u_1}^U \sum_{u_2}^U d_{u_1, u_2} \dots \dots \dots (3.4.5)$$

$$d_{u_1, u_2} = 1 - \frac{|\mathfrak{L}_{u_1} \cap \mathfrak{L}_{u_2}|}{S}$$

ただし、 $S = |\mathfrak{L}_{u_1}| = |\mathfrak{L}_{u_2}|$ である。 IUD は $[0, 1]$ に正規化されたユーザ間の距離の平均と捉えることができ、 0 に近いほどユーザ間の推薦結果に差は無く、 1 に近いほど差が存在することを意味している。

3.5 結果と考察

3.5.1 Perplexity

実験結果における各評価指標を見る前に、まず構築したモデルにおける *Perplexity* を確認する。*Perplexity* は言語モデルの複雑さを表す指標であり、一般的には小さい方が良いとされる。ただし、本実験においては、利用するテキストの違いにより、学習されるモデルの複雑さに違いが生まれるか否かを確認するために用いる。iCTR を対象とした場合の *Perplexity* は、式 (3.5.1) で表される。

$$perplexity(D) = exp \left\{ - \frac{\sum_j \sum_n \log(\sum_k \theta_{jk} \beta_{k,w_{jn}})}{\sum_j \sum_n 1} \right\} \dots\dots\dots (3.5.1)$$

この式で用いられる変数は、3.3.1.1 節で定義したものと同一である。uCTR を対象とする場合は、アイテム j をユーザ i に置き換えれば良い。

今回のデータセットにおける、収束時の各 nlt での *Perplexity* を図 3.4 に示す。この図を見ると、今回のデータセットでは iCTR よりも uCTR はモデルの複雑さが小さいことが分かる。すなわち、uCTR の方が確率的に選択される単語はより限定的であり、生成されるテキストの曖昧性が小さいことを意味している。また、 nlt の値が増加しても *Perplexity* の値はほぼ一定であることから、スパース度による影響も uCTR の方が小さいことが分かる。よって、iCTR と比べて uCTR の方が学習によるモデルの収束が容易で、安定性が高いといえる。ただし、uCTR のモデルの複雑さは非常に小さいということは、ユーザに関する文書が持つ情報はアイテムに関する文書が持つ情報よりも乏しいのではないかという懸念が生じる。表 3.1 に示すように、ユーザの “Profile” と “Interest” はアイテムの “abstract” に比べると文字数が少ない傾向で、これがモデルの複雑さを決める要因になっている可能性が考えられる。そのため、*Perplexity* が低いから推薦性能が良くなるとはいえない。

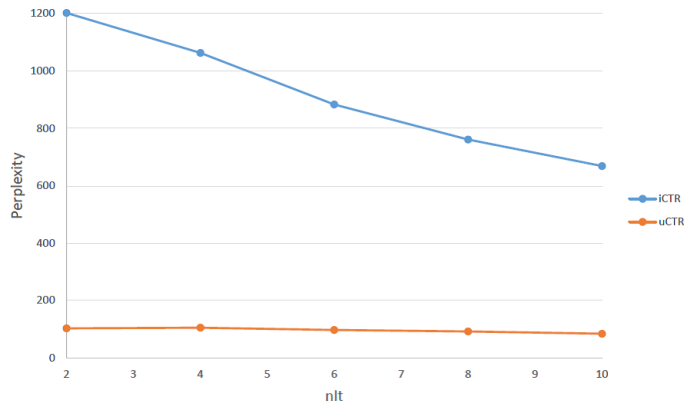


図 3.4 Perplexity ($nlt = 2$ が最もスパース)

3.5.2 再現率

正確性の指標である再現率 (*Recall*) の結果を図 3.5 に示す。このグラフは iCTR, uCTR, MF のモデル間の比較、および 5 段階に分けたスパース度での比較を行っている。横軸は推薦リストに含めるアイテムの数 (推薦数) である。まず同一モデルにおけるスパース度の変化による影響に注目すると、どのモデルにおいてもスパース度が高いほど (nlt の値が小さいほど)、*Recall* の値は小さくなる傾向が見られる。MF はスパース度が高すぎる場合 ($nlt = 2$)、推薦システムとして正常に機能していないことが分かる。興味深いことに、両 CTR と MF ではスパース度の変化に伴う *Recall* の変化量が大きく異っている。推薦数が 100 となる値 (グラフ右端) に注目すると、MF ではスパース度が最大の時 ($nlt = 2$) と最小の時 ($nlt = 10$) の間で *Recall* の差は約 0.55 となっている。対して、iCTR はスパース度が最大の時と最小の時の間で *Recall* は約 0.2 しか変わらず、uCTR は約 0.35 しか変わらない。両 CTR とも、MF に比べるとスパース度による影響が小さく抑えられている。このことから、iCTR と uCTR は、MF に比べてデータの欠損に対してロバストであることが分かる。一方で、データのスパース度が小さい場合は、MF が最も性能が良くなっている。これは、CTR でテキストから抽出されたトピックの情報よりも、評価値行列に存在する潜在的な情報の方がよりリッチであり、トピックの情報がノイズ

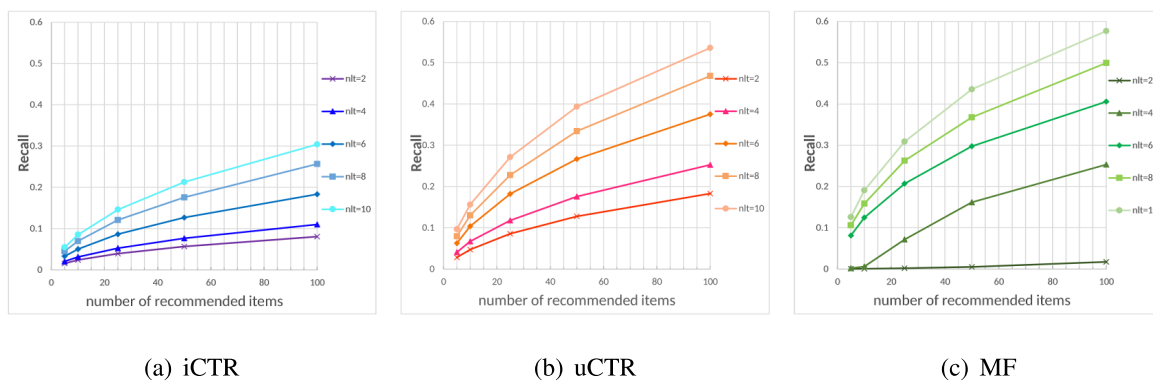


図 3.5 Recall, モデル・スパース度毎の比較 (nlt=2 が最もスパース)

となってしまっているためと考えられる。また、iCTR と uCTR との間にも明らかな差が存在している。特に、iCTR はスパース度が低い場合には、uCTR よりも顕著に *Recall* が低くなっている。この原因を解明することは非常に困難であるが、アイテムのトピックに比べてユーザのトピックを利用する方が推薦の正確性は向上する可能性があることを示す結果である。

以上の結論として、iCTR、uCTR とともに評価値の情報不足をトピックにより補うことができている、またスパース度による影響を抑えることができるといえる。特に uCTR においては、スパース度が高い場合により推薦の正確性を向上させる可能性があることを確認できた。ただし、評価値の情報が十分である場合は逆にノイズとなる可能性があることに注意する必要がある。

3.5.3 被覆率

利便性の 1 つ目の指標として、被覆率を表す Catalogue Coverage (*CC*) の結果を図 3.6 に示す。同一モデルにおけるスパース度の変化による影響は、*Recall* の傾向と似たものとなった。すなわち、スパース度が高いほど (*nlt* の値が小さいほど)、*CC* の値は小さくなる傾向が見られた。スパース度が高い場合と低い場合の *CC* の値の差は、推薦数が増えるにつれて大きくなった。手法間で比較を行うと、iCTR がいずれのスパース度におい

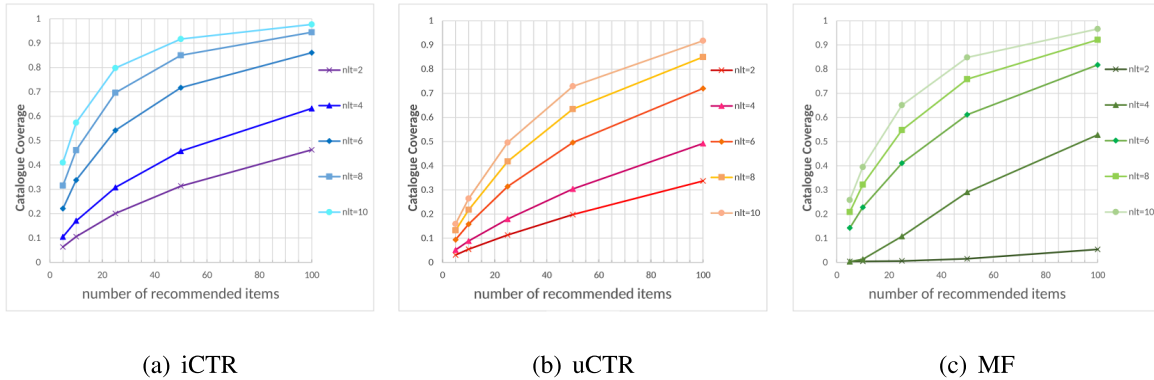


図 3.6 Catalogue Coverage, モデル・スパース度毎の比較 (nlt=2 が最もスパース)

でも、またいずれの推薦数においても、他の 2 手法より優れた結果となっている。uCTR は、スパース度が高い場合 ($nlt = 2, 4$) に MF より優れている。トピックの導入により、スパース度が高い場合に CC が極端に低下してしまうのを抑制することが分かる。推薦を行うための元となるユーザやアイテムに関する情報（評価値の潜在的な情報も含む）が多いほど、モデルのパラメータはより多様なユーザとアイテムを表現できるように学習され、結果として多様なパターンの推薦を行えると考えられる。

3.5.4 多様性

利便性の 2 つ目の指標として、多様性を表す Inter-User Diversity (IUD) の結果を図 3.7 に示す。iCTR はいずれのスパース度においても、またいずれの推薦数においても、 IUD は非常に高い値で安定している。iCTR は 3 つのモデルの中で最も良い結果であり、ユーザごとに個別化された多様な推薦が実現できていると言える。uCTR は、iCTR やスパース度が低い場合の MF より悪い結果となった。スパース度の変化に注目すると、それほど大きな差ではないがスパース度が高いほど IUD も高くなっている。推薦数の変化に注目すると、uCTR では推薦数が 25 を超えてくると 0.8 前後の値を示しており、ある程度高い個別化が行われている。しかし、推薦数が 5 から 25 の間では、推薦数が少ないほど IUD が低くなっている。これは推薦数が少ない時には多くのユーザがブックマーク

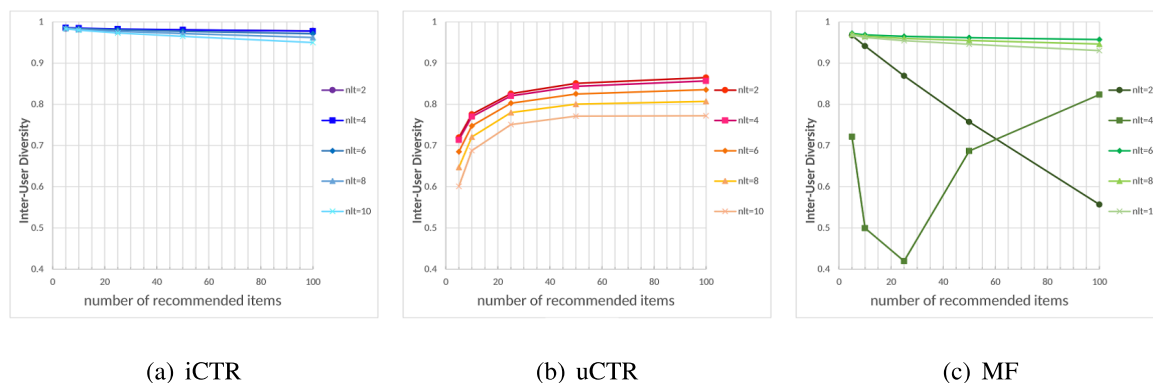


図 3.7 Inter-User Diversity, モデル・スパース度毎の比較 ($nlt=2$ が最もスパース)

しているメジャーなアイテムを推薦し、推薦数が増えると個別化されたアイテムを推薦しているからだと考えられる。すなわち、推薦リストの上位ではどのユーザにも同じアイテムが推薦される傾向があり、推薦システムとしては好ましくない。MF は、スパース度が低い場合、iCTR に匹敵するほどの良い結果となった。しかし、スパース度が高い場合 ($nlt = 2, 4$) に他のモデルとは異なる挙動を示していることが見て取れる。MF ($nlt = 2$) は、急激に単調減少している様子が見られるが、評価値不足のため推薦機構が有効に働かなくなっているためだと考えられる。MF ($nlt = 4$) は、複雑な挙動を示しており、推薦数 25 以降では *IUD* が上昇する様子が見られる。この原因に関しては特定することができなかったが、スパースが高い場合は MF の挙動が不安定になることに注意する必要がある。以上より、MF ではスパース度が高くなると *IUD* の値が急激に減少したり不安定になる傾向があるが、トピックを導入することでその現象を抑制できることが分かる。

3.5.5 推薦例

本節では実際の推薦例を確認することで、各モデルの特徴について考察する。推薦例として、ユーザに関する文書である“Profile”と“Interest”の両方で文字数が十分にある者（ユーザ A）と、“Profile”と“Interest”の両方で文字数が非常に少ない者（ユーザ B）の両方を取り上げる。推薦結果は、スパース度が最も高い場合 ($nlt = 2$) の上位 3 位まで

表 3.3 推薦例 (ユーザ A)

“Profile”	I am a *** Master Student for Natural Language Processing at the University of *** - UK. My main interest is at Opinion Mining and Sentiment Analysis.
“Interest”	Opinion Mining, Sentiment Analysis, Machine Learning, Question Answering, Automatic Summarization, Discourse
評価値の数	22
推薦結果 (iCTR)	<ol style="list-style-type: none"> 1. What is Twitter, a social network or a news media?, Proc. WWW (2010) 2. Why We Twitter: Understanding Microblogging Usage and Communities, Proc. Workshop on Web Mining and Social Network Analysis (2007) 3. Users of the world, unite! The challenges and opportunities of Social Media, Business Horizons, Vol. 53, No.1. (2010)
推薦結果 (uCTR)	<ol style="list-style-type: none"> 1. Latent Dirichlet Allocation, The Journal of Machine Learning Research, Vol.3, No.4-5. (2003) 2. A Training Algorithm for Optimal Margin Classifiers, Proc. Workshop on Computational Learning Theory (1992) 3. The arbitrariness of the genetic code, Biology and Philosophy, Vol. 19, No. 2. (2004)
推薦結果 (MF)	<ol style="list-style-type: none"> 1. What is Twitter, a social network or a news media?, Proc. WWW (2010) 2. Why We Twitter: Understanding Microblogging Usage and Communities, Proc. Workshop on Web Mining and Social Network Analysis (2007) 3. The arbitrariness of the genetic code, Biology and Philosophy, Vol. 19, No. 2. (2004)

のアイテムを示す。ユーザ A の情報 (“Profile” と “Interest”, 評価値の数) と推薦結果を表 3.3 に、ユーザ B の情報と推薦結果を表 3.4 に示す。推薦されたアイテムのタイトルを見ると、いずれの推薦結果においても iCTR は詳細な内容に関する研究論文が多く提示

表 3.4 推薦例 (ユーザ B)

“Profile”	
“Interest”	Cell Biology, Genetics
評価値の数	5
推薦結果 (iCTR)	<ol style="list-style-type: none"> 1. Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genetics, Vol. 4, No. 6 (2008) 2. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1, National Academy of Sciences of the United States of America, Vol. 111, No. 18 (2014) 3. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst, Development (Cambridge, England), Vol. 132, No. 9 (2005)
推薦結果 (uCTR)	<ol style="list-style-type: none"> 1. How To Choose a Good Scientific Problem, Molecular Cell, Vol. 35, No. 6 (2009) 2. The arbitrariness of the genetic code, Biology and Philosophy, Vol. 19, No. 2 (2004) 3. Detecting novel associations in large data sets, Science, Vol. 334, No. 6062 (2011)
推薦結果 (MF)	<ol style="list-style-type: none"> 1. Multiplex genome engineering using CRISPR/Cas systems, Science, Vol. 339, No. 6121. (2013) 2. Sequencing depth and coverage: key considerations in genomic analyses, Nature reviews. Genetics, Vol. 15, No. 2. (2014) 3. A whole-cell computational model predicts phenotype from genotype, Cell, Vol. 150, No. 2. (2012)

されている。一方、uCTR は解説記事や総説論文などの一般論に関する論文が多く提示されている。MF は、評価値の数が多いユーザ A に対しては、iCTR と同じような研究論文が多く示され、評価値の数が少ないユーザ B に対しては、uCTR ほどではないがやや一般論に関する論文が示されているように思われる。このことから、uCTR はユーザの興味

をより一般化する傾向があり、その分野で重要な解説記事や総説論文が推薦される傾向にあることが分かる。これが、uCTR の CC と IUD が iCTR や MF よりも低かった要因だと考えられる。

