

# Twitter の会話ログを利用した複数ユーザに対する話題推薦

蔵内 雄貴<sup>†</sup> 倉島 健<sup>†</sup> 岩田 具治<sup>††</sup> 星出 高秀<sup>†</sup> 藤村 考<sup>†</sup>

<sup>†</sup> 日本電信電話株式会社 NTT サイバーソリューション研究所 〒239-0847 神奈川県横須賀市光の丘 1-1

<sup>††</sup> 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所 〒619-0237 京都府相楽郡精華町光台 2-4

E-mail: †{kurauchi.yuki,kurashima.takeshi,iwata.tomoharu,hoshide.takahide,fujimura.ko}@lab.ntt.co.jp

あらまし 本稿では、ソーシャルメディアにおける人間関係の強化、拡大をより容易にすることを目指し、「2 ユーザの会話が盛り上がる話題」の推薦を行う。「2 ユーザの会話が盛り上がる話題」に対する要件を、1) 2 ユーザの嗜好に共通し、2) 一部のユーザに限られ、3) 議論や雑談を引き出す、4) ニュースや画像などの具体的な情報、という 4 項目と定義した。我々は、Twitter の大規模会話ログとしての価値に着目し、これを話題の自動推薦に活用する手法を提案する。提案手法は 2 ユーザを入力とし、共通の嗜好を持つ一部のユーザ集合をそれぞれ抽出する。そして、そのユーザ集合間において、より多くのユーザがより長く会話している URL やハッシュタグを推薦する。Twitter データを用いた評価実験では、既存手法と比べ提案手法が、要件を満たす話題、すなわち、「2 ユーザの会話が盛り上がる話題」による会話をより正確に予測できることを示した。

キーワード 話題推薦, 会話ログ, グループレコメンデーション

## Recommending Conversation Topics Using Twitter

Yuki KURAUCHI<sup>†</sup>, Takeshi KURASHIMA<sup>†</sup>, Tomoharu IWATA<sup>††</sup>,

Takahide HOSHIDE<sup>†</sup>, and Ko FUJIMURA<sup>†</sup>

<sup>†</sup> NTT Cyber Solutions Laboratories, NTT Corporation

1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan

<sup>††</sup> NTT Communication Science Laboratories, NTT Corporation

2-4 HIKARIDAI, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan

E-mail: †{kurauchi.yuki,kurashima.takeshi,iwata.tomoharu,hoshide.takahide,fujimura.ko}@lab.ntt.co.jp

**Abstract** We investigate the novel problem of recommending conversation topics to liven up the dialog among users, so that they can easily create new and better relationships in social media. We require the system's output to be 1) common to both input users' interests, 2) specific to those interests, 3) appropriate for discussions or chats, and 4) concrete items, such as news and images. We focus on conversation logs on Twitter, and propose a method of recommending conversation topics using Twitter. Our method finds similar-interest groups for each of the input users, and recommends URLs and hashtags which have the most replies from the largest number of users in those groups. We evaluate our method using Twitter data, and conclude that our method can predict conversations about topics which meet the requirements.

**Key words** conversation topic recommendation, conversation log, group recommendation

### 1. はじめに

近年, Twitter<sup>(注1)</sup>, Facebook<sup>(注2)</sup>, mixi<sup>(注3)</sup> などの, ユーザ

同士の会話インタラクションにより, 人と人のつながりを楽しむソーシャルメディアの普及が著しい [15]. このような知人との関係は, 情報伝達の効率化や感情面のサポートなどにおいて有益であり, 社会関係資本 (social capital) としてとらえられ注目されている [2] [4]. しかし, ソーシャルメディアがより多くの知人とのつながりを保てる可能性を提供している一方で, 今ある全ての人間関係を維持し, さらに強化, 拡大するために

(注1): <http://twitter.com>

(注2): <http://www.facebook.com>

(注3): <http://mixi.jp>

は、多くの労力を費やす必要がある [7] [8] .

本稿では、ソーシャルメディアにおける人間関係の強化、拡大をより容易にすることを目指し、「2 ユーザの会話が盛り上がる話題」の推薦を行う。我々は、「2 ユーザの会話が盛り上がる話題」に関する仮説を立て、推薦する話題が満たすべき要件を、1) 2 ユーザの嗜好に共通し、2) 一部のユーザに限られ、3) 議論や雑談を引き出す、4) ニュースや画像などの具体的な情報、という 4 項目と定義した。

話題の自動推薦を実現するため、我々は Twitter の大規模会話ログとしての価値に着目した。本稿では、大規模会話ログデータをユーザ間会話の学習データとして用いることで、話題の自動推薦を実現する手法を提案する。会話ログとは、どのようなユーザとどのようなユーザが、どのような話題でどの程度の長さ会話したかを示すログである。電話やメールといったクローズドな大規模会話ログは過去にも存在したが、オープンな大規模会話ログは、Twitter の登場によって初めて利用可能となったもので、極めて利用価値が高い。また、我々がクローリングした Twitter データにおいて、ユーザ名を含むツイートは 41% を占めることから、Twitter における会話ログの量は非常に多いと言える。

提案手法は、前述した 4 つの要件 (共通性、ローカル性、会話らしさ、具体性) を満たす話題を推薦するために、以下の方法をとる。まず、2 ユーザを入力とし、共通の嗜好を持つ一部のユーザ集合をそれぞれ抽出する。そして、そのユーザ集合間において、より多くのユーザがより長く会話している URL やハッシュタグを推薦する。この手法は、2 ユーザが過去に発信した情報から嗜好を推定し、この嗜好の共通性に基づく推薦を行うことで、共通性を満たす。また、ユーザを嗜好に基づいてクラスタリングし、各嗜好を持つ一部のユーザ集合が会話した話題を推薦することで、ローカル性を満たす。また、会話ログを利用しより多くのユーザがより長く会話している話題を推薦することで、会話らしさを満たす。最後に、URL やハッシュタグを推薦することで、具体性を満たす。

