

Multidimensional Relevance in Microblog Search

Divi Galih Prasetyo Putri

Universit degli Studi di Milano-Bicocca, Dipartimento di Informatica Sistemistica e
Comunicazione, Viale Sarca 336, Milan, 20126, Italy
d.putri@campus.unimib.it

Abstract. In Information Retrieval, the assessment of relevance is affected by multiple criteria. This doctoral research aims to understand which relevance criteria or dimensions can affect the effectiveness of Microblog search systems. The research is based on the hypotheses that depending on the search intention, the user will consider different criteria and give different importance to the various relevance dimensions.

Keywords: Microblog Search · Multidimensional Relevance · Search Task.

1 Introduction

The concept of multidimensional relevance in Information Retrieval has been introduced by several works in the literature [5, 3]. Relevance is in fact a complex notion that relies on several dimensions since a user generally evaluates whether a document is relevant not only based on topicality. However, there has not been much research finalised to explore relevance dimensions in the context of social search. In [15], the authors have characterised the main differences between web search and Microblog search. The huge amount and the characteristics of user-generated content (UGC) constitute a challenge in the process of information seeking, as generally UGC in microblogs are very short and can contain some noise, irregular syntax, slang and misspelled words. Moreover, UGC is not always credible. Recently, some research works proposed a quality indicator of microblogs (Tweet) and exploited it as a relevance criterion in the retrieval model; examples of such indicators are credibility [17], informativeness [13], interestingness [1][11], and opinionatedness[6].

This doctoral research aims to explore the notion of multidimensional relevance in Microblog search, in particular with reference to specific search tasks, related to different search intentions. We postulate that each specific task might need an evaluation process that relies on different relevance dimensions (evaluation criteria or aspects). This implies that we can define which dimensions are important for a given task, and we can develop a method to aggregate the assessment of those dimensions so as to characterize the evaluation of relevance for

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). FDIA 2019, 17-18 July 2019, Milan, Italy.

the different tasks. The tasks we consider are supported by existing datasets. We have conducted a preliminary study where we define relevance dimensions that we have identified base on properties introduced in previous studies about Microblog Retrieval, namely credibility, informativeness, interestingness, and opinionatedness. First, We explore which aspects or dimensions can be useful to estimate relevance. Then, the combination of the dimensions and the importance of each dimension have to be tailored specifically to the various search task we aim to consider.

2 Relevance in Microblog Search

As previously outlined, the multi-dimensionality of relevance has been outlined in several papers, and as a matter of fact, search engines consider multiple dimensions to assess relevance. Among the most used relevance dimensions are Topicality, Novelty, and Reliability [16, 5]. However, in the literature, no study has addressed this issue in the context of Microblog search. During the first period of doctoral research, we have identified some relevance dimensions that could be potentially useful in Microblog search. Moreover, as it will be introduced in Section 3, we have identified different shared task related to Microblog search. We want to study the association between the relevance dimensions and the search tasks.

2.1 Informativeness

Informativeness has several definitions. It has been defined as “the extent to which a tweet meets the general interest of people involved with or tracking the event” [7]. Another study uses a similar definition of interestingness used in [2] as “specific information that people might care about” [4]. In [4], the authors perform human annotation to assess whether a document is relevant or not and whether it is informative or not. The result shows that informativeness and relevance are correlated and *retweeted* tweets tend to be more informative [4]. In [7], the authors eliminate non-informative tweets by adapting another work on tweet ranking and credibility. The result shows that credible users tend to post informative content and the proposed approach can help improve the ranking quality.

2.2 Credibility

Credibility is also used as a quality notion of a document. Previous research utilizes several Microblog features to estimate the credibility of a document [10]. The authors adopt the method proposed by another study on the credibility of a blog post and add several Microblog features such as repost, follower, and recency. They use a language modelling approach and use the credibility as the prior probability. Not only the credibility of the content but also the credibility

of the user is being evaluated to improve retrieval [7]. In this paper, the authors build a network consisting of retweet, mention, and reply to estimate user credibility score.

2.3 Interestingness

Microblogs can be used by many people to communicate with other people about personal matters. However, private and personal messages are not what people are looking for in Microblogs [11]. A tweet is considered useful to a user if the tweet is interesting to the user. Several studies tried to incorporate interestingness as a static quality measure of a tweet [1, 11, 14]. These studies all proposed classification technique (logistic regression and naive bayes) to calculate the interestingness as the quality of a document. A concept related to the interestingness of a tweet is the Retweet (RT). Some works calculated the probability of a tweet being retweeted as the interestingness score [1, 11]. Interestingness also able to improve the search task when it is used to remove tweet with interestingness score below the threshold but gives a lower performance in re-ranking.