3.6 結論

本章では、協調フィルタリングとトピックモデルを組み合わせた代表的モデルである Collaborative Topic Regression (CTR) モデルに注目し、アイテムに関する文書を用いた場合の CTR (iCTR と呼ぶ) と、ユーザに関する文書を用いた場合の CTR (uCTR と呼ぶ)、およびベースラインである Matrix Factorization (MF) の推薦性能の違いを実験により比較した。実験では、評価値の欠損にどれだけ対応できるかという点にも注目し、評価値行列のスパース度を変化させた際の iCTR, uCTR と MF の性能を調べた。それぞれの手法を総合的な観点から評価を行うために、正確性に関する評価指標として再現率を、利便性に関する評価指標として被覆率と多様性を用いた。実験結果から得られた結論、およびモデル採用に関しての指針は 3.6.1 節で説明する。最後に、次章での課題を 3.6.2 節で述べる。

3.6.1 結果の要約

実験結果の要約を表 3.5 に示す。この結果を見ると、iCTR と uCTR のどちらが優れているかについては一概に決めることは難しい。iCTR、すなわちアイテムに関する文書を用いた場合では、利便性の高い推薦が行える傾向がある。uCTR、すなわちユーザに関する文書を用いた場合では、正確性の高い推薦が行える傾向がある。MF はスパース度が低い場合は最も正確性の高い推薦が行えたが、スパース度に対して極めて敏感であるため、利用したいデータセットのスパース度が高いことが見込まれる場合には、iCTR や uCTR の導入を検討したほうが良いことが分かった。また、iCTR と uCTR 共に、MF に比べて評価値の欠損によるスパース性の影響が小さくロバストであることが示せた。

表 3.5 実験結果の要約

		iCTR	uCTR	MF
スパース度 (高)	Recall	○	◎	×
	CC	◎	○	×
	IUD	◎	△	△
スパース度 (低)	Recall	△	◎	◎
	CC	◎	○	○
	IUD	◎	△	◎

iCTR, uCTR, MF の選択基準を整理したガイドラインは以下のようになる。推薦の正確性を重視する方針で、データセットのスパース度が低いことが分かっているならば、文書から抽出したトピックを利用することなく MF を採用すれば良い。データセットのスパース度が高い場合は、uCTR を採用する方が良い。MF は、スパース度によって大きくその正確性が変わってしまうため、システム運用後のスパース度が想定できない場合は、uCTR の方が無難といえる。システム全体として多くの種類のアイテムが推薦対象となることを重視する方針、またはユーザ毎に個別化された推薦ができることを重視する方針ならば、iCTR を採用すれば良い。特に、ユーザへ提示する推薦数が少ない場合、スパース度が想定できない場合に効果的である。原則的に、スパース度が高いことが分かっている場合は MF の採用は避けるべきである。最後に、以上を箇条書きでまとめる。

- 推薦の正確性を重視する場合
 - 評価値行列のスパース度が低い場合
 - * MF を採用する
 - 評価値行列のスパース度が高い、またはスパース度が想定できない場合
 - * uCTR を採用する
- システムが推薦するアイテムの多様性を重視する、またはユーザ毎に個別化された推薦を重視する場合

– iCTR を採用する

3.6.2 今後の課題

本章の実験における結果, 特に *Recall* と *CC* を見ると, スパース度が高い場合にトピックが推薦性能へ与える影響は顕著であることが明確になった. トピックは文書から抽出されるため, 文書中の単語数や使用されている語彙など文書の特徴に着目することで, CTR におけるスパース度の影響をより詳細に分析できるのではないかと考える. スパース度が低い場合に iCTR で *Recall* が低下する現象が見られたが, トピックがノイズとなっていると仮定するなら, 抽出元となる文書とスパース度との関係を分析することは性能向上に繋がる可能性がある. 加えて, トピックが潜在因子ベクトルへ与える影響の大きさを調整するハイパーパラメータ λ と, 評価値が潜在因子ベクトルへ与える影響の大きさを調整するハイパーパラメータ a は, 両者の交互作用だけでなくスパース度との組み合わせも含めて分析する必要があると考える. また, 本章の実験では使用したデータセットのドメインが CiteULike だけであるため, 実験結果から導いた結論の一般性には疑問が残る. CiteULike 以外のドメインからもデータを収集して実験を行い, 発見した傾向の一般性を確かめる必要がある.

第 4 章

スパース度と利用する単語の種類が 推薦性能に与える影響の分析

本章では、1.2.2 節で述べた課題 2 および 3 に対する取り組みとして、推薦性能に影響を与える要因について仮説を立て、複数のデータセットを用いた実験により、それらの要因間の関係性について分析する。

本章の構成は以下となる。研究の背景と動機となる考えや仮定および本章の目的を 4.1 節で述べる。関連研究として、サイド情報を利用する MF における課題を 4.2 節で示す。トピック内容を操作するために行う単語選択の手法についての説明を 4.3 節で行う。実験で使用するデータセットの説明やハイパーパラメータの設定、評価指標など実験設定を 4.4 節で説明する。実験結果の提示と考察を 4.5 節で行う。実験結果から導かれる結論を 4.6 節で述べる。

4.1 背景と目的

CTR はアイテムに関する文書からトピックを抽出し、それを評価値行列から推定される潜在因子ベクトルに反映させることで、評価値行列に欠損値が多い場合においても高い

推薦精度を維持することに成功している。CTR には、評価値行列からの影響とトピックの影響をそれぞれどれくらいの強さで反映するか調整するハイパーパラメータが存在する。CTR のハイパーパラメータは他にも存在するものの、この2つは推薦結果に大きく影響を与えると予想される。また、CTR の動作は与えられた評価値行列のスパース度にも大きく依存している。具体的には、評価値が極端に不足していてスパース度が非常に高い状況では CTR は非常に有効であるが、評価値がある程度存在するスパース度が比較的低い状況になると MF よりも性能が劣ることが3章で確認された。CTR は評価値行列から抽出された潜在因子に文書から抽出されたトピックを反映させて補うモデルである。そのため、スパース度が高い評価値行列から抽出された潜在因子だけで十分な学習ができない場合はトピックにより性能が向上し、逆にスパース度が低い評価値行列から潜在因子が抽出された場合はトピックからの情報はノイズになり性能が低下すると筆者は考えた。

評価値行列のスパース度が高い状況では、評価値行列からアイテム間またはユーザ間の局所的な関係は抽出できるが、アイテムのカテゴリのような大域的な関係は抽出が困難だと考えられる。このような場合は、文書から抽出されたトピックにより大域的な関係を表す情報を補完できればよく、文書中でそのような大域的な情報を持つのは様々な文書に出現する単語（文書頻度の高い単語）だと筆者は考えた。逆にスパース度が低い状況では、評価値行列から大域的な関係は十分に抽出可能であり、大域的な関係を表すトピックはノイズとなってしまう推薦性能に悪影響を及ぼす可能性がある。この場合は、文書頻度の高い単語は選択しない方がよい。よって、CTR を効率的に利用するためには、スパース度に応じて有益なトピックを抽出する必要があると考え、文書頻度に基づいて利用する単語を選択することが重要だと仮定する。3.6.2 節でも言及したように、CTR の性能は以下の要素に依存していると筆者は考えている。

- 評価値行列のスパース度
- トピック抽出で利用する単語の文書頻度
- 評価値の影響力や文書から抽出したトピックの反映度合いを調整するハイパーパ

ラメータ

これらの要素は互いに密接に関連し合っている可能性が高いため、その関係性を明らかにし、CTR を十分に活用できるような知見を整理することは重要である。そこで本研究では、異なる評価値行列のスパース度の下で、文書頻度を基に利用する単語を選択し、また潜在因子ベクトルに影響するハイパーパラメータの組み合わせを変えることで、CTR の性能がどのように変化するかを複数のデータセットで比較し、共通する推薦結果の傾向（良い推薦結果を得るための単語選択方法とパラメータ設定方法）と、文書情報の利用による性能向上が期待できない条件をまとめる。

実験において、ドメインの異なる3種類のデータセットを用意し、結果の一般性を検証する。なお、利用する文書がアイテムの情報を含む文書であるかユーザの情報を含む文書であるかもトピックの内容に影響するが、本研究ではトピック抽出に利用する文書がアイテムに関する文書の場合に注目して精緻に分析を行う。評価指標には、推薦結果がユーザの嗜好に適しているかを評価する正確性、および全体として特定の人気アイテムだけが推薦されるような偏った推薦結果になっていないかを評価する利便性を利用し、この2つの観点で評価を行う。この2つの評価指標はトレードオフの関係にあるため、バランスよく高い値となることが望ましい。最後に本研究の貢献は以下となる。

- いくつかのスパース度の下で、トピック抽出に利用する単語の文書頻度およびハイパーパラメータの組み合わせが性能に与える影響を分析
- ドメインの異なる複数のデータセットによる実験から上記の一般性を検証し、また文書情報による性能向上が期待できない条件を考察

4.2 関連研究

4.2.1 サイド情報を利用する MF における課題

サイド情報を MF に組み込むことを目的として、評価値行列とサイド情報の行列を同時に分解するモデルが多く提案されている [19, 26, 34, 35, 36, 37]. Singh ら [34, 35] は互いに関係を持つ複数の行列に対する汎用的な MF である Collective matrix factorization (CMF) を提案した. CMF は与えられた評価値行列およびアイテムとサイド情報の関係を表す行列を、潜在因子の次元を共有しながら潜在因子行列に分解するモデルである. コールドスタート問題の観点で見ると、評価値が全く存在しない新規アイテムに対しても、サイド情報の行列から抽出された潜在因子を元に評価値を推定することができる. 一方で、CMF は評価値が十分に存在する状況 (ウォームスタート) において MF より性能が低下する可能性があると言われている [38]. ウォームスタート時でも利用するサイド情報の種類とハイパーパラメータの設定次第でこの性能低下は抑えられるものの、CMF は MF に比べて調整の難しいモデルだと指摘されている [38].

サイド情報を利用する CMF ではコールドスタート問題に対する性能向上と引き換えにウォームスタート時に性能低下する問題が見られたが、3章の結論 3.6.2 より、文書をサイド情報として利用する CTR においても同様の問題が生じている. そのため、CTR においてもスパース度を考慮してサイド情報の影響を調整するハイパーパラメータを設定する必要がある. 加えて、CTR はサイド情報を文書として与えるため利用する単語を選択することが可能であり、このようなサイド情報の選択によってもウォームスタート時の性能低下に対処できる可能性がある. すなわち、スパース度のレベルに応じて、利用する単語の種類およびハイパーパラメータの組み合わせが性能に与える影響を分析したことが、本章の研究がこれまでの研究と異なる点である.

4.3 単語選択手法

本節では、トピック内容に大域的な関係を表す情報を含めるか除外するか調整を行う単語選択の手法について説明する。CTR は与えられた文書中の単語を元にトピックを抽出するため、どのような単語を利用するかが重要である。3.3.1.1 で説明したように、トピック k の内容を表す単語比率 β_k は $|W|$ 次元の確率分布であり、語彙 W は文書集合 D に含まれる全ての文書のユニークな単語の集合である。すなわち、語彙 W を構成する単語は単語比率 β_k の表現力に直結しており、語彙 W から単語を取捨選択することでトピックの表現力を調整することができる。本研究では、単語選択により作成されたユニークな単語集合 (語彙) を語彙サブセットと呼ぶことにする。

4.1 節で説明したように、アイテム間またはユーザ間の大域的な関係を表す情報を補完するには、様々な文書に出現する単語が重要であると考えた。様々な文書に出現する単語であるかの定量化には、式 (4.3.1) に示す文書頻度 DF を採用する。大域的な関係を表す情報を含めたい場合は、文書頻度 DF が高い単語から順に選択した語彙サブセットを用いればよい。

$$DF = \frac{\text{その単語が出現した文書数}}{\text{全文書数}} \dots\dots\dots (4.3.1)$$

語彙 W を構成する単語を DF の高い順に並べ、パーセンタイルを基準に、上位 25%、上位 25% から上位 50%、上位 50% から上位 75%、上位 75% から上位 100% のように 4 つに分割した語彙サブセットを作成する。本論文では、例えば上位 25% から上位 50% の語彙サブセットを示すときには、DF[25, 50] のように表記する。また、比較対象としてランダムに指定個数の単語を選んだ語彙サブセットを作成する。例えば DF で分割後の語彙サブセットが 4 つであれば、語彙 W から $|W|/4$ 個の単語をランダムに選択した語彙サブセットを 1 つ作成する。

表 4.1 スパース度の設定 (評価値の削減方法)

original	削減なし (オリジナルの評価値行列)
sample50%	ランダムに 50% 削減
max2u	各ユーザの評価値数が 2 になるまで削減

表 4.2 各データセットにおける評価値行列の統計量

評価値の削減方法	評価値存在率 (%)	ユーザ数	アイテム数
CiteULike			
original	6.0e-4	6,877	52,550
sample50%	3.0e-4	6,664	46,927
max2u	3.6e-5	6,877	10,594
pixiv			
original	2.5e-4	90,565	73,979
sample50%	1.3e-4	90,565	66,040
max2u	6.2e-5	90,565	32,172
Goodreads			
original	3.02e-4	50,000	59,402
sample50%	1.51e-4	50,000	54,200
max2u	9.25e-5	50,000	21,610

4.4 実験設定

4.4.1 データセットとスパース度の設定

本実験では、以下に述べる分野の異なる 3 種類のドメインから取得したデータを用いる。

- CiteULike：ソーシャルブックマークサービス CiteULike における論文・科学記事を推薦する。本研究では、2014 年 9 月から 11 月の間にブックマークを行ったユー

ザとそのユーザが登録したすべての論文・科学記事（アイテム）のタイトルと要約情報を直接クロールすることで収集した。評価値はユーザがあるアイテムにブックマークしたかどうかの2値情報であり、ユーザがあるアイテムに関心がある場合は1、関心がないもしくは存在を知らない場合は0となる。トピックの抽出元となる文書はアイテムのタイトルと要約情報である。

- **pixiv**：イラストのコミュニケーションサービス pixiv におけるイラストを推薦する。本研究では、2019年8月時点でイラスト投稿を行ったことがあるユーザ（シードユーザ）をランダムに選択し、シードユーザが投稿したイラスト（アイテム）に関するメタ情報を直接クロールすることで収集した。それらのアイテムへブックマークを行ったユーザが推薦対象であり、評価値は各アイテムへブックマークしたかどうかの2値情報である。トピックの抽出元となる文書はアイテムのタイトルと紹介文（キャプション）である。
- **Goodreads**：書籍情報や批評を閲覧できるウェブサイト Goodreads における本を推薦する。本研究では、UCSD 公開データセット [46] から5万人のユーザをランダムに選択し、ユーザが読んだ本（アイテム）および本のメタデータを取得した。評価値はユーザがあるアイテムを読んだかどうかの2値情報である。トピックの抽出元となる文書はアイテムのタイトルと説明文である。

各データセットから構築した評価値行列に対して、表 4.1 に示す方法で評価値を削減することで、スパース度が異なる評価値行列を作成する。original は削減前の評価値行列、sample50% はランダムに 50% の評価値を削減した評価値行列である。max2u は各ユーザの評価値の個数が 2 になるまで評価値をランダムに削減した評価値行列で、現実にはサービス開始早期に起こるコールドスタートの状況に近い設定である。なお、表 4.2 中の評価値存在率は式 (3.4.1) で求めた。評価値存在率の値が小さいほど、評価値行列のスパース度は高くなる。

4.4.2 前処理

4.4.1 節で述べた各ドメインから収集したテキストに対して、単語単位に分解し、表記ゆれへの対処やノイズとなる単語の除去を行う。CiteULike と Goodreads の英語のテキストに対しては、英語のテキストに対して一般に用いられる以下の前処理を行う。

- 小文字に統一
- ステミング（単語の語形変化を取り除き、単語表現を揃える処理）
- ストップワード（he, the, this など）の除去

本実験では、ステミングには TreeTagger [80] を利用し、ストップワードの除去には SlothLib ^{*1} のストップワード一覧を利用した。

pixiv の日本語のテキストに対しては、形態素解析ツール MeCab[78] を利用して単語単位に分解する。この際、MeCab で固有名詞や専門用語を扱えるようにするため、MeCab で利用する辞書に以下を登録する処理を行った。

- Wikipedia に存在する項目名
- 専門用語（キーワード）自動抽出モジュール TermExtract[79] で抽出した、tweet 中によく出現する専門用語やフレーズ

加えて、テキストを単語単位に分解した後、品詞が名詞・形容詞・動詞と判定された単語のみを原形で抽出した。

4.4.3 語彙サブセットの設定

アイテムに関する文書集合から求まる語彙 W を用いて、表 4.3 に示す方法により語彙サブセットを作成する。本実験では、4.3 節で説明した文書頻度に基づく単語選択によ

^{*1} <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word>

表 4.3 語彙サブセットの設定 (作成方法)

Vocab All	すべての単語を選択 (語彙 W)
DF[0, 25]	DF 上位 25% までの単語を選択
DF[25, 50]	DF 上位 25% から 50% の単語を選択
DF[50, 75]	DF 上位 50% から 75% の単語を選択
DF[75, 100]	DF 上位 75% 以降の単語を選択
random	ランダムに $ W /4$ 個の単語を選択

り, DF[0, 25], DF[25, 50], DF[50, 75], DF[75, 100] の 4 つに区分した語彙サブセットを作成する. なお, DF[0, 25] が文書頻度の高い単語からなる語彙サブセット, DF[75, 100] は文書頻度の低い単語からなる語彙サブセットである. これまでに CTR に関して単語の文書頻度に関する研究がなく, 文書頻度を細かい単位に分割するほど分割当たりの語彙数が減少してしまい評価指標の結果の誤差が大きくなると考え, まずは文書頻度の大小が CTR の性能にどれだけ影響を与えるのかを調べるため 25% ずつの 4 つに分割した. また, 単語選択を行う前の W を Vocab All と呼ぶことにし, W からランダムに $|W|/4$ 個の単語を選択した語彙サブセットを random とする. 文書全体において, 重複を許して計測した単語の出現回数を総単語数と呼ぶ. 総単語数は語彙サブセット毎に異なるが, Vocab All を除いて語彙サブセットの単語数 (ユニーク単語数) は同じである. 語彙 W および各語彙サブセットの統計量を表 4.4 に示す.