本研究には、会話のサポートという側面と、複数ユーザに対する推薦という側面がある。会話のサポートを行う研究としては、現在の会話内容に関連した画像を提示する研究 [16] や、ソーシャルネットワークを密にするために話すべきユーザを推薦する研究 [12] などがある。複数ユーザに対する推薦を行う研究としては、個人のログや推薦リストを組み合わせる推薦を行う研究 [1] [5] [11] などがある。しかし、話題推薦タスクにおいて、大規模な会話ログデータに基づき統計的にアプローチした研究は過去に存在しなかった。

提案手法が、要件を満たす話題を自動で探し出せるか検証するため、Twitter データを用いた実験を行った。具体的には、2011 年 1 月 1 日から 1 月 15 日において、Twitter ユーザ同士が実際に会話した話題の予測タスクを用いた。話題が満たすべき要件に対応する評価指標で評価した結果、既存手法と比べ提案手法が、要件を満たす話題による会話をより正確に予測できることを示す。

本稿の構成は以下の通りである。2 章で定義について述べ、3

章で要件を満たす話題を推薦するための提案手法について詳述する。4 章で提案手法の有効性を検証するための実験について述べ、5 章で実験結果に対する考察を述べる。6 章で関連研究、7 章で本稿のまとめについて述べる。

## 2. 定 義

本章では、本稿で使用する用語、取り組む課題、話題と手法が満たすべき要件を定義する。

### 2.1 用 語

ツイート: Twitter における投稿。

リプライ: 指定のユーザに向けた返信投稿。Twitter 公式の `in_reply_to_status_id` が付与されたものに限るとする。

投稿ログ: あるユーザの過去のツイートの集合。そのユーザがどのような嗜好を持っているかの解析に利用する。

会話ログ: あるユーザの過去のリプライおよび、それらの起点となったツイートの集合。そのユーザがどのような人と、どのような話題で会話するかの解析に利用する。

### 2.2 課 題

本研究で取り組む技術課題は「ユーザ  $u$  とユーザ  $u'$  が与えられた場合に、2 ユーザに対して話題  $i$  を推薦する」ことである。このとき、ユーザ  $u$  およびユーザ  $u'$  に関する情報として、投稿ログ  $L_u$  および  $L_{u'}$  が与えられるものとする。

### 2.3 話題の要件

ユーザの組み合わせに応じ、2 ユーザに特化した話題の自動推薦を実現するため、まず、Twitter 上に存在する大量の会話ログをユーザ間会話の学習データとして用いる。そして、会話ログから適した話題を抽出し推薦する、という手法をとる。しかし、会話した人数や会話した回数に基づいてシンプルに話題を推薦すると、世間では多く会話されているが、2 ユーザにとってなじみのない話題や、天気のような一般的な話題が推薦され、2 ユーザの会話が盛り上がらないという問題がある。

そこで、(1) 共通の話題である方がお互いが話しやすい、(2) その共通点が、一部のユーザに限られたものである方が、お互いの親近感につながる、(3) 議論や雑談を引き出すような会話らしい話題であるほど、会話が続きやすい、(4) キーワードのような抽象的な情報ではなく、ニュース、投稿、画像、映像など、具体的な情報を推薦した方が、会話のきっかけとなりやすい、という 4 つの仮説に基づき、推薦する話題が満たすべき以下の 4 つの要件を導き出した。

要件 1: 共通な話題

要件 2: ローカルな話題

要件 3: 会話らしい話題

要件 4: 具体的な話題

### 2.4 手法の要件

人間関係の強化だけでなく、人間関係の拡大にも利用できるように、Twitter 上での過去の会話数によらない手法とする。つまり、過去に 2 ユーザが会話をしていなくとも、話題を推薦できるような手法とする。また、実際にサービス化した際に、大量のユーザのリクエストに対し遅延なく応答するため、計算量が少ない必要がある。以上から、手法が満たすべき以下の 2

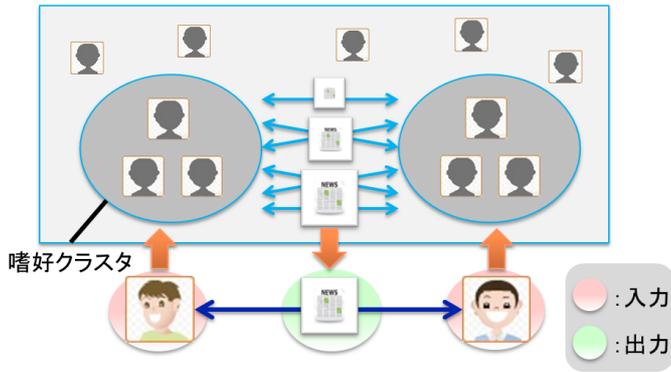


図 1 話題推薦の基本アイデア

つの要件を導き出した。

- 要件 5: Twitter 上で過去に会話が少ない/したことがない  
ユーザ同士に対しても話題を推薦できる
- 要件 6: 計算量が少ない

### 3. 提案手法

会話ログを学習データとし、話題と手法の要件を満たし 2.2 章の課題を解決する手法について述べる。

#### 3.1 手法の概要

提案手法の基本アイデアは、それぞれの入力ユーザと似たユーザ同士の話題を推薦することである。図 1 のように、提案手法は 2 ユーザを入力とし、共通の嗜好を持つ一部のユーザ集合をそれぞれ抽出する。そして、そのユーザ集合間において、より多くのユーザがより長く会話している URL やハッシュタグを推薦する (3.2 章)。このとき、入力された 2 ユーザが共通の嗜好を持つ場合は、そのユーザ集合内で会話されている話題が推薦される。ユーザ集合の抽出は、投稿ログに記述された内容の近さによってユーザ間の類似性を評価することで実現する。なお、ユーザを嗜好によってクラスタリングし、ユーザ集合を生成する処理 (3.3 章) と、各ユーザ集合内またはユーザ集合間において、どの話題がより多くより長く会話されているかを解析する処理 (3.4 章) は、学習フェーズとして事前に行う。