2.4 Opinionatedness

Opinionatedness as a relevance dimension is related to the opinion retrieval task. The aim of this task is to retrieve relevant documents that contain user opinions on a topic while the opinion can be positive or negative. Previous research implemented a lexicon-based approach to estimate the opinionatedness of tweet [8]. The score is combined with social features using L2R to generate the ranking. Another study also estimates the opinionatedness of a document by using the average opinion score and add stylistic-based opinion score [6].

3 Research Question

This doctoral research is focused on relevance assessment in Microblog search. Following the concept of multidimensional relevance, we want to exploit several aspects of relevance that can improve the effectiveness of Microblog search. Specifically, we want to study the concept and behavior of multidimensional relevance related to the search task (Ex: Event Retrieval, Opinion Retrieval, etc.). The main research questions are as follows:

Research Question 1 (RQ1) : *What are the aspects of relevance dimension that important in different microblog search-tasks?*

Several aspects of relevance in Microblog search have been introduced in previous studies such as credibility, informativeness, and etc. Different search task in Microblog might have different search intent. The value used in assessing and ranking UGC can vary depending on the intention. For example, people search for user opinion on a topic in Microblog. In this case, the IR system will also consider opinionatedness as relevance dimension of the content. Only limited studies on specific search task have considered more than topical aspect. This

work is based on the hypothesis that each of the search tasks tends to favor different aspect of relevance. This research questions can extend our knowledge on the application of multidimensional relevance in microblog search.

Research Question 2 (RQ2): *How to model the dimensional importance and how to combine it with another dimension?*

The retrieval status value of a single retrieved document is calculated based on several values representing the considered dimensions. In different search-task, the user might prioritize or give different weight to some dimension. In [5], the authors proposed the use of Multi-Criteria Decision Making to perform prioritized aggregation of several relevance criteria. This concept supports the idea of having different importance in different task or domain. By extending the concept of prioritized aggregation, a model has to be able to change dynamically when different importance is assigned to each relevance dimension.

4 Methodology

In the planned experiments, data from Twitter will be used. In order to examine different relevance dimensions in different search tasks, we will consider the following retrieval tasks, with associated datasets:

1. Disaster Related Retrieval: SMERP 2017 on Text Retrieval (Task 1)¹
2. Opinion Retrieval Data : Previous Work on Opinion Retrieval[8]
3. Cultural Event and Argument Retrieval: CLEF MC2 2017 on Microblog Retrieval (Task 2)²

To understand the impact of each relevance dimension on a specific search task, we make use of a retrieval system using only topical relevance as the baseline system. Then for each of the additional relevance dimension we will consider, the score will be combined with the topicality score by using linear combination.

4.1 Preliminary Result

In this paper we report the preliminary experiment related to two search tasks. We have employed the following datasets: Disaster Related Retrieval- SMERP 2017 on Text Retrieval (Level-1 and Level-2) and Opinion Retrieval Data. In this preliminary experiment we consider only one additional relevance dimension, i.e informativeness. The retrieval system is developed using Apache Lucene. For the baseline, we use Language Model with Dirichlet Smoothing as the retrieval model. The RSV of the baseline is the topicality score RSV_t that will be combined with informativeness score RSV_i by means of a linear combination. We estimate the final RSV using $RSV = \alpha RSV_t + (1 - \alpha)RSV_i$. We tune the weight α in steps of 0.1. Runs were evaluated using *trec-eval*, an evaluation tool provided by TREC.

¹ <https://www.computing.dcu.ie/~dganguly/smerp2017/>

² <https://mc2.talne.eu/>

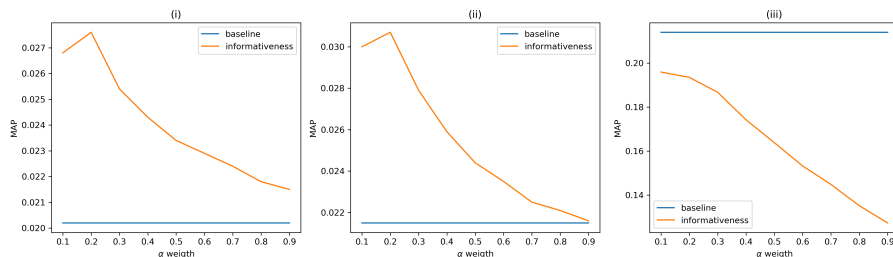


Fig. 1. Preliminary result from the experiment in different dataset. (i) Disaster-related data level-1 (ii) Disaster-related data level-2 (iii) Opinion Retrieval data