4.4.4 ハイパーパラメータの設定

ハイパーパラメータの値はグリッドサーチにより決めるのが一般的である. CTR のハイパーパラメータは 6 個存在し, 各々の役割と探索に設定した値を以下にまとめる.

- K : トピック数. 潜在因子の数もこれと同じになる. [10, 50, 100] で実験.
- $\alpha: \theta$ の事前分布であるディリクレ分布のパラメータ. 1.0 に固定.
- $\lambda_u: u_i$ に対する正規化パラメータ. [1, 10, 100] で実験

表 4.4 語彙サブセットの統計量

語彙サブセットの種類	ユニーク単語数	総単語数
CiteULike		
Vocab All	28,104	4,246,722
DF[0, 25]	7,026	4,017,659
DF[25, 50]	7,026	137,055
DF[50, 75]	7,026	51,920
DF[75, 100]	7,026	40,088
random	7,026	955,922
pixiv		
Vocab All	23,932	1,143,829
DF[0, 25]	5,983	963,572
DF[25, 50]	5,983	103,910
DF[50, 75]	5,983	46,351
DF[75, 100]	5,983	29,996
random	5,983	298,364
Goodreads		
Vocab All	54,648	3,812,877
DF[0, 25]	13,662	3,429,311
DF[25, 50]	13,662	219,733
DF[50, 75]	13,662	95,376
DF[75, 100]	13,662	68,457
random	13,662	993,089

- $\lambda_v : v_j$ にトピック θ_j をどれくらい強く反映するかを調整するパラメータ. [1, 10, 100] で実験.
- a : 評価値の影響力を調整するパラメータ. [1, 5, 10, 50, 100] で実験.
- b : 欠損値の影響力を調整するパラメータ. 0.01 に固定.

表 4.5 MF と CTR のモデル学習にかかる時間の比較

	MF	CTR (Vocab All)	rate (CTR / MF)
K=10	267 [s]	325 [s]	1.22
K=50	704 [s]	1061 [s]	1.51
K=100	2758 [s]	3751 [s]	1.36

モデルの複雑さを調整する K はモデルの表現力に直結するため最も重要である。次に評価値の影響力を調整する a とトピックの影響力を調整する λ_v は、スパース度の影響を強く受けると予想される。本実験では、スパース度との関連が強いと予想される a と λ_v に焦点を置く。 a と b は相対値が重要であるため、 $b = 0.01$ という基準値を決め、それに対して相対的な値を a に設定した。今回使用するデータセットにおいて、 θ に対する仮定は置かないので α は事前情報を持たない一様分布となる 1.0 に固定する。 λ_u は値をどう設定するかは指針が存在しないため、他のハイパーパラメータとの関係を調べて決める必要がある。よって、4.5.1 節では、スパース設定ごとに $K, a, \lambda_v, \lambda_u$ の 4 つのハイパーパラメータに対してグリッドサーチを行う。

テキスト情報を用いる CTR は、テキスト情報を用いない MF よりも処理時間がかかる。ハイパーパラメータの設定によっても処理時間に差が生じ、特にトピック数が多くなると処理時間は長くなる。評価値行列は original で、トピック数 K を 10, 50, 100 としたときの処理時間は表 4.5 のようになる (CPU: Xeon E5-1607 v2, Memory: 64GB で実験)。この表から分かるように、MF よりも CTR の方が時間はかかるが、学習はオフラインで行えることも考慮し、本研究では推薦結果の良さ (すなわち正確性と利便性) にのみ注目する。

4.4.5 評価方法

各ドメインにおいて、収集した評価値集合を 5 分割し、1 つを最終的な評価に用いるためのテスト用データセット、4 つをハイパーパラメータを決めるための学習・検証用

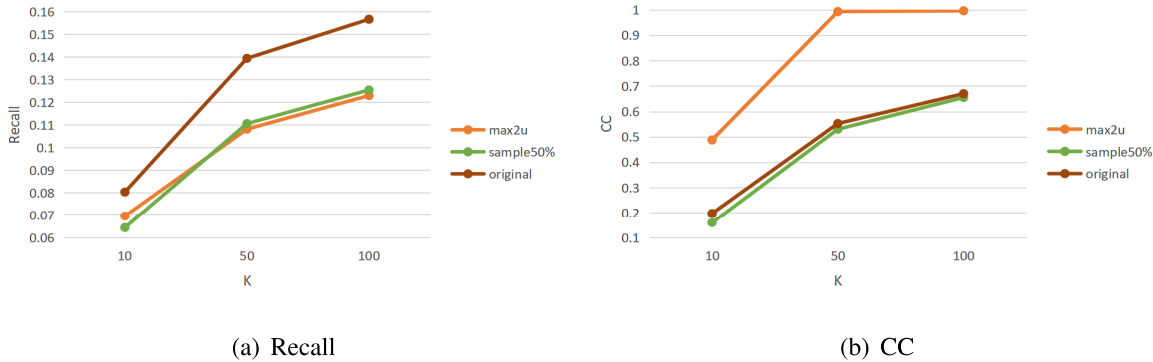


図 4.1 スパース度とトピック数 K の関係 (CiteULike)

データセットとする。学習・検証用データセットでは、ハイパーパラメータを決めるために4分割の交差検定を行う。正確性の評価指標には、3章と同じく再現率を利用する。利便性の評価指標には、推薦システムが偏りなく様々なアイテムを実際に推薦できているかを表す被覆率を採用する。本実験では、推薦アルゴリズムが算出した予測評価値 \hat{r}_{ij} を対象ユーザ i について降順に並べたときの上位 50 個のアイテムで推薦リスト L_i を構成する。

再現率 (*Recall*) は式 (3.4.3) で求められる。最終的なシステムの再現率は、全ユーザの *Recall* の平均を取ることで求める。

被覆率の指標としては、Catalogue Coverage (*CC*) [63] を用いる。CC は式 (3.4.4) で求められる。

4.5 結果と考察

4.5.1 ハイパーパラメータ間の関係性

各スパース度において、各ハイパーパラメータが性能に与える影響を分析する。紙面の都合により、CiteULike の結果のみを代表として掲載する。まずはトピック数 K による *Recall* と *CC* の変化を図 4.1 に示す。なお、 a , λ_v , λ_u は各 K において最も良かった値を設定している。どのスパース度においても K が 100 程度までは値が大きいほど *Recall*,

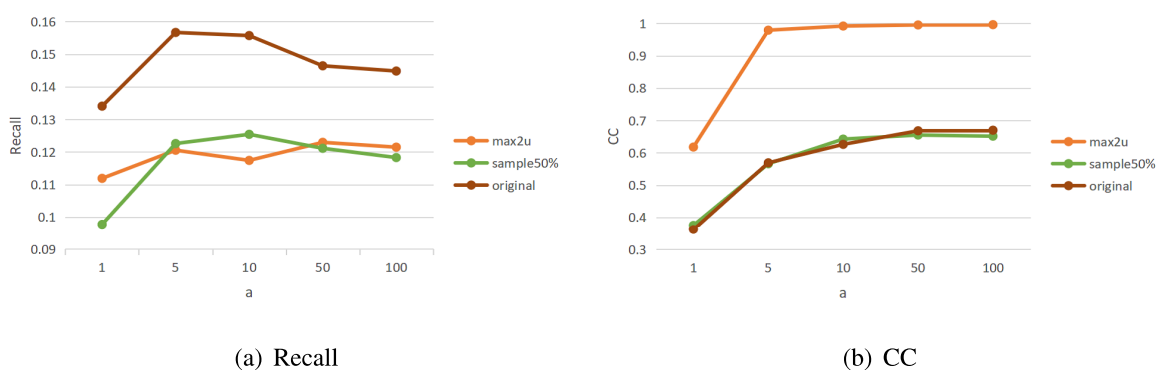


図 4.2 スパース度と評価値の影響力を調整する a との関係 (CiteULike)

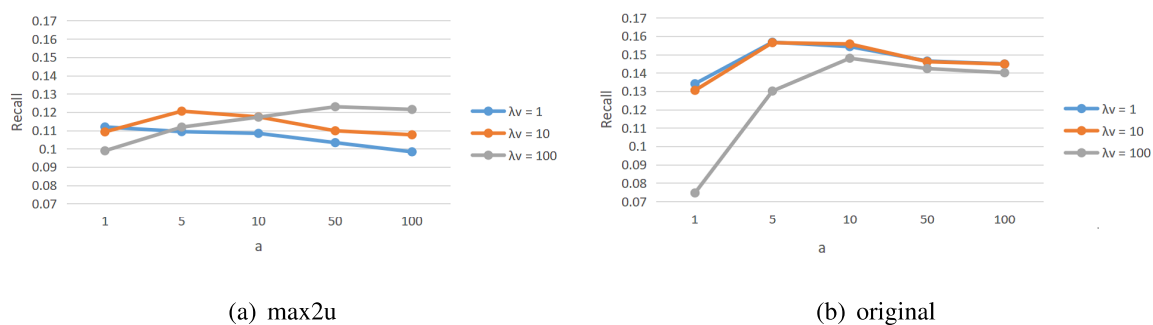


図 4.3 スパース度と評価値の影響力を調整する a とトピックの影響力を調整する λ_v の関係 (CiteULike). 左図はスパース度が非常に高い場合の正確性, 右図はスパース度が低い場合の正確性を表す

CC 共に良くなる傾向が見られる。適切な正則化の設定下においては、モデルの表現力が高い方が正確性と利便性のどちらも優れるのは自明な結果といえる。

正則化パラメータ λ_u による $Recall$ と CC の変化は、紙面の都合上グラフの掲載は省略するが、max2u において $Recall$ が上昇傾向であった点を除くと、他はすべてほぼ一定の値であった。そのため、他のハイパーパラメータに比べて重要度は低いといえる。なお、 $K = 100$ とし、 a と λ_v は各 λ_u において最も良かった値を設定している。

評価値の影響力を表す a による $Recall$ と CC の変化を図 4.2 に示す。なお、 $K = 100$ とし、 λ_v と λ_u は各 a において最も良かった値を設定している。スパース度によらず $Recall$, CC 共に値が小さい $a = 1$ が最も悪く、また $Recall$ は緩やかな凸形状をしているケースもあり a は高いほど良いというわけではなく最適な a を探索する必要がある。

また、 a と λ_v はそれぞれ評価値とトピックの影響力を調整するため、これらは値の相対的な大きさが重要であると考えられる。 a と λ_v の組み合わせによる *Recall* の変化を図 4.3 に示す。これを見ると、 $\lambda_v = 100$ のようにトピックの影響が強い設定では $a = 10$ を下回る辺りから明らかな *Recall* の低下が見られた。特に、スパース度が最も低い original において、 $\lambda_v = 100$ で $a = 1$ の *Recall* 低下が顕著である。この定性的な理由としては、評価値行列から得られる情報を軽視してトピックの比重が大きい潜在因子を学習したことで、元の評価値行列が持つ潜在因子（ある程度正しく学習が進んでいたものと思われる）からかけ離れたためと考えられる。

4.5.2 節以降では、文書頻度に基づく単語選択が CTR の性能にどのような影響を与えるかを明らかにするが、ハイパーパラメータの中でも文書情報をどれだけ MF の潜在因子に反映するかを表す λ_v は、特に結果への影響が強いと考えられる。そこで、4.5.2 節以降では λ_v との関係を中心に分析する。この分析において、 λ_v は 1, 10, 100 の 3 つを用いる。その他のハイパーパラメータはグリッドサーチにて設定しているが、各パラメータの範囲が適切に設定されているかどうか（すなわち最適な値が探索の範囲内に存在するかどうか）は、本節において範囲内でグラフが凸形状になっていることや、反応への変化がほとんどないか頭打ちになっていることを考慮して確認している。

4.5.2 スパース度と単語選択 (CiteULike)

CiteULike の結果に対して、スパース度と語彙サブセットの関係を分析し、その定性的な理由を考察する。使用する語彙サブセットの比較に加えて、トピックの影響力を調整するハイパーパラメータ λ_v を変化させることで語彙の影響力が弱い場合と強い場合を比較する。ただし、文書情報を利用しない MF における λ_v は単純に v_j に対する正規化パラメータとして機能する。なお 4.5.1 節より、これ以降 K と λ_u については他のパラメータに関わらず最良だった $K = 100$, $\lambda_u = 10$ に設定する。 a については λ_v とのバランスを考慮する必要があるため、指定の λ_v に対してグリッドサーチで求めた *Recall* が最も良く

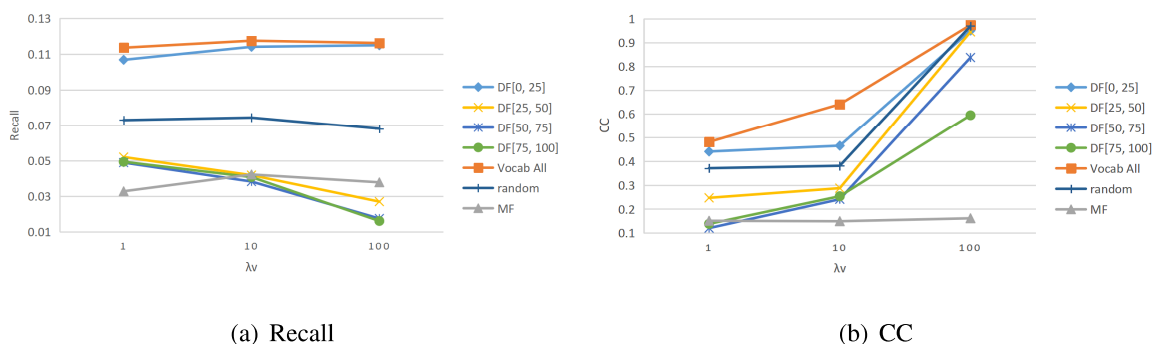


図 4.4 スパース度が非常に高い max2u における正確性と利便性 (CiteULike)

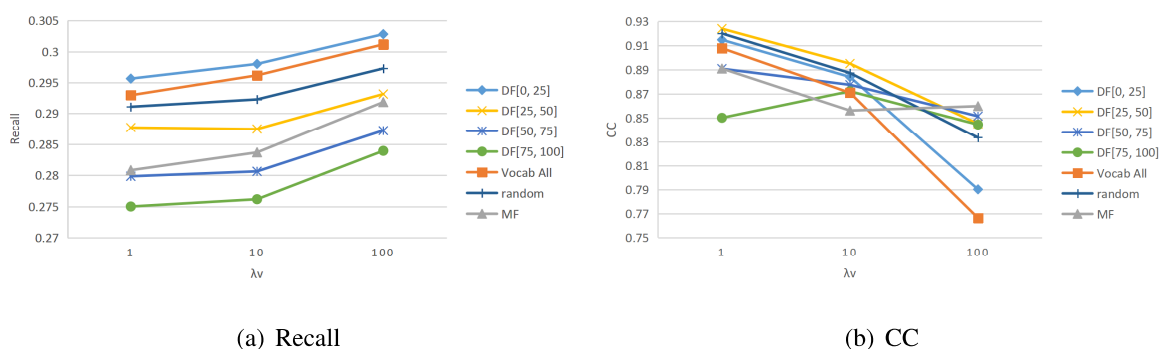


図 4.5 スパース度が非常に高い max2u における正確性と利便性 (pixiv)

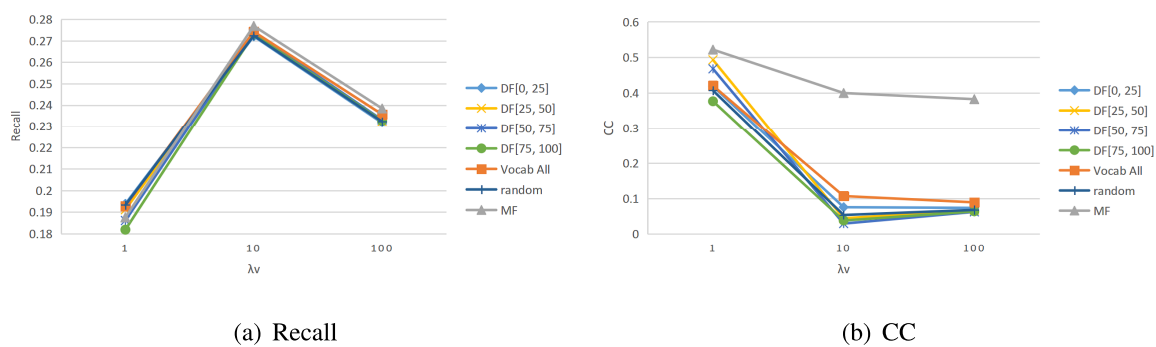


図 4.6 スパース度が非常に高い max2u における正確性と利便性 (Goodreads)