提案手法は、要件を満たす話題を推薦するために考案したものであり、この手法を用いる理由を、話題の要件に沿って順に述べる。2 ユーザに共通な話題の推薦を実現するため、2 ユーザが過去に発信した情報、すなわち投稿ログから嗜好を推定し、この嗜好の共通性に基づく推薦を行う (話題の要件 1)。別の手法として、2 ユーザが過去に会話した内容、すなわち会話ログをもとに話題を推薦する手法がある。しかし、この場合、推薦される 2 ユーザの間に十分な量の会話ログがある必要がある。また、会話ログは Twitter 上に大量に存在するものの、2 ユーザに絞ると極めてスパースであるという特徴をもつ。そこで、2 ユーザの会話ログの有無に依存せず、会話ログのスパースさを吸収する手法をとった (手法の要件 5)。

ローカルな話題を推薦するため、ユーザを嗜好に基づいてクラスタリングし、各嗜好を持つ一部のユーザ集合内で話された話題を優先的に推薦する。推薦される話題は、Twitter 全体で

話題になっているものではなく、ある特定の嗜好を持つ一部のユーザ集合内またはユーザ集合間で話題になっているものである (話題の要件 2)。また、あるユーザが入力された時に、類似したユーザを探すのではなく、類似したユーザ集合を探すだけで良いため、ユーザベースの協調フィルタリングと比べて推論フェーズの計算コストも低いメリットがある (手法の要件 6)。

会話らしい話題を推薦するため、より多くのユーザがより長く会話している話題を推薦する。これは、会話らしい話題すなわち議論や雑談を引き出す話題は、会話ログにおいて、より多くのユーザによってより長く議論や雑談がなされているという考えに基づいている (話題の要件 3)。

具体的な情報を推薦するため、ニュース、投稿、画像、映像などを 1 つのアイテムとして扱える URL やハッシュタグを推薦する。また、ニュースと画像の類似度など、異なるメディア間の類似度計算は困難なため、コンテンツフィルタリングは用いなかった (話題の要件 4)。

#### 3.2 話題の推薦

$P(i|u, u', t)$  は、入力されるユーザ  $u$  と  $u'$  が話題  $i$  を期間  $t$  において話題にする確率であり、以下のように求め、これが高い順に推薦を行う。

$$P(i|u, u', t) = \sum_{z_n \in Z} \sum_{z_m \in Z} P(z_n|u)P(z_m|u')P(i|z_n, z_m, t) \quad (1)$$

ここで、 $P(z_n|u)$  はユーザ  $u$  がクラスター  $z_n$  に属する確率、 $P(z_m|u')$  はユーザ  $u'$  がクラスター  $z_m$  に属する確率である。3.3 章でこれらの算出方法を述べる。 $P(i|z_n, z_m, t)$  は期間  $t$  においてクラスター  $z_n$  と  $z_m$  の間で話題  $i$  による会話が発生する確率である。3.4 章でこの算出方法を述べる。 $Z$  は、全てのクラスター集合である。

#### 3.3 嗜好によるユーザクラスタリング

ここでは、各ユーザが各クラスターに属する確率  $P(z_n|u)$  および  $P(z_m|u')$  を算出する。

話題の要件 1 を満たすため、ユーザを嗜好によってクラスタリングする。その際、人は複数の嗜好を持つため、1 人が 1 つのクラスターに属するハードクラスタリング手法ではなく、1 人が各クラスターに属する確率を算出するソフトクラスタリング手法を用いる。また、新たに現れる嗜好の対象にも対応できるように、特定のクラスターを与えるのではなく、投稿ログを用いてユーザをクラスタリングすることにより、ボトムアップにクラスターを抽出する。

以上から、PLSA を利用し、あるユーザがある潜在クラスターに属する確率を算出する手法を用いる。PLSA の式は以下の通りであり、右辺中の  $P(z|u)$  を推定し利用する。

$$P(w|u) = \sum_{z \in Z} P(z|u)P(w|z) \quad (2)$$

ここで、 $w$  は名詞と URL とハッシュタグを、 $z$  は潜在クラスターを、 $Z$  は潜在クラスター集合を表す。 $P(w|u)$  は、ユーザ  $u$  が  $w$  を選択する確率、 $P(w|z)$  は、潜在クラスター  $z$  から  $w$  が選択される確率である。

ユーザの嗜好によってクラスタリングを行うため、 $w$ として、投稿ログ中で出現する名詞、URL、ハッシュタグを利用する。3つを組み合わせるのは、Twitterは短文かつ口語的な表現が多く、名詞がうまく抽出できない問題に対応するためである。また、ユーザがURLやハッシュタグを利用しない場合にも嗜好を推定できるようにするためでもある。

### 3.4 クラスタ間で各話題が発生する確率の算出

ここでは、クラスタ間で各話題が発生する確率  $P(i|z_n, z_m, t)$  を以下の式で算出する。これは、直感的には、話題  $i$  について会話する人が多いほど、話題  $i$  についてリプライが長く続くほど大きな値をとるものである。

$$P(i|z_n, z_m, t) = \frac{C(z_m, z_n, i, t)}{\sum_{i' \in I} C(z_m, z_n, i', t)} \quad (3)$$

ここで、 $I$  は話題の集合である。 $C(z_n, z_m, i, t)$  は、クラスタ  $z_n$  と  $z_m$  における話題  $i$  の発生数を表し、以下の式で求める。

$$C(z_n, z_m, i, t) = \sum_{v \in U} \sum_{v' \in U} P(v|z_n) P(v'|z_m) N(v, v', i, t) \quad (4)$$

ここで、 $P(v|z_n)$  はクラスタ  $z_n$  からユーザ  $v$  が選択される確率であり、 $U$  は全ユーザの集合である。 $N(v, v', i, t)$  は、期間  $t$  におけるユーザ  $v$  と  $v'$  間の話題  $i$  についての会話回数であり、ユーザ  $v$  と  $v'$  の間で話題  $i$  についてリプライが続いた数と定義する。また、 $t$  は時間窓の長さであり、過去1週間、過去1ヶ月間など、どの程度の長さの期間で出現した話題を推薦対象とするかを表す。

