A Novel Topical Authority-Based Microblog Ranking

Yanmei Zhai[1], Xin Li[1,⋆,⋆⋆], Jialiang Chen[1], Xiumei Fan[1,⋆⋆⋆], and William K. Cheung[2]

[1] School of Computer Science, Beijing Institute of Technology, Beijing, China
[2] Department of Computer Science, Hong Kong Baptist University, Hong Kong

Abstract. The high volume of microblogs produced daily together with their rich social structure makes microblogs' better query and filtering ever challenging. In the literature, most of the existing ranking methods are based on the overall popularity of the authors and the tweets without considering author's expertise. In this paper, we propose the topical authority-based ranking methods for social networks like Twitter and investigate how the underlying topical feature modeling can be optimized for performance boosting. In particular, we present a detailed study on the empirical distributions of the topical features. We propose the use of specific parametric forms for different features, which we believe to be crucial as the value of the cumulative distribution function is explicitly used for topical authority ranking. We applied the extended topical authority-based ranking method to a Twitter dataset for ranking keyword-matched microblogs. The experimental results show that our proposed approach outperforms a number of existing approaches by a large margin which verify the effectiveness of our proposed features and the importance of the topical authority for ranking microblogs.

Keywords: Topical Authority, Feature Distribution, Microblog Ranking.

1 Introduction

The recent proliferation of micro-blogging causes tens of thousands of microblogs produced daily. The availability of the large amount of microblog data, often together with the user profiles, allows a number of data mining tasks possible, e.g., hot topic detection [1], opinion leader detection [2], etc. Yet, this also brings new challenges to microblog search engines like Twitter, in particular the high demand of billions of daily search queries and the need to provide

⋆ Corresponding author.

⋆⋆ The work of Xin Li is partially supported by National Program on Key Basic Research Project under Grant No. 2013CB329605 and the NSFC Grant under Grant No. 61300178.

⋆⋆⋆ The work of X. Fan is supported in part by NFSC under Grant No. 61272509 and 61120106010, BNSF under Grant No. 4132049, and Specialized Research Fund for the Doctoral Program of Higher Education under grant No. 20136118110002.

L. Chen et al. (Eds.): APWeb 2014, LNCS 8709, pp. 105–116, 2014.

high quality microblogs for users. Microblogs are known to be fragmental and ephemeral, making accurate content filtering and retrieval non-trivial and also a hot research area. Various ranking strategies for microblogs have been proposed in the literature. The approaches adopted include the use of the content and specific features (e.g., tags) of the microblogs, as well as the bloggers' social structure (e.g, the author's popularity).

In general, the content-based strategies adopt variations of TFIDF-based cosine similarity to measure the content popularity [3]. And for those strategies based on microblog specific features [4], the number of hashtags, the length of tweet, the presence of URLs, and the number of retweets, etc., have been proposed for ranking tweet data. For the authority-based approach, the basic idea is to rank the tweets of the authors with more followers or more retweets higher. Intuitively, each author has his/her own expertise of some specific areas. For example, Alex J. Smola - the prestigious machine learning researcher has high authority in machine learning related domains, his twitter account "@smolix" distributes lots of useful resources related to ML research but few posts discussing food, trips etc. Thus treating author authority with no topical difference might not be appropriate in microblog ranking.

In [5], the Gaussian ranking algorithm for microblogs topical authority identification was proposed based on a set of so-called topical features which indicate the topical signals over tweets and authors. The use of the approach is mainly for the influential author detection. And its optimality is pretty much relying on the assumption that each feature follows a Gaussian distribution with their parameters estimated from the data.

In this paper, we propose to incorporate author's topical authority to enhance the performance of microblog ranking, with the conjecture that the topical authorities of the authors are good and robust signals to indicate the degree of relevancy with the query keywords. And we develop topical authority-based methodologies and conduct experiments in a Twitter dataset. Detailedly speaking, other than extending the set of topical features proposed in [5], we also adopt different parametric forms of probability distribution for the feature modeling so that the accuracy of our proposed cumulative distribution functions (CDFs) based approaches can be further optimized. The experimental results showed that our proposed approaches can significantly improve the ranking performance measured based on normalized discounted cumulative gain (NDCG) by over 20% when compared to both Gaussian-based and conventional ranking approaches. To the best of our knowledge, we are the first to incorporate topical authority into microblog ranking.