なる a を設定する。なお、結果の図は 4.5.3 節で示すデータセット間の違いを確認しやすくするため、スパース度ごとに全てのデータセットの結果を 1 ページ内で確認できるように提示する。

4.5.2.1 スパース度が非常に高いケース (max2u)

スパース度が非常に高い max2u の結果を図 4.4 に示す。まず *Recall* を見ると、Vocab All と DF[0, 25] では、 λ_v によらずベースラインである MF から非常に大きく改善している。文書頻度の高い DF[0, 25] には “network”, “genome”, “human” といった大まかなカテゴリを表す単語が多く含まれている。スパース度が非常に高い条件下において文書頻度の高い単語が有用である定性的な理由としては、評価値行列が極度にスパースでは協調フィルタリングが局所的にしか働かず、それらの局所間を繋げる大域的な情報を文書頻度の高い単語から抽出したトピックが与えたからだと考えられる。また、Vocab All と random を比較すると、総単語数が多い Vocab All (表 4.3 参照) の方が正確性が高い。一方、DF[25, 50], DF[50, 75], DF[75, 100] は λ_v の値を上げてトピックの影響力を強めると *Recall* が下がっていく。前述の通り評価値行列から抽出される情報が局所的であるのに対し、文書頻度の低いマイナーな単語から抽出されるも局所的であり、これらが競合しているため正確性が低下したと考えられる。

次に *CC* を見ると、文書情報を利用しない MF を除いて λ_v の上昇と共に大きく向上しており、語彙の影響を強めることで推薦対象のアイテムの種類が増えたことを表す。 λ_v の上昇と共に *Recall* も上昇している Vocab All と DF[0, 25] に注目すると、正確性を維持したまま利便性が向上していることが分かる。また *Recall* 同様に、Vocab All と DF[0, 25] は λ_v によらず他よりも優れている。特に、DF[0, 25] はユニーク単語数が少ないにもかかわらず、 $\lambda_v = 100$ では *Recall* と *CC* の両方で Vocab All と同等の性能である。以上より、正確性と利便性のどちらも良くするには、文書頻度の高い単語を選択する、もしくは単語の除外は行わず、 λ_v は大きい値に設定すれば良い。

4.5.2.2 スパース度が最も低いケース (original)

スパース度が最も低い original の結果を図 4.7 に示す。まず *Recall* を見ると、 λ_v が大きい場合はどの設定においても *Recall* が低下しており、その中でも Vocab All と DF[0,

25] が顕著である。これは評価値が十分に存在する状況（ウォームスタート）において MF より性能が低下する可能性がある CMF[38] と似た現象であり、この CTR においては潜在因子に反映されるトピックがノイズとなってしまう MF より性能が低下したと考える。その定性的な理由としては、評価値行列のスパース度が低いと協調フィルタリングが十分に働き局所的な情報も大域的な情報も抽出できており、どの文書頻度の単語においても、評価値行列の持つ情報に対してノイズとなってしまう性能へ悪影響を与えたと考えられる。なお、 $\lambda_v = 100$ での落ち込みは MF でも見られるので、文書情報による影響ではないと考えられるが、この原因は解明できなかった。Vocab All と random を比較すると、 λ_v が大きい場合は Vocab All が大きく劣っている。含まれる単語の文書頻度の比率にはほぼ差がないことを考慮すると、文書頻度によらずユニーク単語数と総単語数（表 4.3 参照）が多いと大きな低下に繋がっているといえる。また、 $\lambda_v = 100$ の時に DF[25, 50] は random より良く、文書頻度が中程度の単語は僅かだが正確性の向上に寄与している。

次に CC を見ると、どの文書頻度においても $\lambda_v = 100$ が良い。文書頻度毎に比較すると、Vocab All と DF[0, 25] が MF よりも悪く、これも *Recall* と同様の理由だと考えられる。一方、DF[25, 50]、DF[50, 75]、DF[75, 100] は大きく向上している。以上より、正確性を維持したまま利便性を得るには、文書頻度が中程度以下の DF[25, 50]、DF[50, 75]、DF[75, 100] の単語を選択し、 λ_v は中程度で *Recall* が低下し始めるポイントを探索すれば良い。

4.5.2.3 スパース度が中間のケース (sample50%)

最後にスパース度が original の約 2 倍である sample50% の結果を図 4.9 に示す。まず *Recall* を見ると、DF[0, 25] は λ_v が中程度の場合に正確性が高く、他の語彙セットと比べても高い。 $\lambda_v = 100$ の時は DF[25, 50] が DF[0, 25] と同程度の高い値になっている。DF[25, 50] 以外の語彙セットは、 λ_v の増加とともに減少する傾向が見られる。文書頻度の高い単語が多いほど正確性が高いという点では、max2u に近い傾向であるが、一方で

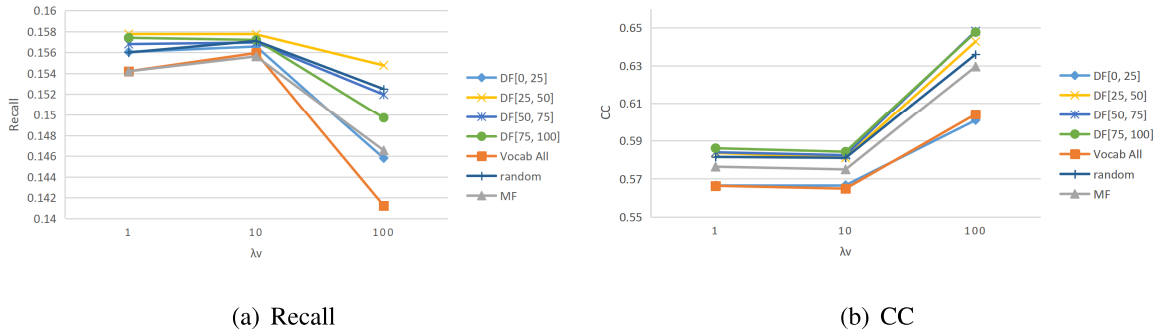


図 4.7 スパース度が最も低い original における正確性と利便性 (CiteULike)

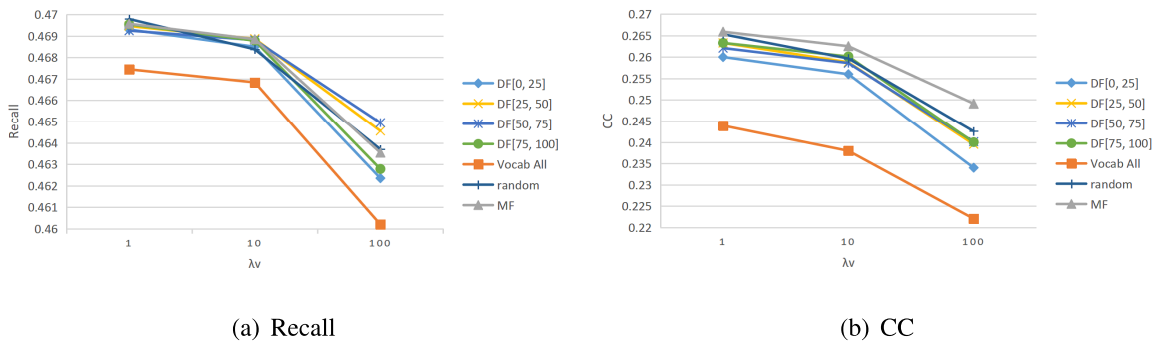


図 4.8 スパース度が最も低い original における正確性と利便性 (pixiv)

λ_v が大きい場合に DF[0, 25] と Vocab All が低下している点は original に近い傾向である。

次に CC を見ると、どの文書頻度においても $\lambda_v = 100$ が良く、特に DF[25, 50], DF[50, 75], DF[75, 100] は他よりも CC が高い。Vocab All と DF[0, 25] が MF よりも悪い点は original に近い傾向である。以上より、正確性と利便性のバランスを重視するなら、文書頻度が中程度の単語を選択し、 λ_v の値は大きいほうが良い。また、正確性を優先して利便性をやや犠牲にするなら、文書頻度の高い単語を選択し、 λ_v の値は中程度にするのが良い。

4.5.3 スパース度と単語選択による影響のデータセット間での比較

前節では CiteULike におけるスパース度と単語選択が推薦結果へ与える影響を見てきたが、他の異なるデータセットにおいても同様の傾向が見られるか検証する。pixiv に関しては、4.5.2 節の CiteULike との傾向の差異に焦点を置き、各スパース度において共通する傾向が存在するか分析を行う。Goodreads に関しては、この後で示す max2u の結果から文書情報による性能向上が見られなかったため、その理由についての分析と考察を行う。なお、ハイパーパラメータの設定に関しては、4.5.2 節と同様である。

4.5.3.1 スパース度が非常に高いケース (max2u)

pixiv の結果を図 4.5 に示す。文書頻度の高い単語を含む Vocab All と DF[0, 25] で *Recall* が良くなる傾向は CiteULike と同じである。これらの語彙頻度において、*CC* は λ_v が小さいほど良い傾向で、*Recall* とはトレードオフの関係になっており CiteULike とは異なる傾向である。*Recall* と *CC* のバランスの良い推薦を行うには、CiteULike では $\lambda_v = 100$ であるのに対し、pixiv では $\lambda_v = 10$ となる。以上より、バランスの良い推薦を行うための λ_v の値はデータセットにより異なるものの、共通点として文書頻度の高い単語を用いることで正確性と利便性のバランスが良い推薦を行えることが分かった。

次に Goodreads の結果を図 4.6 に示す。これを見ると、最も改善が見込めるスパース度が非常に高いケースでさえ MF より劣っている。この理由に関しては、4.5.4 節で分析と考察を行う。スパース度が非常に高い場合でも、*Recall* と *CC* の向上が見られなかったため、スパース度が中間の場合や低い場合にこれらが向上するとは思えない。よって、以降の節では Goodreads の結果は省略する。

4.5.3.2 スパース度が最も低いケース (original)

pixiv の結果を図 4.8 に示す。*Recall* は λ_v が小さい方が良く、文書頻度が中程度の単語が他と比べて良い傾向は CiteULike と同じである。しかし、各文書頻度においてベスト

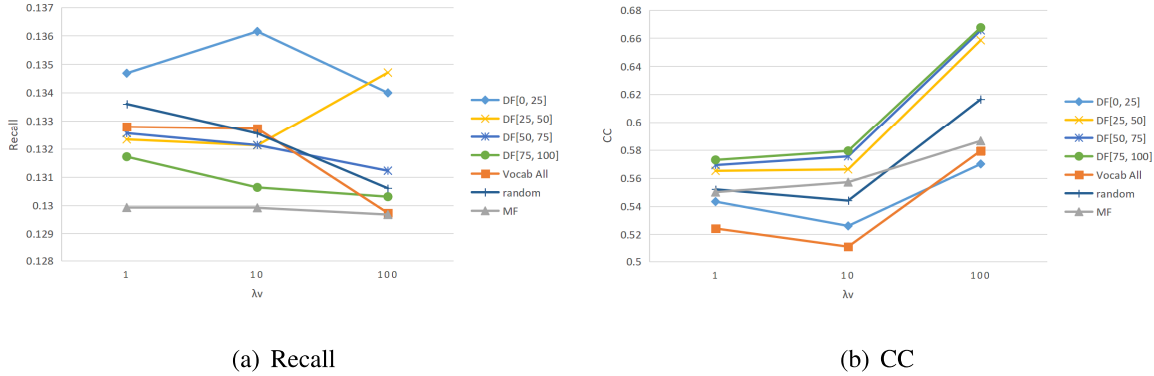


図 4.9 スパース度が中間の sample50% における正確性と利便性 (CiteULike)

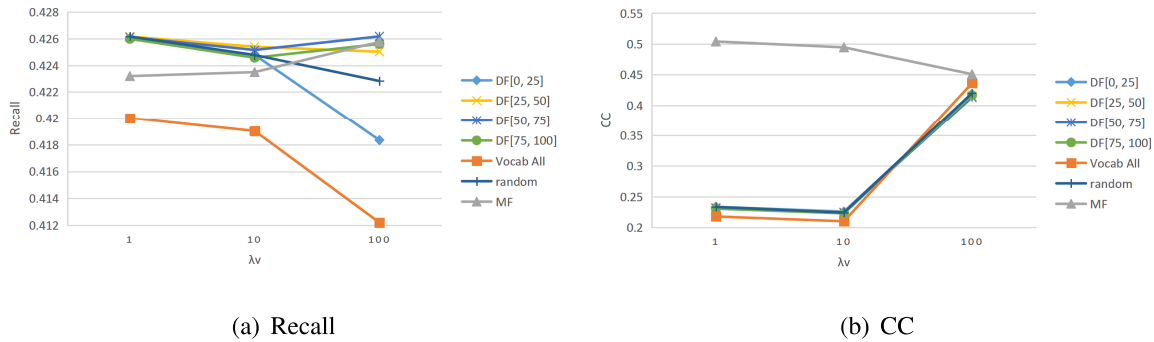


図 4.10 スパース度が中間の sample50% における正確性と利便性 (pixiv)

な値での MF との差は非常に小さく、文書情報を利用することによる性能向上の余地は小さい。CC は CiteULike と傾向が異なり、 λ_v が大きくなるほど低下しており、いずれの文書頻度においても MF より劣っている。Recall と CC のバランスの良い推薦を行うには、CiteULike では DF[25, 50] で $\lambda_v = 1$ または $\lambda_v = 10$ であるのに対し、pixiv では MF の $\lambda_v = 1$ となる。以上より、バランスの良い推薦を行うための λ_v の値はデータセットにより異なるものの、共通点として Recall だけに注目すると文書頻度が中程度の単語が正確性の向上に寄与していることが分かった。

4.5.3.3 スパース度が中間のケース (sample50%)

pixiv の結果を図 4.10 に示す。Recall は文書頻度が中程度以下の DF[25, 50], DF[50, 75], DF[75, 100] が良く、 $\lambda_v = 100$ で DF[25, 50] が良い傾向は CiteULike と同じである。

CC は MF 以外において λ_v が大きいと高くなる傾向は CiteULike と同じである。なお、CC の傾向が MF とそれ以外で大きく異なる原因は、 a の差異にあることが分かった。MF 以外において、 $\lambda_v = 1$ と $\lambda_v = 10$ では *Recall* が最も良くなるのは $a = 10$ であるのに対し、 $\lambda_v = 100$ では $a = 100$ であった。MF においては、どの λ_v でも $a = 100$ であった。*Recall* と CC のバランスの良い推薦を行うには、CiteULike と pixiv 共に DF[25, 50] で $\lambda_v = 100$ となる。以上より、共通点として文書頻度の中程度の単語で λ_v が大きいと正確性と利便性のバランスが良い推薦を行えることが分かった。

4.5.4 文書情報による性能向上が見込めないケース

Goodreads において、文書情報による性能が向上しなかった理由について分析と考察を行う。MF の協調フィルタリングが機能しにくくなる要因として、スパース度の他にもデータセットに依存した要因があると考え、評価値行列におけるアイテムの種類の違いを分析した。すなわち、人気アイテムが多くのユーザに評価されているか、逆に様々なアイテムが満遍なくユーザに評価されているデータセットであるかに着目した。まず初めに、各アイテムがどれだけのユーザに評価されているかをカウントし、ジニ係数を計算することでアイテムの種類の違いを定量化した。ジニ係数の値は 0 から 1 の間をとり、係数が 0 に近づくほど様々なアイテムが満遍なくユーザに評価されていることを意味し、1 に近づくほど一部のアイテムが多くのユーザに評価されていることを意味する。表 4.6 を見ると、pixiv と Goodreads は人気アイテムが多くのユーザに評価されていることが分かるが、Goodreads だけ MF が十分に機能する理由はこれだけでは分からない。

協調フィルタリングに影響を与えるのは評価値としてより多く出現する人気上位のアイテムであるが、ジニ係数だけでは人気上位のアイテムがどの程度の割合のユーザから評価されているか詳細が分からないため、これに焦点を当て分析を行った。図 4.11 は、あるアイテムをユーザ全体の何 % が評価しているかについて、その割合の大きい順にソートしたグラフである。max2u と original は共に、Goodreads は他 2 つのデータセットに

表 4.6 各データセットにおけるジニ係数の比較

	CiteULike	pixiv	Goodreads
max2u	0.16	0.69	0.73
original	0.28	0.77	0.78

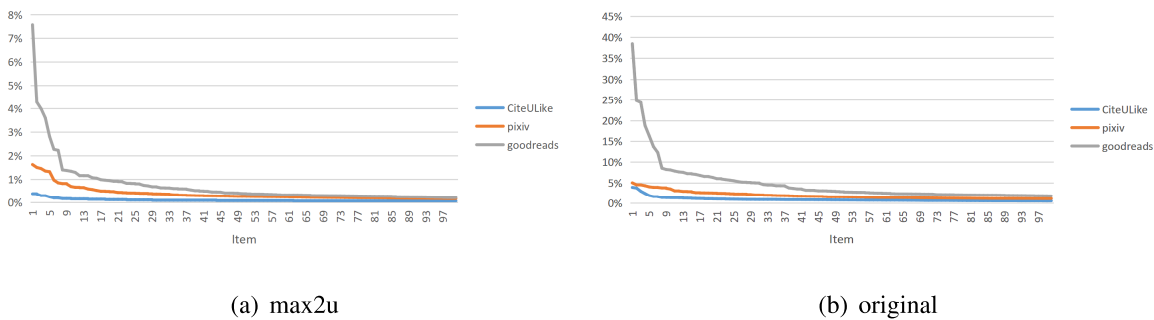


図 4.11 あるアイテムをユーザ全体の何 % が評価しているかについて、その割合の大きい順にソートしたグラフ. 上位 100 個のアイテムを表示

比べてより顕著に人気上位のアイテムが多くのユーザに評価されている。特に max2u はスパース度が高いため大域的な関係の抽出が困難な設定であるが、人気アイテムを通してより多くのユーザ間に関係性が生まれるため、Goodreads ではスパース度が高いことによる悪影響を受けなかったのだと考察する。以上より、ジニ係数が高く、さらに人気上位のアイテムが極端に多くのユーザに評価されているデータセットでは、MF が十分に可能性が高く、文書情報による性能向上は見込めない。