## 4. 実験

### 4.1 概要

提案手法が、要件を満たす話題を自動で探し出せるか検証するため、Twitterユーザ同士が実際に会話した話題を予測できるかというタスクで提案手法を評価した。話題の中でも、特に要件を満たす話題が予測できるかを考慮した評価指標を用いた。

なお、2.3章にて話題の要件4として具体的な話題であることを述べた。URLを推薦することが望ましいが、収集したデータが比較的小規模であり、URLについて学習することが難しかった。そのため、今回の実験においては、推薦する話題としてハッシュタグを用いている。

### 4.2 クロールデータ

会話が密に発生しているデータを収集するよう、以下のステップでクロールを行った。投稿ログの収集には、Twitter statuses/user\_timeline API を利用した。このAPIは、指定したユーザの過去ツイートを、最大3,200件まで取得できる。

- (1) 2011年6月21日以前の@NTTPRの投稿ログを収集
- (2) @NTTPRの投稿ログの投稿内容に含まれるユーザ名(以下、1hopユーザとする)を抽出
- (3) 抽出した全ての1hopユーザの投稿ログを収集
- (4) 抽出した全ての1hopユーザについて以下を繰り返し
  - (4-1) 選んだユーザの投稿ログの投稿内容に含まれるユーザ名(以下、2hopユーザとする)を抽出

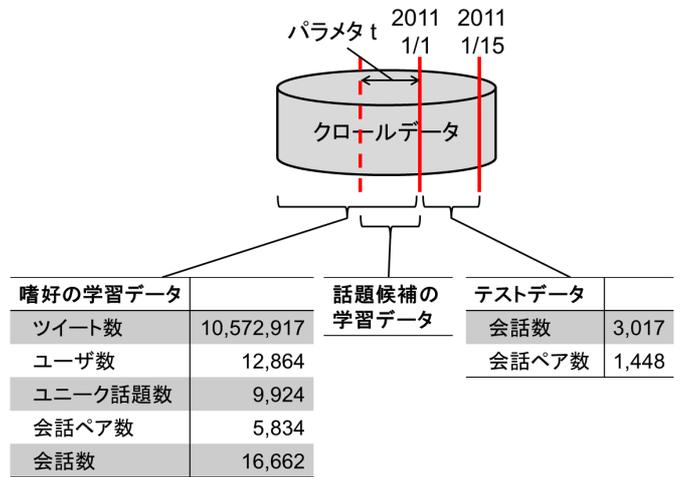


図2 データの詳細

### (4-2) 抽出した全ての2hopユーザの投稿ログを収集

結果として、28,828人のユーザのツイートをクロールした。ただし時間の制約上、@NTTPRの投稿ログの投稿内容に含まれる1hopユーザは769人であるが、実際にステップ(4-1)と(4-2)を繰り返したのはその中の180人であり、2hopユーザを全てクロールできたわけではない。

### 4.3 学習データとテストデータ

利用したデータの詳細を図2に示す。3.3章で述べたクラスタリングには2010年12月31日以前の全てのデータを利用し、3.4章で述べた話題候補の学習には2010年12月31日までのパラメタ  $t$  の期間のデータを利用した。テストデータには、2011年1月1日から2011年1月15日のデータを利用した。ただし、学習データに出現せず、テスト期間において初出となるデータは除外した。また、過去にそのユーザが会話したことがある話題についてはテストデータから除外した。

### 4.4 比較手法

提案手法は、共通かつローカルな話題を推薦するためにユーザの嗜好を考慮するという特徴と、会話らしい話題を推薦するために会話ログを活用するという特徴を持つ。そこで、嗜好を考慮するか、会話ログを活用するかが異なる以下の3つの手法を比較手法として用いた。それぞれの手法の特徴を、表1にまとめる。

出現頻度に基づく推薦: 学習データ中の全ユーザの投稿ログに含まれる話題のうち、その出現頻度が高い順に推薦する手法(以降、頻度ランキング)。

会話回数に基づく推薦: 学習データ中の全ユーザの会話ログに含まれる話題のうち、3.4章において求めた  $P(i|z_n, z_m, t)$  のクラスタ数  $|G|$  を1としたもの、すなわち、各話題の総返信回数が多い順に推薦する手法(以降、話題ランキング)。

2ユーザの嗜好の組合せ: ユーザ  $u, u'$  における話題  $i$  の関連度  $rel(u, u', i)$  を以下の式のように定義し、これが高い順に推薦する[1]で提案されている手法(以降、コンビネーション)。

$$rel(u, u', i) = \frac{1}{2}(P(i|u) + P(i|u')) \quad (5)$$

ここで、 $P(i|u)$  はユーザ  $u$  が話題  $i$  に興味を持つ確率であり、

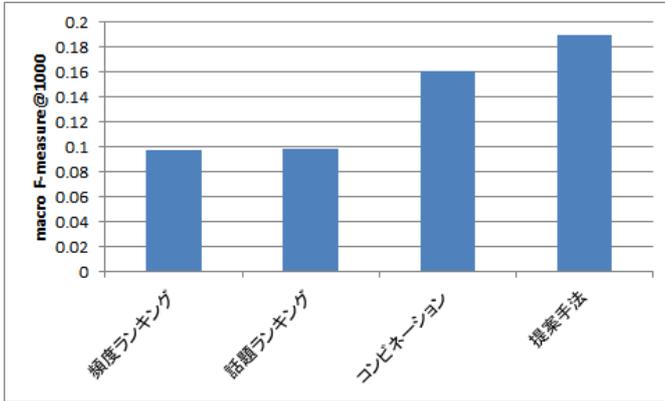


図 3 既存手法との比較 (Macro F-measure@1,000)