The remainder of the paper is organized as follows. Section 2 discusses the related work of microblog ranking and identification of topical authorities. Section 3 presents the feature extraction of all topical and conventional features, followed by several novel authority-based ranking methods. The details of the feature distribution modelling are described in Section 4. Section 5 reports the experimental results and performance evaluation. The conclusion and future work are presented in Section 6.

2 Related Work

In the literature, there exist lots of work on microblog ranking. The importance of considering author authority in Twitter author ranking was first demonstrated in [6]. *TweetRank* and *FollowerRank* [7] were proposed to rank tweets by considering the number of tweets posted by an author and the proportion of the followers in his networks, respectively. Different hybrid approaches that incorporate content relevance, user authority and tweet-specific features have also been considered to support real-time search and ranking [8,4]. In the literature, the existing works have validated the contribution of publishing authority to the microblog ranking approaches. However, the author authority is based on the conventional popularity instead of being evaluated in topics. And ranking the microblogs in consideration of author's topical authority is rarely discussed.

In domains other than microblog ranking, identification of topical author authority was first investigated by Jianshu *et al.* [9]. *TwitterRank* was proposed to identify author authority, which is a PageRank [10] similar approach by adopting both topical similarity and link structure. The topical distribution is constructed using the Latent Dirichlet Allocation (LDA) algorithm [11], and then a weighted user graph is derived accordingly with its edge weight indicating the topical similarity between authors. However the high complexity of the approach can not meet the requirement of real-time ranking. Aditya *et al.* [5] emphasized on real-time performance and first proposed a set of features for characterizing the topical authorities. They performed a probabilistic clustering over the feature space and computed a final list of top-k authors for a given topic by Gaussian-based inner-cluster ranking.

In this paper, we aim to enhance the ranking performance by identifying and incorporating topical authority into microblog ranking scheme. Detailedly, we i) propose two new features based on [5] as the author's topical follower signal and the conventional popularity signal respectively, and ii) adopt different parametric distributions for feature modeling so as to relax the Gaussian distribution assumption. This relaxation is particularly crucial as the cumulative distribution function values are explicitly used to compute the the ranking. Also, we evaluate the effectiveness of topical author authority in microblog ranking.

3 Topical Authority-Based Microblog Ranking

In this section we will present the feature extraction of all topical and conventional features as well as several novel authority-based ranking methods.

3.1 Topical Authority Feature Construction

We adopt and extend the topical authority metrics and features proposed in [5] to further enhance microblog ranking performance. And to make this paper self-contained we include some details of [5].

Topical Metrics in Microblogs. Following the setup of [5], we also utilize a list of metrics extracted and computed for each potential authority. Table 1

tabulates the metrics proposed in [5], where *OT, CT, RT, M* and *G* stand for metrics associated with the original tweets, conventional tweets, repeated tweets, mentions and graph-based characteristics, respectively. All the features indicate the morphology of tweets (the number of embedded URLs, hashtags, etc.), the way they are used (re-tweeting, mentions or conventional tweets), or the signal of author's topical interests. We here propose two additional metrics, $F1$ and $F2$, to indicate author popularity as people tend to have strong interests in celebrities. We then use these two metrics to define new features.

Table 1. List of Metrics of Potential Authority [5]

ID	Metric
OT1	Number of original tweets
OT2	Number of links shared
OT3	Self-similarity score in the words of tweets
OT4	Number of keyword hashtags used
CT1	Number of conversational tweets
CT2	Number of conversational tweets initiated by the author
RT1	Number of retweets of others'
RT2	Number of unique tweets (OT1) retweeted by other users
RT3	Number of unique users who retweeted authors tweets
M1	Number of mentions of other users by the author
M2	Number of unique users mentioned by the author
M3	Number of mentions by others of the author
M4	Number of unique users mentioning the author
G1	Number of topically active followers
G2	Number of topically active friends
G3	Number of followers tweeting on topic after the author
G4	Number of friends tweeting on topic before the author
F1	Number of followers
F2	Number of friends