4.6 結論

評価値行列からだけでなくアイテムやユーザに関する文書から潜在因子を抽出するモデルの代表である CTR は、その性能を十分に活用するためには評価値行列のスパース度に応じて文書中の単語に対する適切な前処理と適切なハイパーパラメータ設定が必要となる。本章で行った実験では、CiteULike と pixiv では結果に共通する傾向が見られた。これらのデータセットは、ユーザ数とアイテム数の比や、評価値存在率に違いはあるが、

極端に人気のアイテムは存在しない特徴がある。このように極端に人気のアイテムは存在しないデータセットにおける、単語の選択方法とパラメータの設定方法を、以下のよう
にまとめる。

- 評価値が極端に不足する場合
 - 正確性と利便性のバランスが良い推薦を行うには、文書頻度が高い単語を選択する
- 評価値がある程度存在する場合
 - 正確性と利便性のバランスが良い推薦を行うには、文書頻度が中程度の単語を選択する
- 評価値が十分に存在する場合
 - 正確性は文書頻度が中程度の単語を選択することで良くなる可能性はあるが、MF と比べて顕著な差はなく性能向上の余地は小さい
 - λ_v の値が大きく、かつハイパーパラメータ a が小さすぎると正確性が極端に低下する

上記のように、スパース度と文書頻度の組み合わせに関しては、CiteULike と pixiv で共通する傾向が存在することが分かった。一方で、トピックの影響力を調整するハイパーパラメータ λ_v は CiteULike と pixiv で傾向が大きく異なり、事例毎に傾向の分析が必要になる。評価値が十分に存在する場合に正確性と利便性のバランスが良い推薦を行うには、CTR では文書頻度が中頻度の単語を選択すべきで、pixiv では MF を使用すべきであった。このように、評価値が十分に存在する場合は、MF を利用すべきか、CTR を利用する際には単語選択をどうすれば良いかについて慎重に実験を行い設定する必要がある。また、本章で行った実験では、Goodreads だけテキスト情報を利用しても、正確性と多様性に向上が見られなかった。CiteULike や pixiv との違いは、極端に人気のアイテムが存在する点にある。このことから、以下のまとめも示す。

- 極端に人気のアイテムが存在するデータセットでは、MF が機能しやすく、文書情報による性能向上は見込めない

推薦アルゴリズムを実サービスに適用する際、想定より性能が得られずそのアルゴリズムの適用を断念するといった事態が起こる可能性は存在する。本研究で得られた知見は CTR を利用する際、サービス開始初期で評価値が十分に得られていないか、ある程度の期間を経て評価値が十分に得られているか等の状況を考慮することで、評価値データのスパース度に起因して性能が十分に発揮されないという問題を解決するのに役立つと期待される。

第 5 章

語彙補完を活用した複数分野に関心を持つユーザを対象とする推薦

本章では、1.2.2 節で述べた課題 3 に対する取り組みとして、アイテムに関する文書とは異なる情報源を用いて文書の内容に合った単語を追加することにより、推薦性能が向上するケースが存在するか分析を行う。筆者は、推薦対象として、一部の分野にだけ深い興味を持つのではなく興味幅が広いユーザに着目する。興味幅が広いユーザは、ユーザが興味を持つ分野における人気の低いアイテムより、それに近い分野における人気の高いアイテムの方を好む傾向があると筆者は考える。論文・科学記事といった専門性の高いアイテムを扱うドメインでは、ユーザの履歴にあるアイテムと同分野で専門性の高いアイテムが推薦される傾向が見られた(3.5.5 節 iCTR の推薦結果がその一例)。そこで、文書に専門度がより低い単語を追加することで、対象の文書に近い分野の文書にも同じ単語が追加されて文書間に類似性が生じ、対象のアイテムに近い分野における人気アイテムが推薦対象に加わると仮定する。

本章の構成は以下となる。研究の動機となる考えや仮定および本章の目的を 5.1 節で述べる。関連研究として、を 5.2 節で示す。言語モデルを用いて、文書に単語を追加する手法についての説明を 5.3 節で行う。実験で使用するデータセットの説明やハイパーパラ

表 5.1 興味の幅が広いユーザがブックマークした論文・科学論文 (1/2)

(a) ユーザ A

- A Global Geometric Framework for Nonlinear Dimensionality Reduction
- A survey of shape analysis techniques
- Efficient Graph-Based Image Segmentation
- Spectral Grouping Using the Nystrom Method
- Matrix computations
- Tensor Decompositions and Applications
- Movement, activity and action: the role of knowledge in the perception of motion

(b) ユーザ B

- A fast local descriptor for dense matching
- Towards optimal bag-of-features for object categorization and semantic video retrieval
- Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories
- Content-based multimedia information retrieval
- Introduction to information retrieval
- Flickr Distance
- Self-Optimizing Memory Controllers: A Reinforcement Learning Approach
- Learning and Leveraging the Relationship between Architecture-Level Measurements and Individual User Satisfaction
- Flash storage memory
- Application Development Productivity Challenges for High-End Computing

メータの設定, 評価指標など実験設定を 5.4 節で説明する. 実験結果の提示と考察を 5.5 節で行う. 実験結果から導かれる結論を 5.6 節で述べる.

表 5.2 興味の幅が広いユーザがブックマークした論文・科学論文 (2/2)

(a) ユーザ C

- Biomedical ontologies: a functional perspective
- Bases de datos bibliograficas especializadas en Biologia
- PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites
- Text Mining
- Introduction to Data Mining
- Multivariate data analysis: The French way
- Learning with kernels : support vector machines, regularization, optimization, and beyond
- Four stages of a scientific discipline; four types of scientist
- How to write consistently boring scientific literature
- How To Choose a Good Scientific Problem

5.1 背景と目的

推薦システムはユーザの情報獲得を支援する方法の一つとして利用されている。年齢や性別といったデモグラフィックな属性の違いだけでなく、興味の異なる様々な人が利用する。論文や科学記事といった分野毎の専門性が高いアイテムの推薦においては、ある一部の分野にだけ深い興味を持つユーザもいれば、それとは逆に興味の幅が広いユーザもいると思われる。ソーシャルブックマークサービス CiteULike における興味の幅が広いユーザの例を表 5.1 と表 5.2 に示す。この表は彼らがブックマークを行った論文・科学記事を示している。ユーザ A は、画像処理分野の物体認識とセグメンテーションを主軸に、それらに必要な行列テンソルの理論や人の運動知覚に関する基礎研究に興味を持っている。ユーザ B は、マルチメディアの情報検索とコンピュータアーキテクチャに興味を持っている。ユーザ C は、データ分析や研究に必要な基礎的な理論や技術に興味を持っ

ているようである。より応用的な内容に興味がある開発者や広い見識が必要な管理者、基礎的な知識を必要としている学生などに対しては、専門的すぎるコンテンツの推薦は適していないと考える。

CTR は各文書に対して、どのような成分の内容で構成されているかを表すトピック比率を文書を構成する単語から推定する。しかし、分野が近くてもそれぞれの文書を構成する単語が非常に専門的だと文書間で共通する単語は少なくなるため、文書間のトピック比率の類似度は小さくなりやすい。トピック比率は潜在因子ベクトルに反映されるため、結果として分野が近くても文書間で共通する単語が少ないと推薦対象として選択されにくくなってしまう。その様な文書に対して、文書の内容に適した専門度のより低い単語を追加することができれば、狭い同分野の文書だけでなく近い分野の文書との類似度が大きくなる。興味の幅が広いユーザは、ユーザが興味を持つ分野における人気の低いアイテムより、それに近い分野における人気の高いアイテムの方を好む傾向があると筆者は考える。

本研究では、専門度のより低い単語をアイテムの文書に追加して CTR が扱う語彙を増やすこと（以降、語彙補完と呼ぶ）で、ユーザの興味に近い分野の人気が高いアイテムが推薦対象に加わり、興味の幅が広いユーザを対象とした推薦性能の向上を目指す。文書への単語の追加は、学習済みの言語モデル GPT-Neo 1.3B[67] で生成した文章に含まれる単語を用いる。言語モデルは与えられた単語列に対して、次に続く単語を確率的に生成することができる。GPT-Neo 1.3B は Web ページから得られた英文テキストで学習されており、生成される文章は論文や科学記事に比べると専門度は低くなる。実験では、論文や科学記事といった専門的な内容の文書に語彙補完を行うことで、興味の幅が広いユーザに対する推薦性能が向上することを検証する。また、4章の結果より、CTR ではスパース度によって性能に影響を与える単語の種類が異なることが判明したため、本実験においてもスパース度は高い場合と低い場合を設定する。評価指標には、推薦結果がユーザの嗜好に適しているかを評価する正確性、および全体として特定の人気アイテムだけが推薦されるような偏った推薦結果になっていないかを評価する利便性を利用し、この2

つの観点で評価を行う。この2つの評価指標はトレードオフの関係にあるため、バランスよく高い値となることが望ましい。以上、論文や科学記事といった専門性の高い文書に語彙補完を行って文書が持つトピックの幅を広げることでユーザの興味に近い分野のアイテムが推薦されるようになり、興味の幅が広いユーザに対してCTRの推薦性能を向上させられることを実験により検証したことが本研究の貢献である。

5.2 関連研究

5.2.1 特定のユーザにより適した推薦を行う手法

ユーザの属性についての情報をもとに推薦候補となるアイテムを限定することで、スケーラブルで精度の高い推薦を行うアプローチが提案されている。Lijiら[68]は、ユーザの年齢、性別、職業といった属性を用いてクラスタリングを行い、各クラスタにおいて、そのクラスタに属するユーザのみを対象としてユーザベースの協調フィルタリングを行った。Levandovski[69]らは、SNSから得られるユーザの位置情報を利用して、その場所や地域に適した推薦を行うモデルを提案した。Shivaramら[70]は、ニュース記事の推薦において、あるトピックでは保守的な記事を好み、別のトピックではリベラルな記事を好むといったように、トピックによって政治的嗜好が異なるユーザに焦点を当てた。トピック特有の政治的嗜好を表す単語を考慮できるよう、Attentionベースの深層学習モデルを提案した。

これらの研究はユーザに関する情報の利用[68, 69]や、トピック毎にユーザの嗜好を推定[70]することで特定のユーザにより適した推薦を行っている。本研究ではユーザが好むアイテムに関する情報を補強してアイテム間の類似性を変化させることで興味の幅が広いユーザに適した推薦を目指す。

5.2.2 学習済み言語モデルを活用した推薦に関する研究

近年、BERT [71] を始め、XLNet [72], GPT3 [73], GPT-Neo [67] といった深層学習のモデルを利用して Web 上で得られる大規模なテキストデータで事前学習を行った学習済み言語モデルが公開されており、自然言語処理の分野だけでなく推薦システムの分野の研究においても活用したケースがいくつか存在する。Gustavo ら [74] は、BERT を利用した学習済みの言語モデルが、本、映画、音楽など推薦されるアイテムについてどの程度の情報を有しているか分析した。会話推薦において、BERT が関連性のある返答であるか区別できること、マルチタスク学習によりファインチューニングすることで性能向上することが分かった。Damien [75] らは、学習済み言語モデルのみを利用して推薦がどの程度行えるか評価した。ユーザが好む映画のタイトルを列挙したテキストに対して、その末尾に推薦候補の映画のタイトルが出現する確率をその映画の予測評価値とした。ユーザ数が 100 未満の MF に相当する性能を有しており、コールドスタート問題に活用できる可能性を示唆した。

学習済みの言語モデルを推薦に活用できるか調査したこれらの研究 [74, 75] は、言語モデルの出力を推薦に直接利用している。本研究では言語モデルを利用して学習データを補強するアプローチをとっており、アイテムに関する情報の補強に学習済みの言語モデルが有用であることを検証する。

5.3 語彙補完手法

アイテムに関する文書に含まれる単語と言語モデルで生成した文章に含まれる単語を合わせることで、語彙補完を行った bag-of-words 形式の文書を作成する。本研究では言語モデルとして、深層学習の一種である Transformer をベースとしたモデルにより自然な文章を生成することができる GPT-Neo の学習済みモデルである GPT-Neo 1.3B [67] を利用する。GPT-Neo 1.3B はオープンソースであるため利用に制限がなく、大規模言語コー

パス Pile で学習されている。Pile [67] は一般の Web サイト、PubMed や ArXiv などアカデミックなサイト、Github, Wikipedia などから得られる様々な内容のテキストで構成されており、アカデミックな内容以外のテキストが大半であるため、生成される文章は論文に比べると専門度は低くなると考えられる。本研究における言語モデルの利用方法としては、言語モデルにクエリとなるテキストを与え、それに続く文章を生成する。例えば、クエリとして “Melos was enraged. He determined he would definitely get rid of the cruel, ruthless king.” を与えると、言語モデルは “He’d make it so that no one would dare to harm him again. He wanted to put an end to all the torment that king had caused him and his clan.” といったクエリの文脈に沿った文章が生成される。語彙補完を行った文書の作成は以下の手順で行う。

1. 各文書に対して,
 - a. クエリとなるテキストを選択する。
 - b. クエリに続く文章を言語モデルで生成する。
2. 生成した文章に対して、TF-IDF で特徴的な単語を抽出する。
3. 抽出した単語とアイテムに関する文書の単語を合わせたものを語彙補完を行った文書とする。

アイテムに関する文書に出現する単語より専門度が低い単語を追加することが語彙補完の目的であるが、あらゆる文書で出現するような一般的な単語は逆にノイズとなる可能性があるため好ましくない。そのため、上記 2. の処理により、様々な文書に出現する単語を除外している。また、CTR の入力として与える文書は bag-of-words 形式であるため、言語モデルで生成した文章に対して単語単位で選択して文書に追加することが可能である。

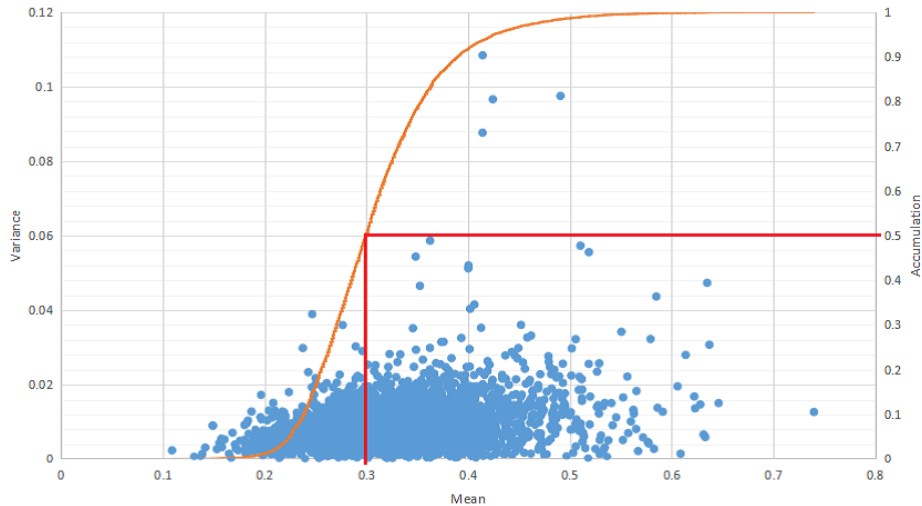


図 5.1 ユーザーの興味の幅を表すスコアの分布. 横軸と縦軸の左側は, それぞれユーザーが持つアイテム間の類似度の平均と分散を表す. 縦軸の右側 (橙色の曲線) はユーザーの累積を表す. 横軸は小さいほどユーザーの興味の幅が広いことを意味する.