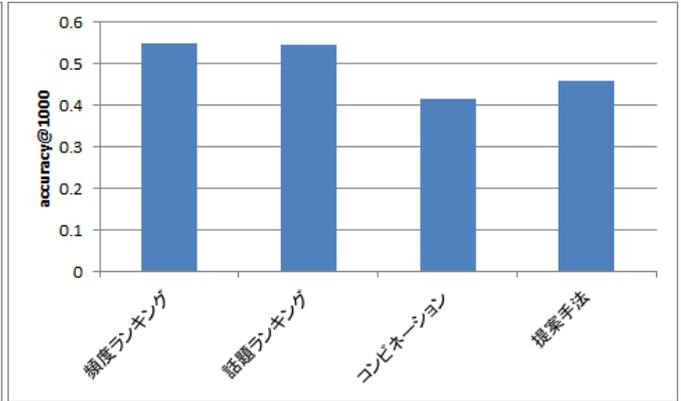


図 4 既存手法との比較 (Accuracy@1,000)

表 1 4 手法の特徴

手法	嗜好の考慮	会話ログの活用
頻度ランキング	×	×
話題ランキング	×	
コンビネーション		×
提案手法		

3.3 章にて PLSA を適用した際に算出した  $P(w|u)$  のうち,  $w$  が話題 (ハッシュタグ) であるものを抜き出して利用した. この際の嗜好の算出には, 4.3 章にて述べたように, 全期間のデータを利用する. その後, パラメタ  $t$  として与えられる期間において出現している話題に絞り, 推薦を行った.

頻度ランキングと話題ランキングの 2 つのランキング手法は, 期間  $t$  のパラメタを持つ. また, コンビネーションと提案手法の 2 つのパーソナライズ手法は, クラスタ数  $|G|$  と期間  $t$  のパラメタを持つ. これらのパラメタは, クラスタ数  $|G|$  が 10,50,100,200,300 のとき, 期間  $t$  が 1ヶ月~12ヶ月, 全期間のときについて実験を行い, それぞれの手法において最適なものを利用した.

#### 4.5 実験結果

本実験で用いた評価指標は Macro F-measure@K である. これは, 各ユーザに対する K 件の話題の出力において, システムが正しく会話を予測できたか? (Precision), もれなく予測できたか? (Recall), 出現頻度が低い話題であっても正しく予測できたか? の 3 つの評価軸をもつ指標である. 前述した通り, 我々は, ローカルな話題で会話することを話題の要件としている. そのため, 出現頻度が低い話題であっても正しく予測できているかを汲み取ることが可能な, Macro F-measure を選択した. Macro F-measure の式を以下に示す.

$$\frac{1}{|I|} \sum_{i \in I} \frac{2R_i}{N_i + C_i} \quad (6)$$

ここで,  $R_i$  は話題  $i$  についてシステムが正解した数,  $N_i$  は話題  $i$  についてシステムが出力した数,  $C_i$  はテストデータに含まれる話題  $i$  についての正解の数である. 今回は推薦可能なユニーク話題数が 9,924 と多いため,  $K=1,000$  とした.

結果を図 3 に示す. 横軸が手法を, 縦軸が Macro F-measure の値を表している. ランキング手法 (頻度ランキング, 話題ラ

ンキング) よりもパーソナライズ手法 (コンビネーション, 提案手法) の方が高い精度を得ており, ランキング手法の中では話題ランキングが, パーソナライズ手法の中では提案手法が高い精度を得た.

## 5. 考察

### 5.1 定量的な評価結果の分析

#### 5.1.1 ランキング手法 vs. パーソナライズ手法

実験の結果, ランキング手法よりも, パーソナライズ手法が高い精度を得た. これは, パーソナライズ手法は, 嗜好を考慮しユーザを絞ることによってローカルな話題を推薦することができたためと予想できる.

これを検証するため, 予測精度の評価に一般的に用いられる Accuracy@K を評価指標とした場合の結果を図 4 に示す. Accuracy@K は, 各ユーザに対し K 件の話題を推薦し, その中に正解が含まれていれば 1, 含まれていなければ 0 とする指標である. Macro F-measure と同様に,  $K=1,000$  とした. 各手法について最終的に算出するスコアは, 全テストユーザ中で正しく予測できた 2 ユーザの割合となる. Macro F-measure との主たる違いは, Accuracy が「出現頻度が低い話題であっても正しく予測できたか?」を評価軸として考慮しない点である.

結果を見ると, パーソナライズ手法よりもランキング手法の方が, ランキング手法の中では話題ランキングが, パーソナライズ手法の中では提案手法が, Accuracy を用いた評価において高い精度を得たことがわかる. ここで注目すべきは, Macro F-measure と Accuracy との間で, ランキング手法とパーソナライズ手法との関係が逆転していることである. この理由として, テストデータ中で観測された話題の出現回数がべき乗分布に従うため, ランキング上位の話題を推薦するとそれだけで高い Accuracy を得ることができたことが挙げられる. つまり, ランキング手法は, いわゆる世間話のような「メジャーな話題」を推薦することで, 世間で話されている話題を再現していると言える. 参考までに, テストデータ中の話題を頻度順に横軸に並べ, 各話題の頻度を縦軸としてプロットしたグラフを図 5 に示す. 両対数をとったときに直線に近い形状をとることから, べき乗分布に従うことがわかる.