Topical Features in Microblogs. Most of the topical features adopted in this paper are again based on [5] as shown in Table 2. Among them, *TS* indicates how much an author is involved with a specific topic. *SS* estimates the originality of author's topical tweets which also indicates author's topical signal. Additionally, $\overline{CS}$ estimates how much an author posts on a topic and how far he wanders from the topic into casual conversations. $\overline{CS}$ is proposed to distinguish real people from the agents or organizations, since people incline to fall into conversations out of courtesy. Referring to λ in $\overline{CS}$, it is used to discount the fact that the author did not initiate the conversation but simply replied back out of politeness. Intuitively, $\overline{CS}$ is less than $\frac{OT1}{OT1+CT2}$, and thus we can solve for λ with this constraint. Empirically, we solve λ to satisfy over 90% of users in our dataset. *RI* considers how many times the author's tweets have been retweeted by others so as to measure the content impact of author. *MI* is used to estimate the mention impact. Feature *ID* is to estimate the diffused influence by the author

in his own networks. And *NS* is to estimate the raw number of topical active users around the author. For *OT21* and *OT41*, they indicate the rate of link and keyword hashtag in original tweets respectively. *OT3* reflects the portion of words an author borrows from his previous posts including both on and off topics, where $S(s_i, s_j) = \frac{|s_i \cap s_j|}{|s_i|}$ is the similarity function defined over the set of words s_i and s_j which are extracted from the author's i^{th} and j^{th} tweets repectively. Moreover, before computing the scores, we should make author's tweets in time order, and apply stemming and stop-word removal.

Intuitively, for a specific area, the more followers an author has, the more influential he is; and the more attention an author receives, the more authoritative he is. Thus, we propose to include feature $F12$ as the conventional popularity signal considering that people tend to have great interests in celebrities' opinions, and also feature $GF1$ to indicate the author's topical follower signal. Both newly added features are found to be effective ones based on our empirical results.

Table 2. List of Features for Each User

	Feature	Description
TS	$\frac{OT1+CT1+RT1}{\lvert\#tweets\rvert}$	Topic Involvement Signal
SS	$\frac{OT1}{OT1+RT1}$	Topical Signal Strength
$\overline{CS}$	$\frac{OT1}{OT1+CT1} + \lambda\frac{CT1-CT2}{CT1+1}$	Non-Chat Signal
RI	$RT2 * log(RT3)$	Retweet Impact
MI	$M3 * log(M4) - M1 * log(M2)$	Mention Impact
ID	$log(G3+1) - log(G4+1)$	Information Diffusion
NS	$log(G1+1) - log(G2+1)$	Network Score
OT21	$\frac{OT2}{OT1}$	Link Rate
OT41	$\frac{OT4}{OT1}$	Keyword Hashtag Rate
OT3	$\frac{2*\sum_{i=1}^{n}\sum_{j=1}^{i-1} S(s_i,s_j)}{(n-1)*n}$	Word Self Similarity
GF1	$\frac{G1}{F1}$	Topical Follower Signal
F12	$\frac{F1}{F2}$	Follower Signal

3.2 Cumulative Distribution-Based Ranking

Since all the topical authority features are assumed to follow Gaussian distribution in [5], which however may not be true as to be discussed in Sec. 4.2. We adopt the feature distribution modelling approach in this paper and compute the cumulative distribution functions (CDF) of the topical authority features to calculate the author's *Authority Score* (AS). For author x_i, AS is defined as:

$$AS(x_i) = \prod_{f=1}^{m} F_f(x_i^f; \Theta_f) \tag{1}$$

where F_f denotes the CDF of feature f with parameter Θ_f and m is the number of features (similarly hereinafter). With the conjecture that the conventional

authority features and topical authority features may carry different weights, one can explore a weighted version of the Authority Score, given as:

$$AS(x_i) = [\prod_{f=1}^{11} F_f(x_i^f; \Theta_f)]^{\beta} [F_{12}(x_i^{12}; \Theta_{12})]^{(1-\beta)} \quad (2)$$

where $\beta \in (0, 1)$ denotes the trade-off parameter between topical authority and conventional authority. In our experiments, the empirical settings of β are over 0.7.

Other than the proposed CDF-based ranking approaches, there exist a number of other possibilities (Seen in Table 3). *Conv_based* corresponds to the ranking method based on conventional author popularity with only feature $F12$ used as the authority score. *Gaus-10* refers to the Gaussian-based ranking method with AS defined as $\prod_{f=1}^{m} \int_{-\infty}^{x_i^f} \mathcal{N}(x; \mu_f, \sigma_f) dx$, where μ_f and σ_f denote the mean and standard deviation of feature f. *SUM-based* defines the AS as $\sum_{f=1}^{m} x_i^f$, and *SUM-12* corresponds to summation over all 12 features shown in Table 2. Similarily, *MUL-based* method defines AS as $\prod_{f=1}^{m} x_i^f$.