表 5.3 スパース度の設定と評価値行列の統計量

	評価値存在率 (%)	ユーザー数	アイテム数	評価値の削減方法
original	6.0e-4	6,877	52,550	削減なし (オリジナルの評価値行列)
max2u	3.6e-5	6,877	10,594	各ユーザーの評価値数が 2 個になるまで削減

5.4 実験設定

5.4.1 データセットとスパース度の設定

本実験では, ソーシャルブックマークサービス CiteULike における論文・科学記事を推薦する. 2014 年 9 月から 11 月の間にブックマークを行ったユーザーとそのユーザーが登録したすべての論文・科学記事 (アイテム) のタイトルと要約情報を直接クロールすることで収集した. 評価値はユーザーがアイテムにブックマークしたかどうかの 2 値情報であり, ユーザーが対象アイテムに関心がある場合は 1, 関心がないもしくは存在を知らない場合は 0 となる. トピックの抽出元となる文書は, アイテムのタイトルと要約情報で

ある。ユーザ数とアイテム数およびスパース度を表 5.3 の original に示す。

次に、ここで用意した CiteULike データセットを元に、興味の幅が広いユーザだけを抽出したデータセットを作成する。ユーザがブックマークを行ったアイテムの分野の偏りを求め、偏りが小さいユーザを興味の幅が広いユーザとして分類する。手順の詳細を以下に示す。

1. アイテムの特徴ベクトルを作成する。
 - Doc2Vec[76] を利用し、文書の分散表現をアイテムに関する文書から作成する。
2. ユーザが持つアイテム間の類似度をすべての組み合わせで計算し、その平均をユーザの興味の広さを表すスコアとする。
 - アイテム間の類似度は、1. で求めた特徴ベクトルのコサイン類似度とする。
3. 図 5.1 に示すユーザのスコアの分布より、ユーザ全体の約半数となる横軸の値が 0.3 以下のユーザを抽出し興味の幅が広いユーザのデータセットとする。

また、4 章の結果より、評価値行列のスパース度と利用する語彙の組み合わせは推薦性能に影響を与えることが分かっている。各データセットから構築した評価値行列に対して、表 5.3 に示す方法で評価値を削減することで、スパース度が異なる評価値行列を作成する。original は削減前の評価値行列、max2u は各ユーザの評価値の個数が 2 になるまで評価値をランダムに削減した評価値行列で、現実にはサービス開始早期に起こるコールドスタートの状況に近い設定である。なお、表 5.3 中の評価値存在率は式 (3.4.1) で求めた。評価値存在率の値が小さいほど、評価値行列のスパース度は高くなる。

5.4.2 語彙補完の設定と文書に対する前処理

5.3 節の語彙補完において、具体的に設定する必要がある箇所については以下のよう

に設定した。

- クエリとなるテキストは、CiteULike の論文・記事のタイトルを用いる。
- クエリに続く文章の長さは、50 単語以上 100 単語以下の制限を設定した。
- 生成した文章に対して行う TF-IDF では、スコア上位 50 単語を抽出した。

また CTR に入力する文書への前処理として、表記ゆれへの対処やノイズとなる単語の除去を行う。CiteULike のテキストに対しては、英語のテキストに対して一般に用いられる以下の前処理を行う。

- 小文字に統一
- ステミング（単語の語形変化を取り除き、単語表現を揃える処理）
- ストップワード（he, the, this など）の除去

本実験では、ステミングには TreeTagger [80] を利用し、ストップワードの除去には SlothLib ^{*1} のストップワード一覧を利用した。

5.4.3 ハイパーパラメータの設定

CTR のハイパーパラメータは 6 個存在し、各々の役割と探索に設定した値を以下にまとめるとめる。

- K : トピック数。潜在因子の数もこれと同じになる。100 に固定。
- α : θ の事前分布であるディリクレ分布のパラメータ。1.0 に固定。
- λ_u : u_i に対する正則化パラメータ。10 に固定。
- λ_v : v_j にトピック θ_j をどれくらい強く反映するかを調整するパラメータ。[1, 10, 100] で実験。
- a : 評価値の影響力を調整するパラメータ。[1, 10, 100] で実験。
- b : 欠損値の影響力を調整するパラメータ。0.01 に固定。

^{*1} <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word>

ハイパーパラメータの値はグリッドサーチにより決めるのが一般的である。4章の結果より、CiteULikeを対象とした実験では、評価値の影響力を調整する a とトピックの影響力を調整する λ_p が推薦性能への影響が大きく、入力によって適切な値が大きく変動するため探索を行う。

5.4.4 評価方法

各データセットにおいて、評価値集合を5分割し、1つを最終的な評価に用いるためのテスト用データセット、4つをハイパーパラメータを決めるための学習・検証用データセットとする。学習・検証用データセットでは、ハイパーパラメータを決めるために4分割の交差検定を行う。正確性の評価指標には、4章と同じく再現率を利用する。利便性の評価指標には、4章と同じく被覆率を採用する。本実験では、推薦アルゴリズムが算出した予測評価値 \hat{r}_{ij} を対象ユーザ i について降順に並べたときの上位50個のアイテムで推薦リスト L_i を構成する。

再現率 (*Recall*) は式 (3.4.3) で求められる。最終的なシステムの再現率は、全ユーザの *Recall* の平均を取ることで求める。

被覆率の指標としては、Catalogue Coverage (*CC*) [63] を用いる。CC は式 (3.4.4) で求められる。

5.5 結果と考察

5.5.1 スパース度が低いケース (original)

本節では、スパース度が低い original のデータセットにおける結果と考察を述べる。興味の幅が広いユーザを対象に行なった結果 (図 5.2) から、語彙補完の有無による影響を分析する。図の凡例については、語彙補完を行わない文書集合を利用する CUL, GPT-Neo による語彙補完を行う文書集合を利用する CUL + GPT-Neo, 文書集合を利用

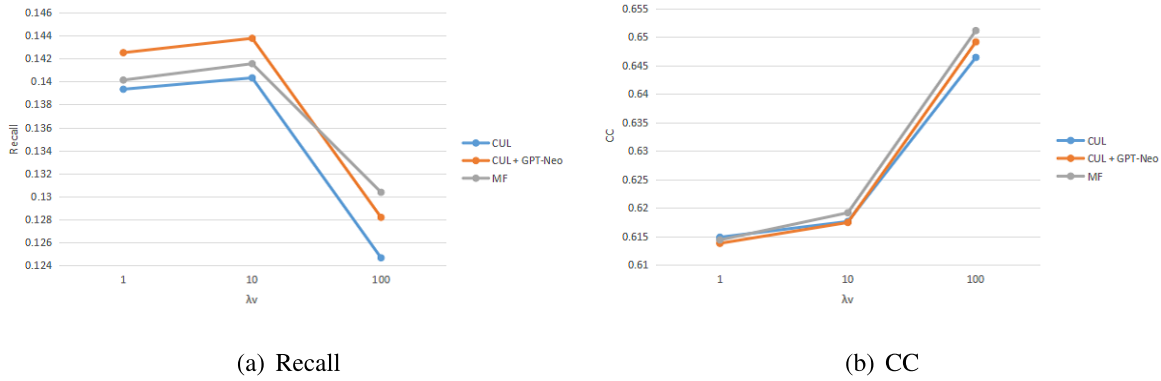


図 5.2 スパース度が低い original における正確性と利便性 (興味幅が広いユーザーが推薦対象)

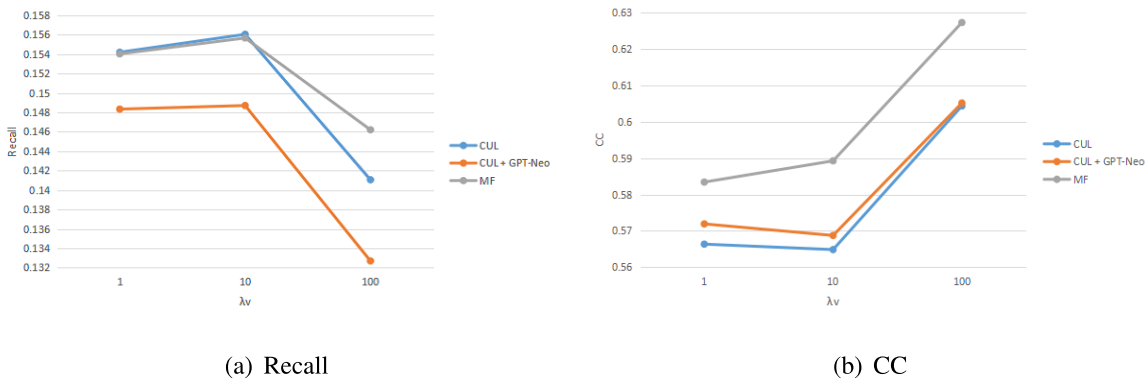


図 5.3 スパース度が低い original における正確性と利便性 (全ユーザーが推薦対象)

しない MF を表す。Recall を見ると、語彙補完の有無による比較では λ_v によらず語彙補完を行う方が優れており、語彙補完を行わない場合に対する語彙補完を行う場合の相対的な Recall の向上は 2.3 ~ 2.8% ほどである。いずれの手法においても値が最も良い $\lambda_v = 10$ で比較すると、語彙補完を行うと MF より向上することから、補完された語彙は正確性の向上に寄与していることが分かる。一方、語彙補完を行わない場合は λ_v によらず MF より劣っていることから、アイテムに関する文書の語彙だけでは正確性が低下することが分かる。次に CC を見ると、語彙補完の有無による比較では $\lambda_v = 100$ で語彙補完を行う場合の相対的な CC の向上は 0.4% ほどで Recall のように明らかな向上であるとは言い難く、 $\lambda_v = 1$ では語彙補完を行わない方が僅かに高い。CC については、語彙補

完の有無による変化はあまりなく利便性は維持されていると分かる。また、*CC* が維持されているので推薦されたアイテムの種類総数は変わっておらず、かつ *Recall* が向上したことから、興味の幅が広いユーザに適さないアイテムの推薦が減り、適したアイテムの推薦が増加したと考えられる。

興味の幅が広いユーザを対象に行なった結果と全ユーザを対象に行なった結果を比較することで、推薦対象のユーザの違いによる語彙補完の影響を調べる。全ユーザを対象に行なった結果を図 5.3 に示す。*Recall* を見ると、語彙補完の有無による比較では λ_v によらず語彙補完を行う方が劣っており、語彙補完を行う場合の相対的な *Recall* の低下は 3.8 ~ 5.9% ほどである。興味の幅が広いユーザを対象に行なった結果 (図 5.2) の *Recall* の傾向を比較すると、語彙補完の有無による優劣が逆転している。これは語彙補完によって、興味の幅が広くないユーザに対しては正確性が悪化したことを意味している。次に興味の幅が広いユーザを対象に行なった結果 (図 5.2) と全ユーザを対象に行なった結果 (図 5.3) の *CC* の傾向を比較すると、 $\lambda_v = 1$ では語彙補完の有無による優劣は逆転している。また、全ユーザを対象に行なった結果 (図 5.3) の *CC* は MF より劣り、 $\lambda_v = 1, 10$ では語彙補完を行う方が少し良くなる程度である。

以上より、スパース度が低いデータセットで興味の幅が広いユーザを対象に推薦を行う場合、言語モデルによる語彙補完は利便性を維持しつつ正確性を向上させることができる。これは興味の幅が広いユーザに適さないアイテムの推薦が減り、適したアイテムの推薦が増加したためと考えられる。一方で、興味の幅が広くないユーザを対象とした推薦を行う場合、正確性は逆に低下してしまうことに注意する必要がある。

5.5.2 スパース度が非常に高いケース (max2u)

本節では、スパース度が非常に高い max2u のデータセットにおける結果と考察を述べる。興味の幅が広いユーザを対象に行なった結果を図 5.4 に示す。なお、このスパース度では評価値の量が非常に少ないため、評価値の情報だけを利用する MF の *Recall* と

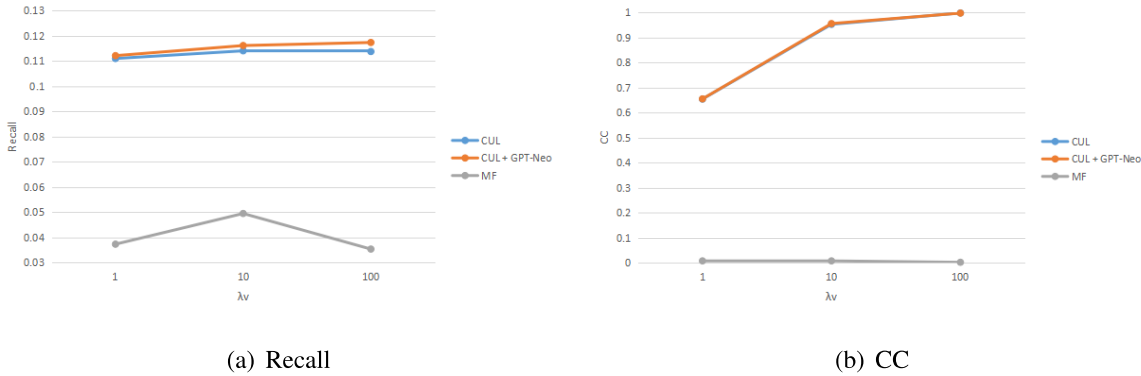


図 5.4 スパース度が非常に高い max2u における正確性と利便性 (興味の幅が広いユーザーが推薦対象)

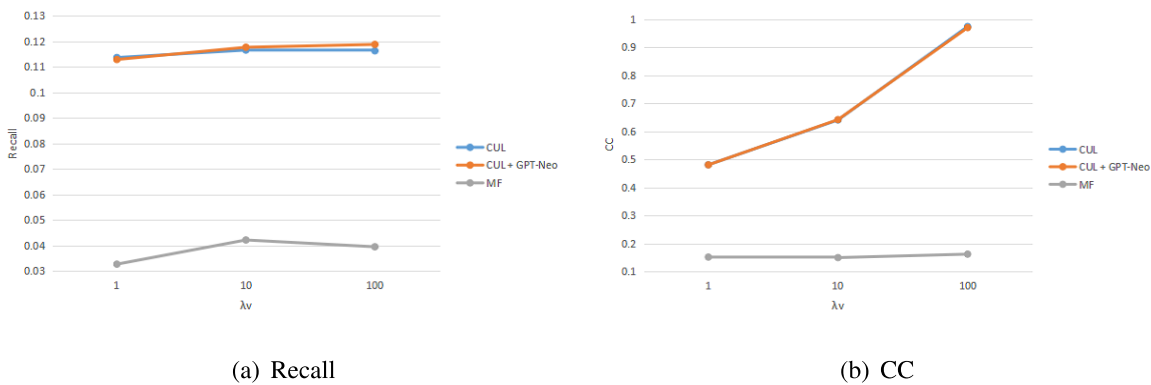


図 5.5 スパース度が非常に高い max2u における正確性と利便性 (全ユーザーが推薦対象)

CC の値はどちらも顕著に低くなっている。Recall を見ると、語彙補完の有無による比較では λ_v によらず語彙補完を行う方が優れており、語彙補完を行う場合の相対的な Recall の向上は 1.1 ~ 2.4% ほどである。次に CC を見ると、語彙補完の有無による比較では λ_v によらず値は同程度である。スパース度が低いケース (図 5.2) と比較すると、Recall は向上の度合いがやや小さくなったものの語彙補完を行う方が優れるのは同じ傾向であり、CC にはあまり変化がない傾向も同様である。

次に、全ユーザーを対象に行なった結果を図 5.5 に示す。Recall を見ると、 $\lambda_v = 1$ を除いて語彙補完を行う方が優れており、語彙補完を行う場合の相対的な Recall の向上は 0.6 ~ 1.3% ほどである。興味の幅が広いユーザーを対象に行なった結果 (図 5.4) と比べる

と、語彙補完による向上の度合いは小さいことが分かる。次に *CC* を見ると、語彙補完の有無による比較では λ_v によらず値は同程度である。

以上より、スパース度が非常に高いデータセットで興味の幅が広いユーザを対象に推薦を行う場合、言語モデルによる語彙補完は利便性を維持しつつ正確性を向上させることができる。ただし、スパース度が低いデータセットに比べると向上の度合いは小さい。

5.5.3 推薦例の比較

語彙補完を行う場合の推薦結果と語彙補完を行わない場合の推薦結果の例を比較することで、語彙補完が推薦結果に与える影響に関して事例の考察を行う。語彙補完の有無での *Recall* の向上度合いが上位のユーザの中から、筆者の専門に近いユーザ（ユーザ D）を選択した。ユーザ D がブックマークしたアイテム一覧を表 5.4 に示す。表 5.4 を見ると、主に“MF に関連する詳細な技術”と“検索の個別化”の分野のアイテムをブックマークしており、それに近い分野である“推薦システム”のアイテムもいくつか見られる。推薦結果において上位 50 個のアイテムを並べた推薦リストを語彙補完の有無でそれぞれ作成し、双方に出現したアイテムを除外した推薦結果の差分リストを作成することで、語彙補完が推薦結果に与える差異を分析する。ユーザ D における、語彙補完を行う場合の推薦結果の差分リストを表 5.5-(a) に、語彙補完を行わない場合の推薦結果の差分リストを表 5.5-(b) に示す。表 5.5-(a) は表 5.5-(b) に比べて、“Recommender”または“Collaborative Filtering”をタイトルに含むアイテムが多いことから、ユーザ D がブックマークした主要分野に近い分野である“推薦システム”に関連するアイテムは語彙補完を行う方が推薦されやすくなったことが分かる。表 5.5-(b) を見ると、“ベイズ統計のモデルや最適化手法”に関連するタイトルが多く、これらは“MF に関連する詳細な技術”の基礎として関連があり、他には web を含むタイトルが見られ、どちらもユーザ D がブックマークした主要分野に合致している。しかし、主要分野のアイテムが多く推薦される一方で、表 5.5-(a) に比べて“推薦システム”のアイテムが推薦されにくくなっている。よって、語彙補完を

行うことで“推薦システム”に関連するアイテムが推薦される傾向になったことにより、テスト用データセットに含まれていたアイテム“Google News Personalization: Scalable Online Collaborative Filtering”が推薦され、*Recall*の向上に寄与している。