我々の考え方は, 「ローカルな話題」を推薦することこそが,

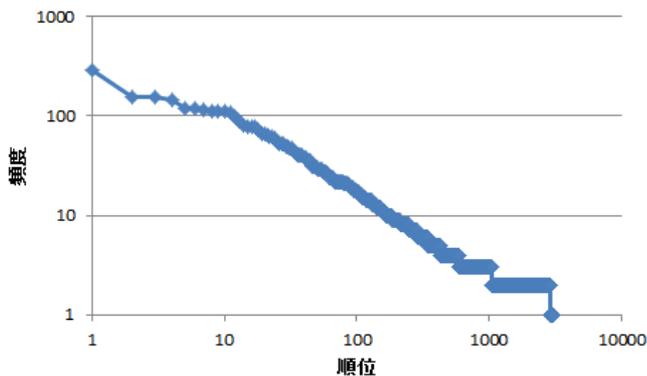


図5 テストデータ中の話題頻度のプロット

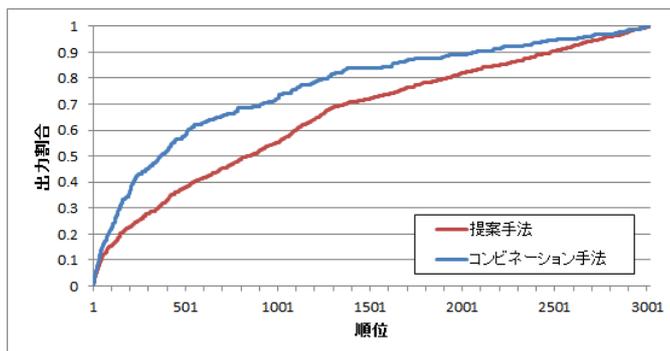


図6 各手法の出力のプロット

2 ユーザの会話が盛り上がるというものである (話題の要件 2) . 出現頻度が低い話題の予測精度を考慮した Macro F-measure の結果と照らし合わせると、ランキング手法が話題の要件 2 を満たしていない一方で、パーソナライズ手法は、嗜好を考慮しユーザを絞ることによって話題の要件 2 を満たしていると結論付けることができる .

### 5.1.2 提案手法 vs. コンビネーション

Macro F-measure と Accuracy による両評価において、提案手法がコンビネーションよりも高い精度を得た . ここで、提案手法とコンビネーションの更なる比較のために、各手法が全テストユーザに対して出力した推薦リストを集計した . 話題を頻度順に横軸に並べ、全出力数に対する各話題の割合を縦軸とした累積分布を図 6 に示す . 図 6 に示すように、提案手法の方がコンビネーションよりもランキング上位における立ち上がりが遅い . つまり、提案手法はコンビネーションよりも、出現頻度が低い話題を出力している . また、出現頻度が低い話題を出力しているにも関わらず、Accuracy における評価でも高い精度を得ていることから、出現頻度が低い話題による会話を予測できたことがわかる . これが Macro F-measure による評価における提案手法のコンビネーションに対する優位性につながったと考える .

## 5.2 定性的な推薦リストの分析

### 5.2.1 ランキング手法 vs. パーソナライズ手法

携帯電話関連に強い嗜好を持つ 2 ユーザに対する、4 手法の推薦リストを表 2 に示す . 各ランキング手法が上位に出力する

話題が、メジャーな話題であると言える . 推薦リストの内容を見ると、特に #ohayo などがメジャーな話題として特徴的である . 一方で、各パーソナライズ手法が #lynx などの携帯電話に関連するローカルな話題を推薦していることがわかる . 各ランキング手法も #androidjpn などの携帯電話に関連する話題を出力しているが、これはメジャーな話題であるために出力されているものである .

### 5.2.2 提案手法 vs. コンビネーション

コンビネーションは、PLSA の特徴から、嗜好に含まれる話題の中でも、出現頻度が高い話題を上位に出力する傾向があると考えられる . 出力を見ても、#xperia や #androidjpn などの、頻度ランキングも出力しているような出現頻度が高い話題から出力していることが見てとれる . 一方で提案手法は、#felicathon という Felica 対応アプリのプログラミングコンテストの話題や、#camangi という、SIM フリーの Android タブレット PC の話題など、出現頻度とは異なる基準によって出力していることがわかる . 以上から、コンビネーションは共通ではあるが、メジャーな話題を出力してしまう可能性が高いと言える .

また、提案手法の出力には、前述の #felicathon、#hikarijpn という光の道論争の話題などの意見が分かれやすい話題や、#lynx という 3D 液晶を世界で初めて搭載したスマートフォンの話題などの話題性が高い話題が含まれており、議論や雑談になりやすい、会話らしい話題が含まれていると言える . 表 1 に示した通り、提案手法が会話ログを活用しているのに対して、コンビネーションは、投稿ログであるか会話ログであるかを区別せずにログを利用する . その結果、提案手法の方が効果的に、会話がしやすい、話題としてふさわしい話題を推薦できたと言える . なお、話題ランキングの出力にも会話らしい話題が含まれていることも、会話ログの利用が有効であったことを裏付けていると言える .

### 5.3 パーソナライズ手法におけるパラメタの影響

ここでは、パラメタを変化させたときの実験結果を利用し、各パーソナライズ手法について考察を行う . 提案手法の結果は図 7、コンビネーションの結果は図 8 の通りである . 縦軸は Macro F-measure の値であり、横軸は時間窓長、色分けはクラスタ数を表している . 実験結果から、提案手法はクラスタ数が 50、時間窓長は過去 4ヶ月が、コンビネーションはクラスタ数が 200、時間窓長は全期間が最適なパラメタであることがわかった .

提案手法は、クラスタ数がある程度の数になると、精度に大きな変化がないことが見てとれる . これは、提案手法においてクラスタ数を増やすということは、他ユーザとの類似度をより詳細な次元で求めるということに相当するためだと考えられる . つまり、ユーザの類似度計算は 50 次元で収束しており、それ以上の詳細さで算出しても変化がないためだと考えられる . ただし、今回のデータは比較的小規模であったため、他のユーザとの類似度を求めるためには 50 次元がベストの結果となったが、データの量が増えれば増えるほど、より多くの嗜好が存在すると考えられるため、それに伴いベストな次元数も増えるこ

表 2 4 手法の出力

出力順位	頻度ランキング	話題ランキング	コンビネーション	提案手法
1	#ohayo	#hikarijip	#xperia	#felicathon
2	#keiba	#xperia	#androidjp	#androidjp
3	#twitbackr	#desire	#android	#hikarijip
4	#nhk	#ohayo	#librahack	#sh03c
5	#nowplaying	#ryomaden	#is01	#lynx
6	#androidjp	#hikari_road	#lynx	#fm600
7	#xperia	#androidjp	#ht03a	#camangijip
8	#nicovideo	#cmam	#nicovideo	#xperia
9	#seiji	#kosodate	#nowplaying	#librahack
10	#fujitv	#kodomomo	#galaxy	#iCre