Table 3. List of Authority Ranking Methods

Ranking Methods	Description
Conv_based	Conventional authority-based ranking by feature $F12$ only
Gaus-10	Gaussian-based over top 10 features in table 2
SUM-12	Summation-based over all 12 features in table 2
MUL-12	Multiplication-based over all 12 features in table 2
CDF-10	CDF-based over top 10 features in table 2
CDF-12	CDF-based over all 12 features in table 2
CDF_weighted	Weighted version of *CDF-12*

4 Optimizing Feature Modeling

In this section, we first present the statistics of the Twitter dataset we used and then suggest better design of the probability distribution for each feature to achieve the model optimality.

4.1 DataSet

We use a Twitter dataset which was collected from June 11st 2009 to October 8th 2009. All collected tweets together with their relationship profiles takes up about $65.8G$ storage space. We select five hot hastags as the keywords. They are *google, healthcare, iran, music* and *twitter*. For each keyword, we collect thousands of most recent and best matched tweets via substring matching and obtain the corresponding authors' relationship.

The statistics of our dataset are shown in Table 4, where |MTN| means the number of matched tweets by keywords, |UTN| indicates the number of unique authors. |UFoN| and |UTFoN| represent the number of unique followers and that of unique topical followers respectively. Similarly, |UFrN| and |UTFrN| indicate the number of unique friends and that of unique topical friends respectively.

Table 4. Dataset Statistics

keywords	google	healthcare	iran	music	twitter
|MTN|	5,371	2,919	4,162	5,175	5,208
|UTN|	4,221	1,949	1,953	4,446	4,651
|UFoN|	788,149	600,355	917,983	634,016	832,140
|UTFoN|	131,281	34,292	57,197	143,870	321,804
|UFrN|	550,980	347,651	388,208	426,138	604,472
|UTFrN|	114,565	30,401	39,763	121,119	272,095

4.2 Feature Distributions of Different Parametric Forms

We first group the features into four categories based on the form of their underlying distributions. The groupings are {*ID, GF1*}, {*TS, F12*}, {*MI, RI, OT41*}, and {*NS, OT3, OT21, CS, SS*}. For each category, they fit the corresponding features with distribution functions of same parametric form. Due to the page limit, we only present some of the features in detail here.

Fig.1 shows the probability distributions of different features. It can be easily noted that features *ID, GF1* and *MI* can be fitted well by Gaussian distribution, and features *TS* and *NS* are unlikely Gaussian. Fig.2 gives the Q-Q plots (where "Q" stands for quantile) of some features based on the Gaussian assumption. It is obvious that only features *ID* and *GF1* end up with good fitting as indicated by having not too many points deviated from the straight line $y = x$. For features *TS* and *NS*, they are found to be better fitted with Lognormal and Gaussian mixture model respectively, as evidenced in Fig.3, compared to Gaussian fitting result shown in Fig.2. For feature *RI*, we can hardly find an appropriate distribution to fit it since its values are too concentrated around zero. And we adopt the method that divides its range into n equal parts first and then turns the discrete probability mass function into a continuous one to calculate its CDF.

For the distributions we adopted, we apply the Maximum Likelihood Estimation to obtain their model parameters. For the sake of brevity, we only present the parameter estimation steps for Gaussian Mixture Model (GMM). GMM is a probabilistic model that assumes all the data points to be generated from a linear superposition of Gaussian components which provides a richer class of density models than the single Gaussian.

Considering that we have n data points $x = \{x_1, x_2, ..., x_n\}$ in d-dimensional space (in our case, $d = 1$), the log likelihood with respect to a GMM can be denoted as:

$$log(p(x|\Theta)) = \sum_{i=1}^{n} log \sum_{k=1}^{K} \pi_k * \mathcal{N}(x_i|\mu_k, \Sigma_k) \tag{3}$$

where $\{\pi_k,\ s.t.\ \sum_{k=1}^{K} \pi_k = 1\}$ is the prior probability over the K Gaussian components, and (μ_k, Σ_k) are mean and standard deviation (model parameters) of the k^{th} Gaussian component. Then we use the Expectation Maximization (EM) algorithm to maximize the log likelihood to estimate the unknown parameters. Due to the page limits, we skip the details of EM process.