5.6 結論

評価値からだけでなくアイテムやユーザに関する文書から潜在因子を抽出してMFに反映するモデルの先駆けであるCTRは、文書に含まれる単語によって性能が大きく変化する。本研究では、CTRにおいて論文や科学記事といった専門性の高い文書に対して専門度のより低い単語を追加することで、興味の幅が広いユーザを対象とした推薦性能の向上を目指した。具体的には、CiteULikeからクローリングを行い収集した文書に対して、webページで学習した言語モデルが生成する文章を用いて語彙補完を行ったところ、興味の幅が広いユーザに対する推薦の正確性が向上することを確認した。評価値行列のスパース度による影響も分析し、スパース度が高く十分な量の評価値が存在するデータセットでは改善が見られ、スパース度が低くコールドスタート状態のデータセットでも効果は小さいものの同様の傾向が見られた。一方で、本研究における語彙補完は、興味の幅が広くないユーザに対しては原理的に適していない。実際に、推薦対象を興味の幅が広いユーザに限定しない場合、専門度のより低い単語での語彙補完は正確性の低下に繋がる結果となった。実用に際しては、広い視点を必要とする上流設計に携わるエンジニアやプロジェクトマネージャなどの立場にあるユーザに対して適用することで、推薦のミスマッチを減らすことが期待できる。

表 5.4 ユーザ D がブックマークしたアイテム一覧

(a) 学習・検証用データセット

- Diffusion kernels on graphs and other discrete input spaces
- Structured Machine Learning: The Next Ten Years
- Orthogonal nonnegative matrix t-factorizations for clustering
- Automatic identification of user interest for personalized search
- Evaluating the Effectiveness of Personalized Web Search
- Bayesian Sets
- A large-scale evaluation and analysis of personalized search strategies
- User language model for collaborative personalized search
- Interest-based personalized search
- Learning and Evaluating Classifiers Under Sample Selection Bias
- Document Clustering Based on Non-negative Matrix Factorization
- Online-updating regularized kernel matrix factorization models for large-scale recommender systems
- Shallow parsing with conditional random fields
- Implicit user modeling for personalized search
- Matrix Factorization Techniques for Recommender Systems
- Yes, there is a correlation: - from social networks to personal behavior on the web
- The wisdom of the few: a collaborative filtering approach based on expert opinions from the web
- Persona: a contextualized and personalized web search
- A Fast Learning Algorithm for Deep Belief Nets

(b) テスト用データセット

- Learning the parts of objects by non-negative matrix factorization
- Algorithms for Non-negative Matrix Factorization
- Personalized Web search for improving retrieval effectiveness
- Feature-based recommendation system
- Google News Personalization: Scalable Online Collaborative Filtering

表 5.5 ユーザ D に対する語彙補完の有無それぞれの推薦結果において、双方に出現するアイテムを除外した推薦結果の差分リスト

(a) 語彙補完あり

テスト用データセットに含まれていたか	タイトル
○	- Evaluating Collaborative Filtering Recommender Systems
	- Hybrid Recommender Systems: Survey and Experiments
	- Scalable collaborative filtering using cluster-based smoothing
	- Recommender systems
○	- Learning the parts of objects by non-negative matrix factorization
	- Beyond Recommender Systems: Helping People Help Each Other
	- Trust-aware recommender systems
○	- Google News Personalization: Scalable Online Collaborative Filtering

(b) 語彙補完なし

テスト用データセットに含まれていたか	タイトル
	- Learning to cluster web search results
	- Probabilistic latent semantic indexing
	- Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions
	- A View of the EM Algorithm that Justifies Incremental, Sparse, and other Variants
	- Unsupervised learning of finite mixture models
	- A Taxonomy of Recommender Agents on the Internet
	- Learning to Rank Using Gradient Descent
	- Clustering the tagged web

第 6 章

結論

本章では、これまで各章で述べてきた個別分析の結論を課題毎の形で整理する。本論文の冒頭 1.2.2 節では、本研究における 3 つの課題を定義した。6.1 節では、2 章から 5 章で行った研究から判明した、それぞれの課題に対する結論を順に述べる。最後に、本研究をふまえて筆者が考える今後の展望を 6.2 節で述べる。

6.1 課題と結論

6.1.1 課題 1 に対する結論

本研究では、課題 1 “ユーザに関する文書を用いる場合の推薦性能についての詳細な分析” への取り組みとして、ユーザに関する文書から抽出されるトピックの分析 (2 章) とトピック抽出元となる文書の種類が推薦性能に与える影響の分析 (3 章) を行った。2 章では、ソーシャル・ネットワーキング・サービス (SNS) の一種である Twitter から取得したデータを分析することで、ユーザに関する文書から抽出したトピックへの理解を深めるための分析を行った。アイテムに関する文書、ユーザに関する文書、アイテムに対する擬似的なレビュー文書という 3 つの情報源において、トピックを構成する上位単語の分析および可視化結果の妥当性と発見性を比較することでトピックの差異を分析した。

結果、ユーザに関する文書から抽出されたトピック内容はユーザが興味を持っている対象やユーザの属性であり、これらはユーザ層という観点を表していることが分かった。これはアイテムに関する文書から抽出されるアイテムのジャンルを表すトピックとは明確に内容が異なり、なおかつユーザに関する文書を用いた場合の可視化結果は妥当性と発見性に優れていたため、CTR のトピック抽出のための新たな情報源として有益である可能性が高いと考えた。次に、3章では、論文・科学記事のソーシャルブックマークサービス CiteULike から取得したデータを使用して、ユーザに関する文書を用いる場合の推薦性能は従来のアイテムに関する文書を用いる場合と比べてどのような差異が生じるか分析した。実験において、総合的な観点から評価を行うために、正確性に関する評価指標として再現率を、利便性に関する評価指標として被覆率と多様性を用いた。評価値の欠損にどれだけ対応できるかという点にも注目し、評価値行列のスパース度を変化させた際の性能を調べた。結果、アイテムに関する文書を用いた場合では利便性の高い推薦、ユーザに関する文書を用いた場合では正確性の高い推薦を行う傾向が分かった。ユーザに関する文書からはユーザ層という観点を表すトピックが得られるため、あるユーザ層はどのようなアイテムを好むかという観点で推薦が行われることで、推薦されるアイテムの多様性は減ったもののユーザの嗜好から外れる可能性が低い無難な推薦を行っているのではないかと考察する。加えて、ユーザに関する文書にはそのユーザの興味が明示されていることが多く、ユーザの嗜好を直接的に得ることができる。この2つの影響により uCTR では正確性が向上したと考えられる。また、評価値行列以外の情報である文書を用いることで、文書の種類（アイテム、ユーザ）によらず、評価値の欠損によるスパース性の影響が小さくロバストな推薦が行えることが分かった。

6.1.2 課題2 に対する結論

本研究では、課題2 “評価値行列が持つ特徴と文書自体が持つ特徴およびハイパーパラメータ間の関係性が推薦性能に与える影響の分析” への取り組みとして、スパース度と

利用する単語の種類が推薦性能に与える影響の分析（4章）を行った。筆者はスパース度に応じて有益なトピックを抽出する必要があると考え、文書頻度に基づいて利用する単語を選択することが重要だと仮定した。実験において、異なる評価値行列のスパース度の下で、文書頻度を基に利用する単語を選択し、また潜在因子ベクトルに影響するハイパーパラメータの組み合わせを変えることで、CTRの性能がどのように変化するかを分析した。この実験は、結果の一般性を検証するためにドメインの異なる3種類のデータセットを用意して行った。結果、スパース度と文書頻度の組み合わせに関しては、2つのデータセットで共通の傾向が存在することが分かった。スパース度が高い場合は大域的な情報が不足しており、様々な文書に出現する単語を含む文書から抽出したトピックは大域的な情報を持っていると考えられるため、文書頻度の高い単語が正確性と利便性の向上に非常に効果的である。スパース度が低い場合は大域的な情報は十分に存在しており、トピックが持つ大域的な情報がノイズとならないよう配慮が必要であるため文書頻度が高い単語は使用しない方が良く、文書頻度が中程度の単語は大域的な情報を持たないため正確性の向上する余地があるのだと考えられる。トピックの影響力を調整するハイパーパラメータに関しては、2つのデータセットで傾向が大きく異なり、事例毎に傾向の分析を行い調整する必要があると分かった。また、極端に人気のアイテムが存在するデータセットでは、MFが十分に機能する可能性が高く、文書情報による性能向上は見込めないことが分かった。これは人気アイテムを通してより多くのユーザ間に関係性が生まれるため、スパース度が高いことによる悪影響を受けなかったのだと考えられる。

6.1.3 課題3に対する結論

本研究では、課題3“文書中の語彙を制御することによる推薦性能に与える影響の分析”への取り組みとして、スパース度と利用する単語の種類が推薦性能に与える影響の分析（4章）と語彙補完を活用した複数分野に関心を持つユーザを対象とする推薦（5章）を行った。4章では、6.1.2節で述べたように、スパース度に応じて有益なトピックを抽

出す方法として、文書頻度に基づいて利用する単語を選択することが有効であることが分かった。また、5章では、推薦対象に応じて推薦性能を向上させる方法として、アイテムに関する文書とは異なる情報源を用いて文書の内容に合った単語の追加（語彙補完）を検討し分析を行った。筆者は推薦対象として、一部の分野にだけ深い興味を持つのではなく興味幅が広いユーザに着目した。文書に専門度のより低い単語を追加することで、その文書に近い分野の文書にも同じ単語が追加されて文書間に類似性が生じ、その文書（アイテム）に近い分野における人気が高いアイテムが推薦対象に加わると仮定した。実験において、論文や科学記事といった専門的な内容の文書に対して、web ページで学習した言語モデルが生成する文章を用いて語彙補完を行い、推薦対象を限定しない場合と興味幅が広いユーザに限定した場合で比較を行った。結果、専門的な内容の文書に対する web ページで学習した言語モデルによる語彙補完を行うことで、興味幅が広いユーザに対して推薦性能が向上することが分かった。

6.2 今後の展望

推薦システムの分野において、近年ではサイド情報が注目されている。サイド情報として利用される情報は、アイテムに関する文書情報やアイテムに対するレビュー情報、ユーザ間のネットワーク情報が多く、これらに比べてユーザに関する文書情報を利用した研究は多くない。サービスを提供している企業以外では、ユーザに関する文書の入手が難しいことが要因の一つだと推測する。本研究の分析により、ユーザに関する文書を利用した場合は正確性に優れた推薦が行える可能性を示した。そのため、ユーザに関する文書が取得できるのであれば検証対象に加えるべきであり、多数の検証事例が生じることで企業のサービスにおいてもユーザに関する文書の活用が広がるのではないかと筆者は考える。

ユーザの嗜好の時間的な変化に着目した研究も近年では多く見られる。ユーザの嗜好の変化をとらえるには、評価値の時系列順の履歴を利用するのが一般的である。SNSにお

けるプロフィール文には自分の社会的情報や趣味・興味が記述されていることが多いため、ユーザのプロフィール文の時系列的な変化が利用できれば、より明示的にユーザの嗜好の変化が分かるのではないかと考えられる。ただし、ユーザのプロフィール文の履歴はサービス提供側でなければ取得が困難である。代替方法としては、ユーザの発言履歴の利用が挙げられ、その当時にユーザが興味を持っていた内容を発言から抽出できればよい。トピック抽出を行うことでその発言で話題にしている分野を推定し、文章の感情極性がポジティブであればユーザの嗜好に合う興味である可能性が高い。また、発言量の多さと頻度によっては、プロフィール文より詳細にユーザの嗜好の変化をとらえることが可能だと考えられる。

本研究で行った実験はユーザの過去の履歴を利用したオフライン評価であり、ユーザの履歴中には存在しないが実際に提示されれば好む可能性のあるアイテムかを評価することができていない。推薦されたユーザにとって未知のアイテムがユーザの好みに合うかを評価すること、そして推薦されたアイテムをきっかけにユーザの新しい嗜好を開拓できたかを評価することが、より実用的な推薦システムの評価として必要だと考える。この評価を行うためには、被験者を募って彼らの履歴を元にその場で推薦を行って評価してもらう動的なオンライン評価を採用する必要がある。ただし、オンライン評価の被験者を個人で募る場合は大規模な評価は困難であり、被験者を募って作業を行ってもらう Web サービスを利用する場合は実験内容が制限されてしまうことや研究実施面でのコストが問題になる。E コマースや SNS などのサービスを提供している企業であれば、サービスの一部として推薦機能を実装することで多くのユーザを被験者候補にでき、なおかつユーザにインセンティブを与えることで評価に協力してもらえらるだろう。そのため、大規模な被験者でオンライン評価を行うのは企業こそが適していると考えられる。先述した、企業のサービスにおいてユーザに関する文書の活用についての検証が、オンライン評価で行われることが理想的である。ユーザの新しい嗜好を開拓することは、推薦システムの存在意義を高めるものであり、企業のサービス価値を高めることにも繋がる。

CTR をはじめとしたサイド情報として文書を利用するアプローチにおいて、取得したテ

キストに対する単語の追加や除去の処理は、ストップワードの除去や評価値行列の状況に関係なく文書頻度が一定以下の単語を除去するといった軽度な前処理にとどまっていた。これは、どのような手法や基準を用いて追加や削除すれば良いか明確でなかったからだと推測する。本研究で提案する、スパース度に応じて文書頻度を基準に単語を削除する手法と、推薦対象に応じて言語モデルを用いて単語を追加する手法は、このことに対する一つの回答となる。本研究で提案するこれらの手法を取り入れることで、文書情報を利用する研究や実用においてより一層の性能向上が実現できるのではないかと筆者は考える。また、文書の内容が推薦性能に与える影響を考慮すると、文書に対する処理として本研究では見出すことができなかった新たな着眼点での性能向上の余地は存在するだろう。思いつく例としては、深層学習で用いられる技術の一種である Attention は入力データで注目すべき箇所を学習する機構であるため、スパース度と文書から注目すべき単語を推定して文書中の単語選択を行うことが可能になれば、文書頻度の区間を具体的に設定する必要がなくなりルールベースから脱却できる。本研究が、今後の文書の内容が推薦性能に与える影響に関する研究への取っ掛かりとなることを期待する。

謝辞

博士課程での研究を進めるにあたって、ご指導いただいた土方嘉徳教授、佐藤宏介教授に厚く御礼申し上げます。土方嘉徳教授からは、研究計画、調査活動、論文執筆において親身にご指導およびご助言をいただきました。佐藤宏介教授からは、博士に必要な素養や博士論文としてまとめる際の学術的ポイントについてのご助言をいただき深く学ばせていただきました。本論文の審査を行ってくださった、副査の飯國洋二教授、長井隆行教授には、貴重なコメントをいただきましたことを感謝申し上げます。本研究に対して貴重なご指摘をいただきました、西田正吾教授、原田研介教授、酒田信親准教授、岩井大輔准教授に感謝申し上げます。また、事務手続きでお世話になった是國ゆう子氏、企業に勤めながら博士課程に進む機会を与えてくださった上司の大原秀一郎氏、角田大氏に感謝申し上げます。

これまでを振り返ると、土方先生には10年以上お世話になっております。最初にご指導いただいたのは学部2年生の電子システム学セミナーでデータマイニングに関する本の輪読を行う授業でした。当時の私はAIや機械学習に薄っすらとした興味を持っていましたが、今振り返ってみるとこの授業で興味の方向性が固まったように思います。西田研究室に入ってからは、研究の進め方や論文・プレゼンテーションの書き方など今に繋がる基礎をご指導くださりました。修士課程を終えて就職した後、修士で行った研究が論文誌に採録されたことをきっかけに博士課程へのお誘いをいただけたことも、私の人生にとって大きな転機だったと感じています。企業での研究開発の業務と博士課程の研究を並行して行うことは、予想していたより大変なものでした。思ったような研究成果が

出ない日々は辛く、そのような状況が2年ほど続いた間、土方先生からいただいた粘り強いご指導と励ましには感謝の念が尽きません。

最後になりましたが、遠方で一人暮らしをしている私の心身を心配して頻繁に連絡をくれた家族の存在は心の支えであり、たまの帰省では暖かく迎えてくれたことに深く感謝いたします。以上、これまでお世話になった皆様へ、心からの感謝を申し上げます。

参考文献

- 1) 神嶋 敏弘: 推薦システムのアルゴリズム (2), 人工知能学会誌, Vol.23, No.1, pp.89-103 (2008)
- 2) 神嶋 敏弘: 推薦システムのアルゴリズム (3), 人工知能学会誌, Vol.23, No.2, pp.248-263 (2008)
- 3) Burke, R.: Hybrid recommender systems: Survey and experiments, *Journal of User-Modeling and User-Adapted Interactions*, Vol.12, No.4, pp.331-370 (2002)
- 4) Deldjoo, Y., Schedl, M., Hidasi, B. and Knees, P.: Multimedia Recommender Systems, *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys)*, pp.537-538 (2018)
- 5) Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An open architecture for collaborative filtering of netnews, *Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW)*, pp.175-186 (1994)
- 6) Basu, C., Hirsh, H., Cohen, WW and Nevill-Manning, C.: Technical paper recommendation: A study in combining multiple information sources, *Journal of Artificial Intelligence Research*, Vol.14, pp.231-252 (2001)
- 7) Yu, K., Schwaighofer, A., Tresp, V., Ma, W. Y. and Zhang, H.: Collaborative Ensemble Learning: Combining Collaborative and Content-Based Information Filtering via Hierarchical Bayes, *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence (UAI)*, pp.616-623 (2003)