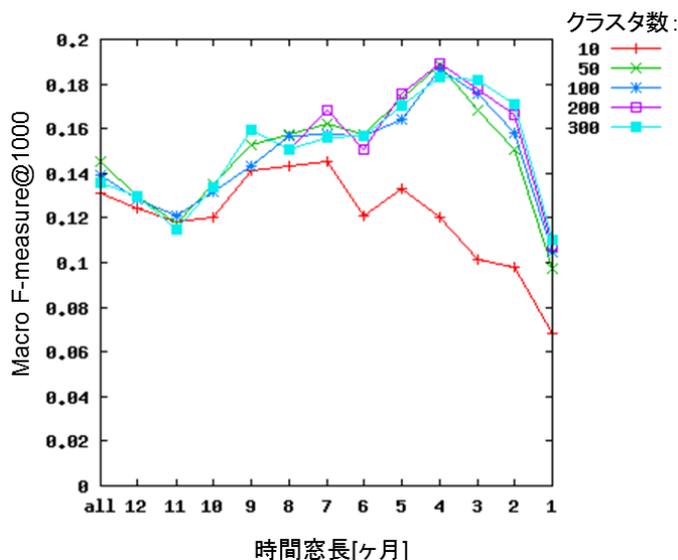


図 7 提案手法におけるパラメタの影響 (Macro F-measure@1,000)

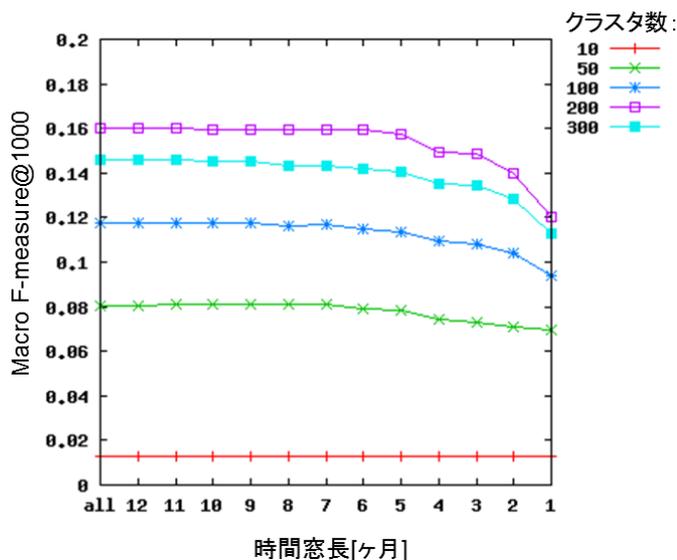


図 8 コンビネーションにおけるパラメタの影響 (Macro F-measure@1,000)

とが予想できる。

提案手法は、時間窓長を 4ヶ月とした際に最も精度が高かった。これは、データ数が増えれば増えるほど正解できる話題の数が増え、精度が上がるという現象と、古いデータから推薦しても、古い話題で会話するユーザが少ないために精度が下がるという現象が同時に起きていると考えられる。今回のデータでは、過去 4ヶ月という結果となったが、データの量が増えれば増えるほど、より短い期間でも高い精度を得ることができると考える。

コンビネーションについては、時間窓長を全期間とした際に最も精度が高かった。これは、コンビネーションの特徴によるものであると考えられる。コンビネーションは、全期間において話題への嗜好を算出しているため、時間窓長を変化させても、ユーザ  $u, u'$  における話題  $i$  の関連度  $rel(u, u', i)$  は変化しない。つまり、コンビネーションにおいて時間窓長を短くすることは、推薦できるアイテムの数を減らすだけにすぎない。そのため、提案手法は与えられた時間窓における話題を出力できているのに対して、コンビネーションは全期間における嗜好と合う話題を出力しているのみで、時間窓を有効に利用できていないと言える。また、提案手法は、期間が短い際にコンビネーションよりも高い精度を得ていることから、より流行に沿った

話題を出力することができるというメリットもあると言える。

## 6. 関連研究

本研究には、会話のサポートという側面と、複数ユーザに対する推薦という側面がある。それぞれの分野で行われている研究について、順に述べる。

### 6.1 会話のサポート

ユーザの入力に対し、適した返答をする対話システムの構築を目的として、会話のモデリングが研究されている [6] [9]。我々の研究は、会話のやりとりではなく、会話のきっかけとなるような話題の推薦を目指している点と、ユーザとシステムの会話ではなく、ユーザ同士の会話をサポートする点で異なる。

Zubiaga らは、会話のトリガとは、ニュース、イベント、噂、記念であると定義し、SVM を利用して、ツイートがこの 4 クラスに分類している [14]。また、Rowe らは、ツイートの言語的特徴とユーザの特徴を素性として、SVM を利用して会話の回数を予測している [10]。これらの研究はともに、会話につながるツイートの抽出を目的としていると考えられる。しかし、これらの研究においては、推薦されるユーザにとって話題となるかどうか、というパーソナライズが行われていない。また、Chen らは、会話の長さ、嗜好の近さ、ソーシャルネットワーク

上のユーザ間距離を考慮し、「一連のユーザ間会話」を推薦している [3]。この研究では、ユーザに対するパーソナライズが行われているものの、一人のユーザに対する推薦である。我々の研究は、複数のユーザに対してパーソナライズした話題の推薦を目的としており、Twitter の会話ログを活用することで、ユーザ間の会話をモデリングするものである。