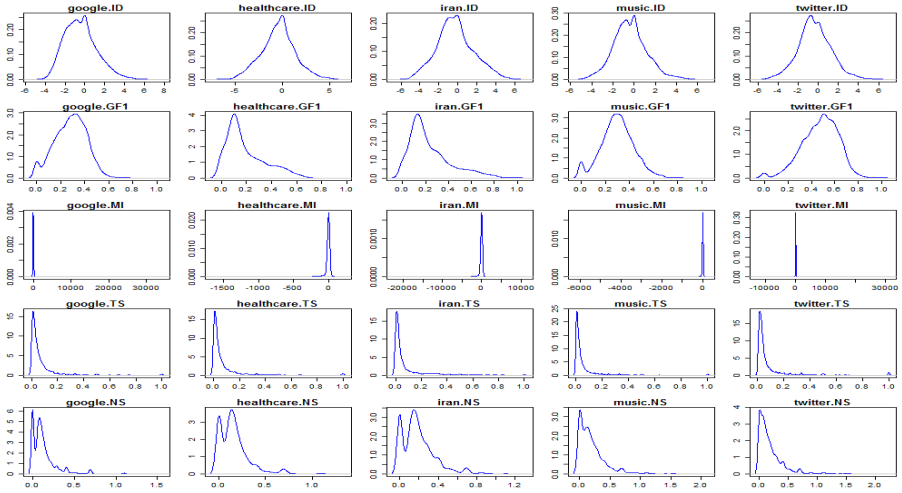

Fig. 1. Probability distributions of feature $ID, GF1, MI, TS, NS$ under different topics

Recall that we have proposed Authority Score(AS) based on the features' CDF in Sec.3.2. For feature f fitted by GMM, its CDF value of author x_i is defined as,

$$F_f(x_i) = \sum_{k=1}^{K} \int_{-\infty}^{x_i^f} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)\, dx \tag{4}$$

Figs.4-6 give the plots of the empirical densities of some features together with their fitting results of "google" dataset based on different models. We can observe that univariate Gaussian and Lognormal fitting have achieved good performance for feature *ID* and *TS* respectively. Fig.6 shows the GMM-based fitting and Gaussian-based fitting of feature NS. Obviously, GMM-based approach achieves more accurate fit than the univariate Gaussian-based one.

5 Experimental Evaluation

While preparing for the evaluation of our proposed ranking approaches, we manually labelled each tuple <query keyword, tweet> with a method which is similar to 3-point Likert scale, considering how relevant the tweet is to the query keyword and the amount of information it carries.

5.1 Evaluation Metric

To evaluate the ranking results, we adopt Normalized Discounted Cumulative Gain (NDCG) as the metric which is based on DCG [12]. NDCG measures the effectiveness of the ranking methods by penalizing the position from the result list with normalization. It is defined as:

