

論文 / 著書情報
Article / Book Information

題目(和文)	ハッシュタグコミュニティ検出に基づくニュース・トピック関連ソーシャルメディアマイニング
Title(English)	News-Topic Oriented Social Media Mining Based on Hashtag Community Detection
著者(和文)	肖峰
Author(English)	Feng Xiao
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第9601号, 授与年月日:2014年6月30日, 学位の種別:課程博士, 審査員:徳田 雄洋,米崎 直樹,佐伯 元司,権藤 克彦,西崎 真也
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第9601号, Conferred date:2014/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

専攻： Department of	計算工学	専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of	（ 学術 ）
学生氏名： Student's Name	肖 峰		指導教員（主）： Academic Advisor(main)	徳田 雄洋	
			指導教員（副）： Academic Advisor(sub)		

要旨（和文 2000 字程度）

Thesis Summary (approx.2000 Japanese Characters)

本論文は“News-Topic Oriented Social Media Mining Based on Hashtag Community Detection”と題し、7 章から成っている。

第 1 章“Introduction”では、本研究の動機と目的について述べている。近年ソーシャルメディアでは、ユーザによりニューストピックに関する情報収集や情報発信が活発に行われている。本研究では最も有力なソーシャルメディアの 1 つである Twitter を対象としている。Twitter ユーザは他のユーザをフォローすることにより、関心あるニューストピックの情報を取得できる。しかしながらニューストピックに関する情報取得のためにどのユーザをフォローすべきかという問題は、一般的ユーザ推薦やキーワードを含むツイートを発信するユーザ推薦に比べ、ツイート 140 文字制限下では非常に困難な問題となる。本論文は、ニューストピックに関する情報取得のためにフォローすべきユーザの新しい推薦方法を与えていると述べている。

第 2 章“Background and Related Work”では本研究の背景と関連研究について述べている。本研究は 3 つの研究領域に関係している。特徴的共起語、タグ推薦、重要ユーザ推薦の 3 研究領域である。既存の特徴的共起語の検出法は対称的方法か、非対称的方法であるが連語検出に留まっている。既存のタグ推薦法は内容類似性やタグ共起関係に基づくという限界を持っている。既存の重要ユーザ推薦法は、トピックに独立か、重要ユーザのタイプや行動を無視していると述べている。

第 3 章“Detection of Characteristic Co-Occurrence Words from News Articles”では、与えられた対象語について、ニュース記事中で特徴的に共起する語を検出する新しい手法として、PIOLog 法を提案している。特徴的共起語は、対象語を含む記事に多く出現するが、対象語を含まない記事には少なく出現するという観察に基づき、対象語を含む記事とそれ以外の記事での出現確率の比を利用して共起語のスコア付けを行うと述べている。

第 4 章“News-Topic Oriented Hashtag Recommendation in Twitter”では、ニューストピックに関連するハッシュタグ推薦の新しい手法を提案している。ニューストピックを表す特徴語ベクトルとハッシュタグを表す特徴語ベクトルを生成し、その類似度を計算してハッシュタグ推薦を行う。ニューストピックを表す特徴語ベクトルの生成には、第 3 章で提案した PIOLog 法で検出した特徴的共起語を利用する。ハッシュタグを表す特徴語ベクトルの生成には、PIOLog 法を拡張した PIOLogH 法を利用する。PIOLogH 法では、対象のハッシュタグを含むツイート中とそれ以外のツイート中の出現確率の比を利用する。文字数に制限のあるツイートだけでなくニュース記事も利用することにより、既存手法における情報量不足の問題を克服している。

第 5 章“Finding News-Topic Oriented Influential Twitter Users”では、ニューストピックに関する重要な Twitter ユーザを発見する新しい手法を提案している。本論文では、重要 Twitter ユーザを、ニューストピックに関連する有益な情報を発信するユーザとニューストピックにおいて権威あるユーザの 2 種類に分類し、それぞれを発見する手法として RetweetRank 法と MentionRank 法を示している。本手法では、第 4 章のハッシュタグ推薦手法で獲得したハッシュタグを利用してニューストピックに関連するツイートを収集し、リツイートとメンションの関係を表すユーザグラフをそれぞれ構築する。RetweetRank 法はリツイート関係を表すグラフをもとに有益な情報を発信するユーザを発見し、MentionRank 法はメンション関係を表すグラフをもとに権威あるユーザを発見する。グラフを利用する既存手法の PageRank 法と異なり、ユーザ間のハッシュタグ利用の類似性やニューストピックとの関連性をエッジの重みや遷移確率に反映する点が提案手法の特徴であると述べている。

第 6 章“Experiments and Evaluation”では、本論文で提案した特徴的共起語検出、ハッシュタグ推薦、Twitter ユーザ推薦の手法と既存手法の比較実験と評価を行っている。特徴的共起語検出では Jaccard 係数法と対数尤度比法を比較対象とし、提案手法が優れていると述べている。Jaccard 係数法は対称性を持ち、非対称の特徴的共起語を検出する手法としては適さない。対数尤度比法は、連語を検出する目的には適しているが本論文で扱う特徴的共起語の検出には適さないとしている。ハッシュタグ推薦では、TF-IDF 法を比較対象とし、提案手法が優れていることを示している。Twitter ユーザ推薦では、ツイート数、被リツイート/メンション数、PageRank 法を比較対象とし、提案手法が優れていることを示している。この違いは既存方法がユーザの行動やニューストピックの関連を考慮していないためとしている。

第 7 章“Conclusion”では、研究成果のまとめと今後の研究の方向について議論している。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

(博士課程)
Doctoral Program

論文要旨

THESIS SUMMARY

専攻：	計算工学	専攻
Department of		
学生氏名：	肖 峰	
Student's Name		

申請学位 (専攻分野)：	博士	(学術)
Academic Degree Requested	Doctor of	
指導教員 (主)：	徳田 雄洋	
Academic Advisor(main)		
指導教員 (副)：		
Academic Advisor(sub)		

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

This thesis presents two new methods to find two types of influential Twitter users about news topics related to the target word. RetweetRank is proposed to find content-based influential Twitter users who often post valuable tweets about the news topic. MentionRank is proposed to find authority-based influential Twitter users who have high authority on the news topic. To collect tweets related to news topics, hashtags which are highly relevant to the news topic are detected by our news-topic oriented hashtag recommendation approach. As basic components of researches above, two new methods (PIOLog and PIOLogH) to detect characteristic co-occurrence words with the target word and hashtags are proposed. Newly proposed PIOLog and PIOLogH methods could detect characteristic co-occurrence words which often co-occur with the target word in news articles for the same topic and the hashtag while they are less likely to appear in other news articles or co-occur with other hashtags. These detected words are topic-specific and informative for the news topic and hashtag. Existing methods are either symmetric method which is not topic-oriented, or used to detect word collocation which is different from our purpose. We also propose a new approach to recommend news-topic oriented hashtags using characteristic co-occurrence words detected by PIOLog and PIOLogH methods. Hashtags which are highly relevant to the news topic are recommended to users who want to use hashtag in their tweets about the news topic. Existing approaches for hashtag recommendation are not news-topic oriented. They often use the similarity between tweets while there is far not enough information in a single tweet due to its length limitation. Our approach concatenates on tweets having the same hashtag, and the hashtag could be represented by more information. RetweetRank and MentionRank methods are proposed to find content-based and authority-based influential Twitter users from all Twitter users using news-topic oriented hashtags. They consider not only link structures of different type among Twitter users, but also the relevance between Twitter users and the news topic. Some existing methods ignore the link structure, others do not consider link structures of different type or user's topic relevance, which bring negative affection to their results.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).