- 8) Su, X. and Khoshgoftaar, T.M.: A Survey of Collaborative Filtering Techniques, *Journal of Advances in Artificial Intelligence*, Vol.4 (2009)
- 9) Koren, Y., Bell, R. and Volinsky, C.: Matrix Factorization Techniques for Recommender Systems, *IEEE Computer* Vol.42, No.8, pp.30-37 (2009)
- 10) Koren, Y.: Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model, *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.426-434 (2008)
- 11) 土方 嘉徳: 推薦システムのオフライン評価手法, *人工知能学会誌*, Vol.29, No.6, pp.658-689 (2014)
- 12) Rashid, A., Karypis, G. and Riedl, J.: Learning preferences of new users in recommender systems: an information theoretic approach, *ACM SIGKDD Explorations Newsletter*, Vol.10, Issue.2, pp.90-100 (2008)
- 13) Sun, M., Li, F., Lee, J., Zhou, K., Lebanon, G. and Zha, H.: Learning multiple-question decision trees for cold-start recommendation, *Proceedings of the sixth ACM international conference on Web search and data mining (WSDM)*, pp.445-454 (2013)
- 14) Gope, J. and Jain, S.K.: A survey on solving cold start problem in recommender systems, *Proceedings of the International Conference on Computing, Communication and Automation (ICCCA)*, pp.133-138 (2017)
- 15) Paterek, A.: Improving Regularized Singular Value Decomposition for Collaborative Filtering, *Proceedings of KDD Cup and Workshop*, pp.39-42 (2007)
- 16) Wang, C. and Blei, D.M.: Collaborative Topic Modeling for Recommending Scientific Articles, *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.448-456 (2011)
- 17) McAuley, J. and Leskovec, J.: Hidden Factors and Hidden Topics : Understanding Rating Dimensions with Review Text, *Proceedings of the 7th ACM conference on Recommender systems (RecSys)*, pp.165-172 (2013)

- 18) Bao, Y., Fang, H., Zhang, J.: Topicmf: Simultaneously exploiting ratings and reviews for recommendation, Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI), pp.27-31 (2014)
- 19) Yu, Y., Wang, C., Wang, H. and Gao, Y.: Attributes coupling based matrix factorization for item recommendation, Journal of Applied Intelligence Vol.46, pp.521-533 (2017)
- 20) Ding, X., Jin, X., Li, Y. and Li, L.: Celebrity Recommendation with Collaborative Social Topic Regression, Proceedings of the Twenty-Third international joint conference on Artificial Intelligence (IJCAI), pp.2612-2618 (2013)
- 21) Purushotham, S., Liu, Y. and Kuo, C.C.J.: Collaborative Topic Regression with Social Matrix Factorization for Recommender System, Proceedings of the 29th International Conference on International Conference on Machine Learning (ICML), pp.691-698 (2012)
- 22) Kang, J. H. and Lerman, K.: LA-CTR: A Limited Attention Collaborative Topic Regression for Social Media, Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI), pp.1128-1135 (2013)
- 23) Wang, H., Chen, B. and Li, W.J.: Collaborative Topic Regression with Social Regularization for Tag Recommendation, Proceedings of the Twenty-Third international joint conference on Artificial Intelligence (IJCAI), pp.2719-2725 (2013)
- 24) Chen, C., Zheng, X., Wang, Y., Hong, F. and Lin, Z.: Context-Aware Collaborative Topic Regression with Social Matrix Factorization for Recommender Systems, Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI), pp.9-15 (2014)
- 25) Wang, H. and Li, W.J.: Relational collaborative topic regression for recommender systems, Journal of IEEE Transactions on Knowledge and Data Engineering, Vol.27, pp.1343-1355 (2015)
- 26) Li, J., Chen, C., Chen, H. and Tong, C.: Towards Context-aware Social Recommen-

- dation via Individual Trust, *Journal of Knowledge-Based Systems*, Vol.127, pp.58-66 (2017)
- 27) Wang, H., Wang, N. and Yeung, D.-Y.: Collaborative Deep Learning for Recommender Systems, *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1235-1244 (2015)
- 28) Bansal, T., Belanger, D. and McCallum, A.: Ask the GRU: Multi-task Learning for Deep Text Recommendations, *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)*, pp.107-114 (2016)
- 29) Jin, M., Luo, X., Zhu, H. and Zhuo, H.: Combining Deep Learning and Topic Modeling for Review Understanding in Context-Aware Recommendation, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Vol.1, pp.1605-1614 (2018)
- 30) Wu, L., Quan, C., Li, C., Wang, Q., Zheng, B. and Luo, X.: A Context-Aware User-Item Representation Learning for Item Recommendation, *Journal of ACM Transactions on Information Systems*, Vol.37, No.2 (2019)
- 31) Chen, W.H., Hsu, C.C., Lai, Y.A., Liu, V., Yeh, M.Y and Lin, S.D.: Attribute-Aware Recommender System Based on Collaborative Filtering: Survey and Classification, *Journal of Frontiers in Big Data*, 10.3389/fdata.2019.00049 (2019)
- 32) Zhen, Y., Li, W.J. and Yeung, D.Y.: Tagicofi: Tag Informed Collaborative Filtering, *Proceedings of the third ACM conference on Recommender systems (RecSys)*, pp.69-76 (2009)
- 33) Bauman, K., Liu, B. and Tuzhilin, A.: Aspect Based Recommendations: Recommending Items with the Most Valuable Aspects Based on User Reviews, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.717-725 (2017)
- 34) Singh, A.P. and Gordon, J.G.: Relational Learning via Collective Matrix Factorization,

- Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.650-658 (2008)
- 35) Singh, A.P. and Gordon, J.G.: A Bayesian Matrix Factorization Model for Relational Data, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI), pp.556-563 (2010)
 - 36) Saveski, M. and Mantrach, A.: Item Cold-start Recommendations: Learning Local Collective Embeddings, Proceedings of the 8th ACM Conference on Recommender systems (Recsys), pp.89-96 (2014)
 - 37) Sedhain, S., Menon, A.K., Sanner, S., Xie, L. and Brazhunas, D.: Low-Rank Linear Cold-Start Recommendation from Social Data, Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI), pp.1502-1508 (2017)
 - 38) Cortes, D.: Cold-start Recommendations in Collective Matrix Factorization, arXiv preprint, abs/1809.00366 (2018)
 - 39) Balraj, K., Neeraj, S. and Shubhangi, s.: Collaborative Topic Regression-Based Recommendation Systems: A Comparative Study, Proceedings of ICRIC 2019, vol.597, pp.723-737 (2019)
 - 40) Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet Allocation, Journal of Machine Learning Research, Vol.3, pp.993-1022 (2003)
 - 41) Blei, D.M.: Probabilistic Topic Models, IEEE Signal Processing Magazine, vol.27, pp.55-65 (2010)
 - 42) 佐藤 一誠, 奥村 学: トピックモデルによる統計的潜在意味解析, 自然言語処理シリーズ 8, コロナ社 (2015)
 - 43) Hofmann, T.: Probabilistic latent semantic indexing, Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR), pp.50-57 (1999)
 - 44) Salakhutdinov, R. and Mnih, A.: Probabilistic Matrix Factorization, Proceedings of the

- 20th International Conference on Neural Information Processing Systems (NeurIPS), pp.1257-1264 (2007)
- 45) Hu, Y., Koren, Y. and Volinsky, C.: Collaborative Filtering for Implicit Feedback Datasets, Proceedings of the Eighth IEEE International Conference on Data Mining (ICDM), pp.263-272 (2008)
- 46) Wan, M. and McAuley, J.: Item Recommendation on Monotonic Behavior Chains, Proceedings of the 12th ACM Conference on Recommender Systems (RecSys), pp.86-94 (2018)
- 47) Jey, L., Karl, G., David, N. and Timothy, B.: Automatic Labelling of Topic Models, Proceedings of the 49th Annual Meeting of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp.1536-1545 (2011)
- 48) Seunghyun, P., Jichul, J. and Chihyung, O.: Analyzing Twitter to explore perceptions of Asian restaurants, Journal of Hospitality and Tourism Technology, Vol.7, pp.405-422 (2016)
- 49) Leung, X., Bai, B., and Stahura, K.: THE MARKETING EFFECTIVENESS OF SOCIAL MEDIA IN THE HOTEL INDUSTRY: A COMPARISON OF FACEBOOK AND TWITTER, Journal of Hospitality and Tourism Research, Vol.39 (2015)
- 50) Taecharungroj, V.: Starbucks' marketing communications strategy on Twitter, Journal of Marketing Communications, Vol.23, pp.1-19 (2016)
- 51) 山本 祐輔, 浅井 洋樹, 上田 高德, 秋岡 明香, 山名 早人: テレビ番組に対する意見をもつ Twitter ユーザのリアルタイム検出, データ工学と情報マネジメントに関するフォーラム (2013)
- 52) Meng, X., Wei, F., Liu, X. and Zhou, M.: Entity-centric topic-oriented opinion summarization in twitter, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.379-387 (2012)

- 53) Park, S.J., Lim, Y.S., Sams, S., Sang, M.N. and Park, H.W.: Networked Politics on Cyworld: The Text and Sentiment of Korean Political Profiles, *The Journal of Social Science Computer Review*, Vol.29, No.3, pp.288-299 (2011)
- 54) 池田 和史, 服部 元, 松本 一則, 小野 智弘, 東野 輝夫: マーケット分析のための Twitter 投稿者プロフィール推定手法, *情報処理学会論文誌コンシューマ・デバイス&システム*, Vol.2, No.1, pp.82-93 (2012)
- 55) Wang, Z., Hale, S., Adelani, D.I., Grabowicz, P., Hartmann, T., Flöck, F. and Jurgens, D.: Demographic Inference and Representative Population Estimates from Multilingual Social Media Data, *Proceedings of the World Wide Web Conference (WWW)*, pp.2056-2067 (2019)
- 56) Guy, I., Jacovi, M., Perer, A., Ronen, I. and Uziel, E.: Same Places, Same Things, Same People? Mining User Similarity on Social Media, *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW)*, pp.41-50 (2010)
- 57) Koga, H. and Taniguchi, T.: Developing a user recommendation engine on twitter using estimated latent topics, *Proceedings of the 14th international conference on Human-computer interaction: design and development approaches (HCII)*, pp.461-470 (2011)
- 58) 藤本 拓, 秋永 和計, 柴藤 稔: 潜在トピックモデルを利用したユーザプロファイリング技術, *NTT DoCoMo テクニカル・ジャーナル*, Vol.19, No.3, pp.37-41 (2011)
- 59) 松尾 豊, 友部 博教, 橋田 浩一, 中島 秀之, 石塚 満: Web 上の情報から人間関係ネットワークの抽出, *人工知能学会論文誌*, Vol.20, No.1, pp.46-56, (2005)
- 60) 芹澤 翠, 小林 一郎: 潜在的ディリクレ配分法に基づくトピック類似度を考慮したトピック追跡, *データ工学と情報マネジメントに関するフォーラム (DEIM)* (2011)
- 61) Griffiths, T.L. and Steyvers, M.: Finding Scientific Topics, *Proceedings of the National Academy of Sciences*, Vol.101, pp.5228-5235 (2004)
- 62) Bertsekas, D.P.: *Nonlinear Programming Second Edition*, Athena Scientific, Belmont, Massachusetts (1999)

- 63) Ge, M., Delgado-Battenfeld, C. and Jannach, D.: Beyond accuracy: Evaluating Recommender Systems by Coverage and Serendipity, Proceedings of the fourth ACM conference on Recommender systems (RecSys), pp.257-260 (2010)
- 64) Adomavicius, G. and Kwon, Y.: Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques, Journal of IEEE Transactions on Knowledge and Data Engineering, Vol.24, No.5, pp.896-911 (2012)
- 65) Zhou, T., Kuscsik, Z., Liu, J.G., Medo, M., Wakeling, J.R. and Zhang, Y.C.: Solving the Apparent Diversity-accuracy Dilemma of Recommender Systems, Proceedings of the National Academy of Sciences, Vol.107, No.10, pp.4511-4515 (2010)
- 66) Lathia, N., Hailes, S., Capra, L. and Amatriain, X.: Temporal Diversity in Recommender Systems, Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR), pp.210–217 (2010)
- 67) Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N. and others: The Pile: An 800GB Dataset of Diverse Text for Language Modeling, arXiv preprint arXiv:2101.00027 (2020)
- 68) Liji, U., Chai, Y. and Chen, J.: Improved personalized recommendation based on user attributes clustering and score matrix filling, Journal of Computer Standards and Interfaces, Vol.57, pp.59-67 (2018)
- 69) Levandoski, J., Sarwat, M., Eldawy, A. and Mokbel, M.: LARS: A location-aware recommender system, Proceedings of International Conference on Data Engineering, pp.450-461 (2012)
- 70) Shivaram, K., Liu, P., Shapiro, M., Bilgic, M. and Culotta, A.: Reducing Cross-Topic Political Homogenization in Content-Based News Recommendation, Proceedings of the 16th ACM Conference on Recommender Systems (RecSys), pp.220-228 (2022)
- 71) Devlin, J., Chang, M.W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Proceedings of the 2019 Confer-

- ence of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp.4171-4186 (2019)
- 72) Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. and Le, Q.V.: XLNet: Generalized Autoregressive Pretraining for Language Understanding, Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS), pp.5753-5763 (2019)
- 73) Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D.: Language Models are Few-Shot Learners, Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS), pp.1877-1901 (2020)
- 74) Gustavo, P. and Claudia, H.: What Does BERT Know about Books, Movies and Music? Probing BERT for Conversational Recommendation, Proceedings of the 14th ACM Conference on Recommender Systems (RecSys), pp.388-397 (2020)
- 75) Damien, S., Wout, V. and Robbe, R.: Zero-Shot Recommendation as Language Modeling, Proceedings of Advances in Information Retrieval: 44th European Conference on IR Research, pp.223-230 (2022)
- 76) Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, Proceedings of the 31st International Conference on International Conference on Machine Learning (PMLR), pp.1188-1196 (2014)
- 77) echoes: データからみる Twitter ユーザー実態まとめ
(<https://service.aainc.co.jp/product/echoes/voices/0014>), (参照 2022-09-07)
- 78) MeCab: Japanese morphological analyzer, 入手先 (<http://taku910.github.io/mecab/>), (参照 2022-11-11)

-
- 79) TermExtract: 専門用語（キーワード）自動抽出用 Perl モジュール, 入手先
〈<http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>〉, (参照 2022-11-11)
- 80) TreeTagger: a language independent part-of-speech tagger,
〈<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>〉, (参照 2022-11-11)

業績

学術雑誌論文

1. 西村章宏, 土方嘉徳, 三輪祥太郎, 西田正吾: 一般ユーザの観点に基づく Twitter から的人物関係の可視化と事例の考察, 情報処理学会論文誌, Vol.56, No.3, pp.972-982 (2015)
2. 西村章宏, 土方嘉徳, 酒田信親: 潜在トピックを利用した協調フィルタリングにおけるトピック情報源の違いに関する調査, 情報処理学会論文誌, Vol.58, No.5, pp.1102-1112 (2017)
3. 西村章宏, 土方嘉徳, 佐藤宏介: 潜在トピックを利用した協調フィルタリングにおけるスパース度合いに応じた性能評価, 電気学会論文誌 C, Vol.142, No.9, pp.1048-1059 (2022)

学術雑誌論文 (投稿中)

1. 西村章宏, 土方嘉徳, 佐藤宏介: 潜在トピックを用いた協調フィルタリングに対する語彙補完, 情報処理学会論文誌

国際会議 (査読あり)

1. Akihiro Nishimura, Yoshinori Hijikata, Kosuke Sato: Recommender Performance on Collaborative Filtering using Latent Topics under Several Sparsity Levels, Proceedings of the 7th Workshop on Complex Methods for Data and Web Mining (CMDWM'20), Melbourne, Australia (2020)
2. Akihiro Nishimura, Yoshinori Hijikata and Kosuke Sato: Recommender Performance for Users with Interest in Several Fields on Collaborative Topic Regression through Vocabulary Completion, Proceedings of the 17th International Conference on Ubiquitous Information Management and Communication (IMCOM 2023), Suwon, South Korea (2023)

国内会議 (査読あり)

1. 西村章宏, 土方嘉徳：潜在トピックを利用した協調フィルタリングにおけるトピック情報源の違いに関する分析, 情報処理学会, 第 8 回 Web とデータベースに関するフォーラム (WebDB Forum 2015), pp.24-31 (2015)

国内会議 (査読なし)

1. 西村章宏, 土方嘉徳, 三輪祥太郎, 西田正吾：一般ユーザの観点に基づく Twitter からの人物関係の可視化, 第 159 回 DBS・第 115 回 IFAT 合同研究発表会 (2014)
2. 西村章宏, 土方嘉徳, 三輪祥太郎, 西田正吾：一般ユーザ目線での Twitter からの人物関係の可視化, ARG 第 5 回 Web インテリジェンスとインタラクション研究会 (ARG WI2 研究会), pp.89-94 (2014)

受賞

1. 学生奨励賞: 西村章宏, 土方嘉徳, 三輪祥太郎, 西田正吾: 一般ユーザの観点に基づく Twitter からの人物関係の可視化, 第 159 回 DBS・第 115 回 IFAT 合同研究発表会 (2014)