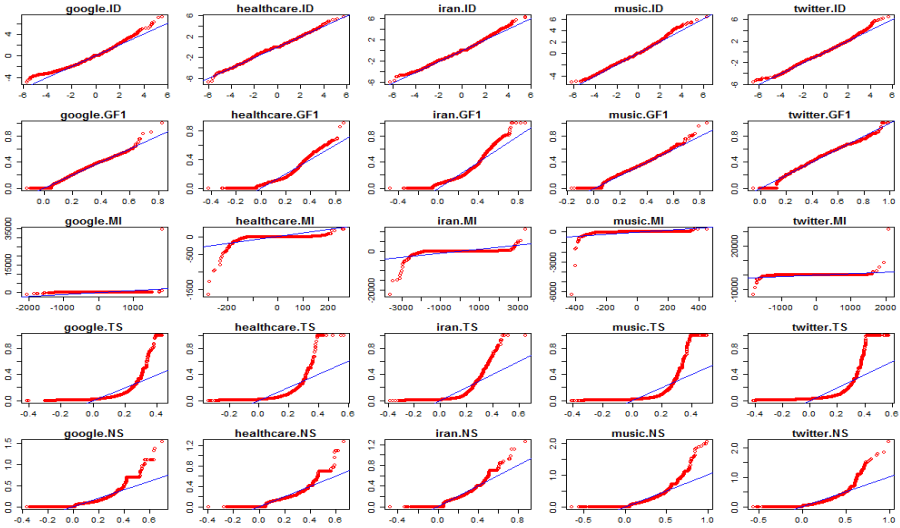

Fig. 2. Q-Q plots of features with Gaussian Fitting

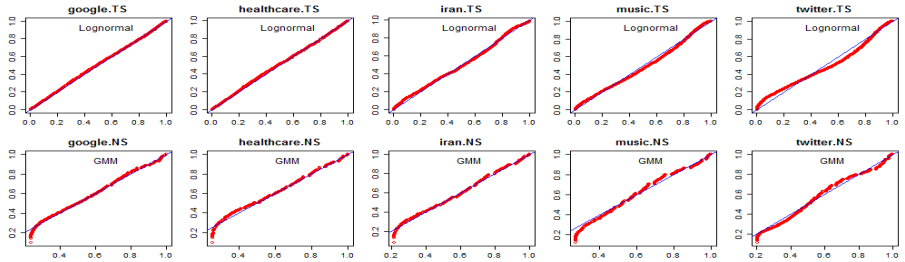

Fig. 3. Q-Q plots of features TS and NS with Lognormal and GMM Fitting

$$NDCG_n = Z_n \sum_{i=1}^{n} \frac{2^{G_i} - 1}{\log_2(i+1)} \tag{5}$$

where G_i is the label of i^{th} tweet in the final ranked list, and Z_n is a normalization factor, which is used to make the value of NDCG of the ideal list to be 1.

5.2 Evaluating Author Ranking Results

In Table 5, we present the top 10 authors of each dataset selected by *CDF-12* approach. With careful manual effort checking with Twitter, we find that the top-10 list is dominated by celebrities, popular bloggers and organizations. Besides, our method also discovers those authors who focus on certain areas and have a small number of followers (denoted in bold font).

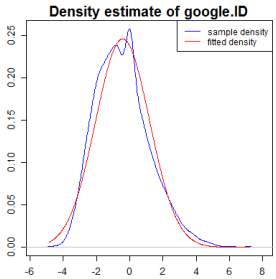

Fig. 4. Univariate Gaussian based fitting of feature ID for topic "google"

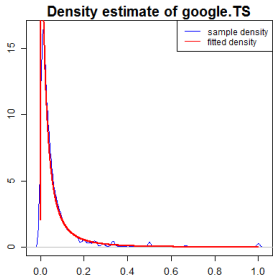

Fig. 5. Lognormal based fitting of feature TS for topic "google"

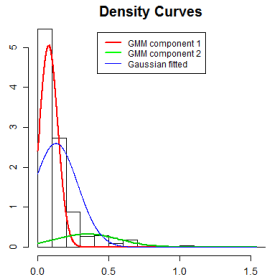

Fig. 6. GMM and Univariate Gaussian based fitting of feature NS for topic "google"

Table 5. Top 10 Authors From Query Datasets

google	**healthcare**	**iran**	**music**	**twitter**
programmableweb	healthcareintl	iranhr	showhype	dehboss
paulkbiba	**hcrepair**	jricole	nytimesmusic	chito1029
omarkattan	hcdmagazine	**newscomirancvrg**	variety_music	louer_voiture
morevisibility	**notmaxbaucus**	**jerusalemnews**	**im_musiclover**	twithority
wormreport	bnet_healthcare	jewishnews	digitalmusicnws	trueflashwear
followchromeos	healthnewsblogs	dailydish	musicfeeds	**twedir**
digg_technews	vcbh	haaretzonline	wemissmjblog	jointhetrain
webguild	presidentnews	guneyazerbaycan	411music	robbmontgomery
junlabao	chinahealthcare	**ltvx**	**radioriel**	youtubeprofits
redhotnews	ilgop	reuterskl	jobsinhiphop	thepodcast

5.3 Evaluating Microblog Ranking Results

We re-rank our dataset according to author Authority Score(AS) which is calculated by the methods described in Sec.3.2. In this section, we present the results of only two of the five topics ("google" and "healthcare") due to the page limit. Figs. 7 and 8 show the top-k ranking performance in terms of $NDCG_k$ (seen in Sec.5.1), where k varies from 5 to 1,000.

It is obvious that the *CDF_weighted* approach outperforms the others. According to Fig.7, we observe that the performance of *Conv_based* ranking method drops sharply with the increasing value of k in general. For another topic ("healthcare") that corresponds to Fig.8, the *Conv_based* method also underperforms our proposed approaches by a large margin. The phenomenon further demonstrates the effectiveness of the adoption of the topical authority in microblog ranking.

