Finding Topic Experts in the Twitter Dataset Using LDA Algorithm

Ashwini Anandrao Shirolkar, Department of Technology, SU, Kolhapur, India
R. J. Deshmukh, Department of Technology, SU, Kolhapur, India

ABSTRACT

In microblogging services like Twitter, the expert judgment problem has gained increasing attention in social media. Twitter is a new type of social media that provides a publicly available way for users to publish 140-character short messages (tweets). However, previous methods cannot be directly applied to twitter experts finding problems. They generally rely on the assumption that all the documents associated with the candidate experts contain tacit knowledge related to the expertise of individuals. Whereas it might not be directly associated with their expertise, i.e., who is not an expert, but may publish/re-tweet a substantial number of tweets containing the topic words. Recently, several attempts use the relations among users and twitter list for expert finding. Nevertheless, these strategies only partially utilize such relations. To address these issues a probabilistic method is developed to jointly exploit three types of relations (i.e., follower relation, user-list relation and list-list relation) for finding experts. LDA algorithms are used for finding topic experts. LDA is based upon the concept of searching for a linear combination of variables (predictors) that best separates two classes (targets). Semi-supervised graph-based ranking approach (SSGR) to offline calculate the global authority of users. Then, the local relevance between users and the given query is computed. Then, the rank of all the users is found and the top-N users with the highest-ranking scores. Therefore, the proposed approach can jointly exploit the different types of relations among users and lists for improving the accuracy of finding experts on a given topic on Twitter.

KEYWORDS

Expert Finding, Graph-Based Ranking Approach, Hashtag, Semi-Supervised, Sentiment Analysis, Twitter

1. INTRODUCTION

The purpose of the data mining process is to extract information from a data set and convert it into an acceptable structure for further use. Data mining is widely used in diverse areas. There is a number of the commercial data mining system available today, and yet there are many challenges in this field. Data mining is widely used in intrusion detection, biological data analysis, scientific applications, social networking.

Mining in social networking is an important data mining task with broad applications. Expert finding problem has gained increasing attention in social media, it builds a weighted graph by

DOI: 10.4018/IJAEC.2019040103

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
considering both the topical similarity between two users and follower graph, and then employ page Rank algorithm to find topic-specific influential users. Pal et al. (2011) extract user’s features from the follower graph, and users posted tweets, and then employ a Gaussian-based mixture model to cluster users for ranking. The Twitter Rank and pal’s work only consider user-user single relation.

Gosh et al. (2012) propose to utilize Twitter List to analyze the attributes of Twitter users. In their subsequent work, they develop a system named Cognos (Ghosh et al., 2012) to infer the topical expertise of users by utilizing only user-list relation in Twitter List, which captures the wisdom Twitter crowds. Cognos represents each user by the meta-data of Twitter lists that contain the user and then employs a similarity measure to compute the similarity score between each user and topical query, which is used to rank users for search.

The proposed method jointly exploits three types of relations (i.e., follower relation, user-list relation and list relation) for finding experts. LDA algorithm is used for finding topic experts. LDA is based upon the concept of searching for a linear combination of variables (predictors) that best separates two classes (targets). A Semi-Supervised Graph-based Ranking approach (SSGR) to offline calculate the global authority of users. In SSGR, employ a normalized Laplacian regularization term to jointly explore the three relations, which is subject to the supervised information derived from Twitter crowds. Then online compute the local relevance between users and the given query. By taking advantage of the global authority and local relevance of users, rank all of the users and find top-N users with the highest-ranking scores.

The proposed approach jointly exploits the different types of relations. Two types of information to target Twitter expert finding problem, namely: i) Local Relevance: the similarity between users published tweets and the given query; and ii) Global Authority: the global expertise scores of users on a given topic in Twitter. The proposed approach successfully integrates different types of user-related information (i.e., the crowdsourced Lists information, follower graph and user profiles) into a unified ranking framework for accurately inferring the topical expertise of users.

2. RELATED WORK

Ghosh, Sharma, Ganguly and Gummadi (2012) propose and investigate a new methodology for discovering topic experts in the popular Twitter social network. The methodology relies on the wisdom of the Twitter crowds; it leverages twitter lists, which are often carefully curated by individual users to include experts on topics that interest them and whose metadata (list names and descriptions) provides valuable semantic cues to the experts’ domain of expertise. In this list information to build Cognos, a system for finding topic experts in Twitter. Cognos infer user expertise more accurately than a state-of-art system that relies on the user’s bio or tweet content. Cognos scales well due to a built-in mechanism to update its expert's database with the new user.

Weng, Lim, Jiang and He (2010) focus on the problem of identifying influential users of microblogging services. Twitter, one of the most notable micro-blogging services, employs a social-networking model called “following,” in which each user can decide whom she wants to “follow” to receive tweets from without requiring the latter to give permission first. The Twitterrank algorithm an extension of the Page Rank algorithm, to measure the influence of users on Twitter. Twitterrank measures the similarity between users and the link structure into account. The Twitterrank works in two steps. First, it employs the Latent Dirichlet Allocation (LDA) model to detect the topics of individuals based on their tweets. Second, for each topic, it builds a weighted graph by considering both the topical similarity between two users and the follower graph, and then employ the Page-Rank algorithm to find topic-specific influential users.

Pal and Counts (2011) focus on the problem of identifying topical authorities in microblogs. Content in microblogging systems such as Twitter is produced by tens to hundreds of millions of users. This diversity is a notable strength but also presents the challenge of finding the most exciting and authoritative authors on any given topic. In this proposes an approach that extracts user’s features
Concept and Definition of Complexity  
[www.igi-global.com/chapter/concept-definition-complexity/24185?camid=4v1a](www.igi-global.com/chapter/concept-definition-complexity/24185?camid=4v1a)

Positioning Against Other Theories  
Lars Taxén (2010). *Using Activity Domain Theory for Managing Complex Systems* (pp. 135-161).  
[www.igi-global.com/chapter/positioning-against-other-theories/39676?camid=4v1a](www.igi-global.com/chapter/positioning-against-other-theories/39676?camid=4v1a)