Using Emotion to Diversify Document Rankings

Y. Moshfeghi, G. Zuccon and J.M. Jose
{yashar, guido,jj}@dcs.gla.ac.uk

School of Computing Science
University of Glasgow, Scotland, UK

Abstract. The aim of this paper is to investigate the role of emotion features in diversifying document rankings to improve the effectiveness of Information Retrieval (IR) systems. For this purpose, two approaches are proposed to consider emotion features for diversification, and they are empirically tested on the TREC 678 Interactive Track collection. The results show that emotion features are capable of enhancing retrieval effectiveness.

1 Introduction

Emotion is considered to be an important factor influencing overall human behaviour, including rational tasks such as reasoning, decision making, communication and interaction. Although emotion is subjective, it is presented in some objectively deducible ways in written documents [1]. News and user-generated content such as blogs, reviews, and tweets contain emotionally rich data and several studies have attempted to automatically extract these features from such data [1]. The use of emotion features has been shown to improve retrieval system effectiveness in collaborative search [2]. However, the effectiveness of emotion features when diversifying document rankings has yet to be studied.

Given a query, IR systems generate rankings according to the relevance of documents. Diversity in the ranking results has been shown to be useful in improving the effectiveness of IR systems. This is because diversity avoids redundancy, resolves ambiguity and effectively addresses users' information needs [3].

Diversity has been addressed through mathematical models [4] and through the use of external evidence [5]. We propose to use emotional features to enhance the diversity of the retrieved results. We believe that emotion features serve as beneficial information for diversifying document rankings. This is motivated by the fact that IR systems strive to gather conceptual information about a document through an indexing process, e.g., by representing documents as a bag of words. However, such a process ignores the fact that documents are not only vehicles for transmitting information, but also convey meanings and emotion. Here we focus on emotion and propose that diversifying document rankings based on emotion features allows us to better overcome this issue. We posit that relevant documents belonging to different subtopics may differ with respect to their conveyed emotion. For example, documents relevant to subtopic "diseases entering UK" of topic 352i ("British Chunnel impacts") imply different emotion than documents relevant to "increased tourism anywhere on British island": we thus expect that diversifying document rankings based on emotion will yield improvements in performance.

2 Approach

In the following, we outline the diversification approaches used in this work and discuss how emotion features are blended together with estimations of document relevance. Then the emotion extraction technique is explained.

2.1 Diversifying Document Rankings

In order to diversify document rankings, we adopt Maximal Marginal Relevance (MMR) [4] as it is an effective and popular approach. Let $sim(d, q)$ denote a measure of similarity between document d and query q; this can be regarded as a measure of relevance of d to q. Also let $esim(e(d), e(d'))$ represent the similarity between the emotion vector representations (see Section 2.2) of documents d and d'. We consider the situation where $|R|$ documents have been ranked, and the ranking function considers which document has to be ranked next. Following MMR, the next document to be ranked (i.e., d^*) is selected such that:

$$d^* = \arg\max\,[\lambda sim(d, q) - (1-\lambda)\max_{d' \in R} esim(e(d), e(d'))]$$

where λ is a parameter that controls the impact of emotion similarity on the selection of document d^*: if $\lambda = 1$, emotion similarity has no impact on the selection of documents; while if $\lambda = 0$, emotion similarity is the only criterion used for ranking documents.

We further generalise the MMR approach such that the similarity between the candidate document and the query is interpolated with the *average* emotion similarity between the candidate document and those that have been ranked at previous positions. Thus, under the average interpolation approach (AVG-INT), d^* is ranked at rank position $|R| + 1$ if

$$d^* = \arg\max\,[\lambda sim(d, q) - (1-\lambda)\sum_{d' \in R} \frac{1}{|R|} esim(e(d), e(d'))]$$

In contrast to MMR, the AVG-INT approach considers the average similarity between a candidate document and documents ranked in the previous $|R|$ ranks.

Several similarity functions can be used for computing $esim(e(d), e(d'))$. We test our ranking strategies using the cosine similarity and Pearson's correlation as similarity function to measure document relationships with respect to emotion. Other measures can be used (e.g., KL divergence, L1 norm, etc.): we plan to investigate the impact of different functions on empirical results in future works.

2.2 Construction of Emotion Vectors

There are multiple views of what emotion is and how it should be represented. Ortony, Clore and Collins regard emotion as consequences of events, actions of agents, and aspects of objects. They introduced the OCC model which specifies 22 emotion types and two cognitive states[1] [6], in contrast to sentiment analysis which categorises text into binary classes (i.e. positive/negative), in turn

[1] The emotion categories are: joy , distress, happy-for, sorry-for, resentment, gloating, hope, fear, satisfaction, fears-confirmed, relief, disappointment, shock, surprise, pride, shame, admiration, reproach, gratification, remorse, gratitude, anger and the two cognitive states are love and hate [6].

Table 1. α-nDCG values of Language Model (LM), MMR with text features (MMR(t)), MMR with emotion features (MMR(e)) and AVG-INT with emotion features (AVG-INT(e)) are reported and percentages of improvement over LM are presented in brackets. The best performing approach at each rank is highlighted in bold. Due to space constraints, for MMR(t) we only report results when re-ranking the top 20 documents: other settings obtain results that exhibit similar trends. Performance of AVG-INT(t) and MMR(t) are similar, and we therefore report the latter.