To estimate the score of Informativeness, we exploit an annotated dataset that publicly available [12]. This dataset labeled based on the information source and information type (informativeness). In this dataset, they define a tweet as informative if it helps to understand the situation. We build a model using logistic regression to predict the probability and the informativeness value of a new tweet. The features used in this study is based on the previous study [9]. We use 5-fold cross validation in the experiment and the model is able to get the precision of 73,73%, recall of 85,19%, and f-measure of 78,99%. Fig. 1 illustrates the results of our experiment using Language Model as the baseline. The x-axis represents the weight of the topicality value and y-axis shows the MAP score. We use this figure to highlight the changes in MAP score with respect to the weight. We can see in Fig.1 that adding informativeness value can improve the performance of disaster-related tweet retrieval. Both in SMERP level-1 and SMERP level-2 dataset, the combination outperforms the baseline when the weight is 0.1 and 0.2. However, combining informativeness does not improve the result of the opinion retrieval system. We can infer that informativeness has an impact on disaster-related retrieval task but not on opinion retrieval task.

5 Conclusion and Future Work

In this work, we focus on exploring the use of multidimensional relevance in different tasks of Microblog search. We combine topicality with another relevance dimension and evaluate the system using two different types of dataset (Disaster-related retrieval task and opinion retrieval task) to see the impact in each search task. In future work, we plan to implement other aspects of relevance such as credibility, interestingness, and opinionatedness, and also to evaluate the system using other dataset related to other tasks in Microblog search.

References

1. Alhadi, A.C., Gottron, T., Kunegis, J., Naveed, N.: Livetweet: Microblog retrieval based on interestingness and an adaptation of the vector space model. In: TREC

- (2011)
2. Alonso, O., Carson, C., Gerster, D., Ji, X., Nabar, S.U.: Detecting uninteresting content in text streams. In: SIGIR Crowdsourcing for Search Evaluation Workshop (2010)
 3. Borlund, P.: The concept of relevance in ir. *Journal of the American Society for information Science and Technology* **54**(10), 913–925 (2003)
 4. Choi, J., Croft, W.B., Kim, J.Y.: Quality models for microblog retrieval. In: Proceedings of the 21st ACM international conference on Information and knowledge management. pp. 1834–1838. ACM (2012)
 5. da Costa Pereira, C., Dragoni, M., Pasi, G.: Multidimensional relevance: Prioritized aggregation in a personalized information retrieval setting. *Information processing & management* **48**(2), 340–357 (2012)
 6. Giachanou, A., Harvey, M., Crestani, F.: Topic-specific stylistic variations for opinion retrieval on twitter. In: European Conference on Information Retrieval. pp. 466–478. Springer (2016)
 7. Huang, H., Zubiaga, A., Ji, H., Deng, H., Wang, D., Le, H., Abdelzaher, T., Han, J., Leung, A., Hancock, J., et al.: Tweet ranking based on heterogeneous networks. Proceedings of COLING 2012 pp. 1239–1256 (2012)
 8. Luo, Z., Osborne, M., Wang, T.: An effective approach to tweets opinion retrieval. *World Wide Web* **18**(3), 545–566 (2015)
 9. Mahata, D., Talburt, J.R., Singh, V.K.: From chirps to whistles: discovering event-specific informative content from twitter. In: Proceedings of the ACM web science conference. p. 17. ACM (2015)
 10. Massoudi, K., Tsagkias, M., De Rijke, M., Weerkamp, W.: Incorporating query expansion and quality indicators in searching microblog posts. In: European Conference on Information Retrieval. pp. 362–367. Springer (2011)
 11. Naveed, N., Gottron, T., Kunegis, J., Alhadi, A.C.: Searching microblogs: coping with sparsity and document quality. In: Proceedings of the 20th ACM international conference on Information and knowledge management. pp. 183–188. ACM (2011)
 12. Olteanu, A., Vieweg, S., Castillo, C.: What to expect when the unexpected happens: Social media communications across crises. In: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing. pp. 994–1009. ACM (2015)
 13. Surdeanu, M., Ciaramita, M., Zaragoza, H.: Learning to rank answers to non-factoid questions from web collections. *Computational linguistics* **37**(2), 351–383 (2011)
 14. Tao, K., Abel, F., Hauff, C., Houben, G.J.: Twinder: a search engine for twitter streams. In: International Conference on Web Engineering. pp. 153–168. Springer (2012)
 15. Teevan, J., Ramage, D., Morris, M.R.: # twittersearch: a comparison of microblog search and web search. In: Proceedings of the fourth ACM international conference on Web search and data mining. pp. 35–44. ACM (2011)
 16. Xu, Y., Chen, Z.: Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology* **57**(7), 961–973 (2006)
 17. Zhao, C., Xu, L., Huang, H.: Exploiting user activities for answer ranking in q&a forums. In: International Conference on Collaborative Computing: Networking, Applications and Worksharing. pp. 693–703. Springer (2017)