Figs. 7 and 8 show that *CDF-10* performs much better than its Gaussian version (i.e., *Gaus-10*), which verifies the benefit brought by the more accurate feature modeling in the CDF-based method. Also, we can observe that *CDF-12* outperforms *CDF-10* except for the top-5 case of the topic "google". This demonstrates the benefit brought by the two newly proposed authority features.

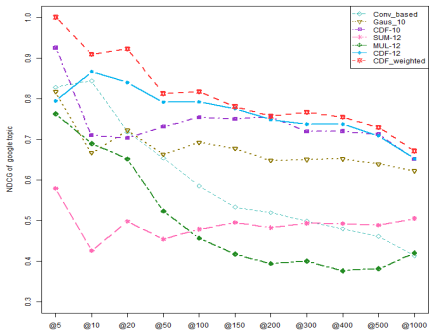

Fig. 7. A Plot of NDCG of the Top-k Results for the topic "google"

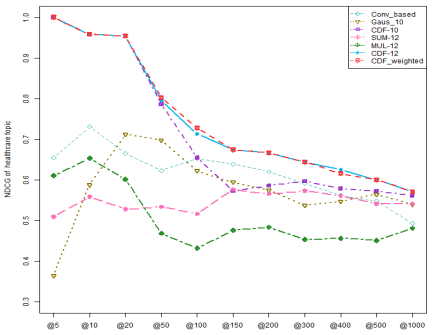

Fig. 8. A Plot of NDCG of the Top-k Results for the topic "healthcare"

Furthermore, the *CDF_weighted* approach further boosts the ranking quality by making an appropriate trade-off between the conventional popularity feature and the topical authority features. In addition, we adopt *SUM-based* and *MUL-based* approaches for benchmarking. And the *CDF-based* ones perform much better than the *non-CDF-based* approaches.

To summarize, our proposed *CDF_weighted* approach enhances the ranking performance significantly and perform best among all the proposed approaches. Quantitative analysis over the performance of the approaches show that *CDF_weighted* achieves more than 20% enhancement as compared to the conventional method as well as the Gaussian-based ranking method.

6 Conclusion and Future Work

In this paper, we first proposed to adopt the topical authority in microblog ranking and investigated to what extent the topical feature modeling can be optimized for boosting the performance of topical authority-based microblog ranking. Our attempts include extending the set of features considered and improving the feature modelling step. We applied the proposed extensions to a Twitter data set and compared the corresponding tweet ranking results with a number of existing methods for benchmarking. The experimental results validated the effectiveness of our proposed approaches and showed that the weighted version of CDF-based method outperforms other ones.

For future work, we will further investigate how the trade-off weight can be optimized for enhancing the microblog ranking quality. In addition, we are also interested in incorporating more features, e.g., content-based features, to further improve the microblog ranking quality.

References

1. Chen, Y., Amiri, H., Li, Z., Chua, T.-S.: Emerging topic detection for organizations from microblogs. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 43–52. ACM (2013)
2. Dalrymple, K.E., Shaw, B.R., Brossard, D.: Following the leader: Using opinion leaders in environmental strategic communication. Society & Natural Resources 26(12), 1438–1453 (2013)
3. Ravikumar, S., Balakrishnan, R., Kambhampati, S.: Ranking tweets considering trust and relevance. In: Proceedings of the Ninth International Workshop on Information Integration on the Web, p. 4. ACM (2012)
4. Duan, Y., Jiang, L., Qin, T., Zhou, M., Shum, H.-Y.: An empirical study on learning to rank of tweets. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 295–303. Association for Computational Linguistics (2010)
5. Pal, A., Counts, S.: Identifying topical authorities in microblogs. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, pp. 45–54. ACM (2011)
6. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web, pp. 591–600. ACM (2010)
7. Nagmoti, R., Teredesai, A., De Cock, M.: Ranking approaches for microblog search. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, pp. 153–157. IEEE (2010)
8. Cheng, F., Zhang, X., He, B., Luo, T., Wang, W.: A survey of learning to rank for real-time twitter search. In: Zu, Q., Hu, B., Elçi, A. (eds.) ICPCA 2012 and SWS 2012. LNCS, vol. 7719, pp. 150–164. Springer, Heidelberg (2013)
9. Weng, J., Lim, E.-P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 261–270. ACM (2010)
10. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web (1999)
11. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. The Journal of Machine Learning Research 3, 993–1022 (2003)
12. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of ir techniques. ACM Transactions on Information Systems (TOIS) 20(4), 422–446 (2002)