		LM	***MMR(t)***	***MMR(e)***			***AVG-INT(e)***		
			Top 20	Top 20	Top 50	Top 100	Top 20	Top 50	Top 100
α-nDCG	@5	0.520	0.554 (+7%)	**0.568 (+9%)**	0.555 (+7%)	0.545 (+5%)	0.561 (+8%)	0.559 (+8%)	0.539 (+4%)
	@10	0.532	0.559 (+5%)	0.560 (+5%)	**0.567 (+6%)**	0.551 (+4%)	0.554 (+4%)	0.555 (+4%)	0.547 (+3%)
	@20	0.545	0.556 (+2%)	0.556 (+2%)	0.564 (+4%)	0.546 (+0%)	0.555 (+2%)	**0.565 (+4%)**	0.559 (+3%)

providing potentially more information for diversification. Here we follow the OCC model because it has been considered as a superior view by the cognitive psychology community. Based on this model, Shaikh et al. [1] developed a state-of-the-art text-based emotion extraction system. In this work, we use our own implementation of Shaikh et al. approach which is shown to be more accurate than other state-of-the-art emotion extraction systems.

Our emotion extraction method is sentence-based and makes a binary decision about the presence of each emotion for a given sentence. Since the emotion extractor is rule-based there is no need for training the model. In order to extract emotions from a retrieved document, we consider the following procedure. Let S denote a set of sentences associated to a document d. For each sentence s in S, we construct a 24 dimension vector where each component can take value 1 if the emotion is present in the sentence and 0 otherwise. Then, in order to represent the emotion contained in d, we give equal importance to each sentence by averaging the emotion vectors of the sentences in d.

3 Experiment and Results

Implementation. Documents were indexed using the Lemur toolkit (`http://www.lemurproject.org/`). Standard stop-word removal and stemming techniques were applied at indexing time to both documents and query topics. The top n documents (with n = 20, 50, 100, 200) were retrieved in answer to each query using a unigram language model with Dirichlet smoothing, where the smoothing parameter was set according to standard values (i.e., $\mu = 2000$). The ranking of the top n retrieved documents formed the baseline (identified as LM in Table 1) against which we compared their re-ranked version according to the approaches presented in Section 2.1, where $sim(d, q)$ was estimated according to the scores returned by LM and $esim(e(d), e(d'))$ was computed by the cosine similarity or Pearson's correlation between the emotion vectors representing the documents. We also tested MMR and AVG-INT considering only text features (i.e., *MMR(t)* and *AVG-INT(t)*): these are based upon the diversification approaches presented in Section 2.1, but use term vector representations of documents instead.

Experiment Settings. We tested our approaches on the TREC 678 Interactive Track collection containing 20 topics which also have been used for diversity task evaluation [7]. Ranking approaches were evaluated according to α-nDCG [3] at different rank positions. Results were similar both when using the cosine similarity and the Pearson's correlation: we only report the former due to space limits. For all the diversification approaches, we varied λ in the range $[0, 1]$ with granularity of 0.05. We report the results obtained selecting parameter values that maximise α-NDCG@10 for each query.

Results. The results[2] reported in Table 1 show that considering emotion features improves retrieval effectiveness. Emotion-based approaches display better performances than LM. We found that emotion-based diversification obtained substantial gains (about 20%) for more than 30% of queries over LM. For example, for topic 446i, "tourists, violence", diversifying rankings based on emotion, provides substantial increments at all levels of diversification (i.e. for all λ values). Emotion-based approaches also provide better performance than the *MMR(t)* approach, which employs text features. Whilst the average effectiveness gains are marginal in this preliminary study, there is a case for using emotion features to diversifying document rankings.

4 Conclusions

In this paper we investigated the effectiveness of using emotion features when diversifying document rankings. We adapted existing models (i.e. MMR and AVG-INT) to exploit emotional features. The results are encouraging and show improvement when including emotion features for re-ranking retrieved results. This work is a foundation towards future research that employs emotion features to improve IR systems. Future work will consider combining both text and emotion features building more elaborate diversity models.

References

1. Shaikh, M.A.M., Prendinger, H., Ishizuka, M.: A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text. Affective Information Processing (2009)
2. Moshfeghi, Y., Jose, J.M.: Role of emotional features in collaborative recommendation. In: ECIR'11. (2011)
3. Clarke, C.L., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: SIGIR'08. (2008)
4. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: SIGIR'98. (1998)
5. Yin, X., Huang, J.X., Zhou, X., Li, Z.: A survival modeling approach to biomedical search result diversification using wikipedia. In: SIGIR'10. (2010)
6. Ortony, A., Clore, G., Collins, A.: The cognitive structure of emotions. Cambridge University Press (1990)
7. Zhai, C.X., Cohen, W.W., Lafferty, J.: Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In: SIGIR'03. (2003)
8. van Rijsbergen, C.J.: Information Retrieval, 2nd Ed. Butterworth (1979)

[2] Since the topic set is small (i.e., 20 queries), performing significance tests would not be appropriate [8, pages 178–180]. Moreover we do not report results obtained for $n = 200$ for space limits.