## 6.2 複数ユーザに対する推薦

これまでの推薦タスクは、主に、一人のユーザに対する推薦を実現するものであった。例えば Yang らは、友人関係が近いユーザ同士はアイテムに対する嗜好も近いと仮定し、購買ログとソーシャルグラフを併用することで、アイテム推薦を行う手法を提案している [13]。一方で、複数のユーザに対してアイテムを推薦するグループレコメンデーションも広く研究されている [1] [5] [11]。これらは、個人の嗜好モデルを組み合わせ、複数人の共通な嗜好を間接的に導きだしてアイテムを推薦するものである。しかし、2人で過去にどんな映画を見て、どんなレストランに行ったかといったユーザ同士のインタラクション情報を用いていない。そのため、ユーザ同士の関係性によっては、ユーザ同士がともに興味を持つ話題を推薦することが難しかった。我々の研究は、ユーザ同士のインタラクション情報として会話ログを利用することで、複数ユーザに対する推薦を高精度に実現している。

中茂らは、会話の最中に、会話内容の関連画像をテーブルトップに映写する手法を提案している [16]。この研究は、我々の研究と同様に、複数ユーザに対する情報提示を行うものだが、その目的は、今ある話題の会話を円滑に進行させることである。我々の研究は、話題となるアイテムを推薦することを目的としており、すでに会話が発生している必要もない。

## 7. まとめと今後の予定

本稿では、Twitter 上に含まれる会話ログを利用した、2ユーザに対する話題推薦手法を提案した。2ユーザの会話が盛り上がるために話題が満たすべき4つの要件と、実サービスとして運用するために手法が満たすべき2つの要件を洗い出し、これらを満たす提案手法を考案した。評価実験においては、Twitter ユーザ同士が実際に会話した話題を予測するタスクをもとに、提案手法が、要件を満たす話題を自動で探し出せるかを検証した。話題の満たすべき要件を意識し、出現頻度が低い話題であっても正しく予測できるかを考慮した評価指標を用いることで、既存手法と比べ提案手法が、要件を満たす話題、すなわち、「2ユーザの会話が盛り上がる話題」を予測できることを示した。

ただし、クローリングしたデータは比較的小規模であるため、URLを用いた予測実験を行うことができなかった。今後は、クローリング方法を工夫するなどして、より大規模なデータセットを利用した会話 URL 予測実験を行う。

今回の実験においては、提案手法が「2ユーザの会話が盛り上がる話題」を予測できることを示した。しかし、実際のユーザに対し話題推薦した結果、会話が誘発されるかは未検証である。そのため、ユーザに対する実証実験を通じて提案手法の効果を検証する必要があると考える。具体的には、Twitter ユー

ザに対して提案手法を用いて話題の推薦を行った結果、過去に会話したことのあるユーザと会話したか (人間関係強化の観点)、過去に会話したことのないユーザと会話したか (人間関係拡大の観点) を評価することを予定している。

## 文 献

- [1] S. Amer-Yahia, S.B. Roy, A. Chawlat, G. Das and C. Yu, "Group recommendation: semantics and efficiency", Proceedings of the VLDB Endowment, vol. 2, no. 1, pp. 754-765, 2009.
- [2] M. Burke, C. Marlow and T. Lento, "Social network activity and social well-being", Proceedings of the 28th international conference on Human factors in computing systems, pp. 1909-1912, 2010.
- [3] J. Chen, R. Nairn, E. Chi, "Speak little and well: recommending conversations in online social streams", Proceedings of the 2011 annual conference on Human factors in computing systems, pp. 217-226, 2011.
- [4] J. Coleman, "Social capital in the creation of human capital", The American journal of sociology, vol. 94, no. 1, pp. 95-120, 1988.
- [5] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra and K. Seada, "Enhancing group recommendation by incorporating social relationship interactions", Proceedings of the 16th ACM international conference on Supporting group work, pp. 97-106, 2010.
- [6] J. Henderson, O. Lemon and K. Georgila, "Hybrid reinforcement/supervised learning of dialogue policies from fixed datasets", Computational Linguistics, vol. 34, no. 4, pp. 487-511, 2008.
- [7] IT用語辞典バイナリ, "mixi 疲れとは", <http://www.sophia-it.com/content/mixi%E7%96%B2%E3%82%8C>.
- [8] IT用語辞典バイナリ, "twitter 疲れとは", <http://www.sophia-it.com/content/Twitter%E7%96%B2%E3%82%8C>.
- [9] Y. Minami, R. Higashinaka, K. Dohsaka, T. Meguro and E. Maeda, "Trigram dialogue control using POMDPs", Spoken Language Technology Workshop 2010 IEEE, pp. 336-341, 2010.
- [10] M. Roew, S. Angeletou, H. Alani, "Predicting discussions on the social semantic web", The Semantic Web: Research and Applications, pp. 405-420, 2011.
- [11] S. Seko, M. Motegi, T. Yagi, S. Muto, M. Abe, "A Contents Filtering Method for Group Recommendation using Life Log", FIT: Forum on Information Technology, vol. 9, no. 4, pp. 505-510, 2010.
- [12] Y. Tian, Q. He, Q. Zhao, X. Liu and W. Lee, "Boosting social network connectivity with link revival", Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 589-598, 2010.
- [13] S. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng and H. Zha, "Like like alike: joint friendship and interest propagation in social networks", In Proceedings of the 20th international conference on World wide web, pp. 537-546, 2011.
- [14] A. Zubiaga, D. Spina, V. Fresno and R. Martínez, "Classifying trending topics: a typology of conversation triggers on twitter", In Proceedings of the 20th ACM international conference on Information and knowledge management, pp. 2461-2464, 2011.
- [15] 齊藤 徹, "mixi, Twitter, Facebook, Google+, LinkedIn 2011年11月最新ニールセン調査 in the loop", <http://media.loops.net/saito/2011/12/20/mixi-twitter-facebook-google-linkedin/>, 2011.
- [16] 中茂 睦裕, 中山 彰 and 小林 稔, "テーブルトップへの画像刺激による遠隔コミュニケーションの分析", ヒューマンインタフェースシンポジウム 2006 論文集 2, pp. 695-700, 2006.