Learning to Rank for Joy

Claudia Orellana-Rodriguez, Wolfgang Nejdl
L3S Research Center
Hannover, Germany
{orellana, nejdl}@L3S.de

Ernesto Diaz-Aviles*
IBM Research
Dublin, Ireland
e.diaz-aviles@ie.ibm.com

Ismail Sengor Altingovde
Middle East Technical University
Ankara, Turkey
altingovde@ceng.metu.edu.tr

ABSTRACT

User-generated content is a growing source of valuable information and its analysis can lead to a better understanding of the users needs and trends. In this paper, we leverage user feedback about YouTube videos for the task of affective video ranking. To this end, we follow a learning to rank approach, which allows us to compare the performance of different sets of features when the ranking task goes beyond mere relevance and requires an affective understanding of the videos. Our results show that, while basic video features, such as title and tags, lead to effective rankings in an affective-less setup, they do not perform as good when dealing with an affective ranking task.

Categories and Subject Descriptors : H3.3 [**Information Search and Retrieval**]; K.4 [**Computer and Society**]
General Terms: Human Factors, Experimentation
Keywords: Sentiment Analysis; Social Media Analytics; YouTube

1. INTRODUCTION

Users' information needs are complex and depend on their context (e.g. time of the day, mood, location). In this paper, we leverage social feedback for the task of affective video ranking. According to [1], while video titles and tags, so called *basic features*, are effective for ranking; social feedback (e.g., views, likes, dislikes, comments) is useful to further improve the retrieval quality. Here, we take one step forward and suggest that the social feedback and its conveyed emotions and polarity can be even more effective in scenarios where the information need is less general. For instance, if a user is looking for videos about "panda bears", it is possible that basic features will suffice in the search of relevant videos, but if the user is looking for "*happy* videos of panda bears", basic features can prove not to be as effective, because now, besides checking whether a video is relevant for the query of interest, there is the inherent need to explore if it is also a happy video, which may not be captured by the basic features alone. In this work, we seek to address the following research question: *Can features extracted from social feedback perform better than basic features when the retrieval task is within an affective context?* We go beyond the relevance ranking of videos and consider targeting a scenario on which the user's information need is highly related to her individual context and affective needs.

*Ernesto Diaz-Aviles was a senior research scientist at the L3S Research Center when contributing to this work.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2744-2/14/04.
http://dx.doi.org/10.1145/2567948.2576961.

2. APPROACH

We seek to leverage social feedback for the task of *learning to rank for joy.* For this, we consider a standard learning to rank framework for Information Retrieval [1, 3]. For each video, we construct a vector consisting of three sets of features: *basic, social* and *sentic. Basic features* are those created by the video uploader (video title and tags). Once a video is uploaded, YouTube users can watch, like/dislike, favorite, and comment; we call *social features* the ones extracted from these interactions [1]. For *sentic features*, we first build a profile consisting of all the comments collected for the corresponding video, then, we perform part-of-speech tagging on each comment to extract the nouns and adjectives, and finally, we use a term-based matching technique to associate each term with emotion and polarity values.[1]

Affective Representation of Videos. In our study, we use EmoLex, a large set of words described in [4] and annotated according to Plutchik's psychoevolutionary theory [5]. Plutchik considered the existence of eight *primary emotions* forming four opposite pairs, joy-sadness, anger-fear, trust-disgust, and anticipation-surprise. Based on [2], we define the sentic vector, s_v, for video $v \in V$ as follows: Let T_p be the set of terms extracted from the video's profile p, and T_m the set of all terms in EmoLex annotated with emotion m, where $m \in M$; $M := \{$*joy, sadness, anger, fear, trust, disgust, anticipation, surprise*$\}$. Then, the m^{th} dimension of sentic vector $s_v \in \mathbb{R}^{|M|}$ is given by $s_v[m] := \sum_{t \in T_p} \mathcal{I}_m(t)$, where $\mathcal{I}_m(t)$ is an indicator function that outputs 1 if the term $t \in T_p$ is associated to emotion m, and 0 otherwise. Finally, we normalize s_v to produce a probability vector $\hat{s_v} = \frac{s_v}{N_M}$, where N_M is a normalization constant corresponding to the total number of terms $t \in T_p$ associated to an emotion. Similarly, we compute the polarity tuple (*positive, negative*) and append it to the sentic vector. In this vector, each emotion and polarity is represented as a real value denoting their degree of presence in the videos, this gives us insights on how *emotional* is a video and what is the relation among the emotions themselves. To each video, a label is assigned representing

[1]Complete list of features: `http://www.l3s.de/~orellana/info.pdf`

its relevance with respect to a given query; however, we need each label to also represent the presence of the emotion m_1 in a *higher degree* than the presence of its opposite emotion m_2, e.g., if the relevance label of a video v with respect to query q is 1 (v is relevant to q), *and* according to human annotations, the emotion m_1 is associated to v in a higher degree than m_2, then the video relevance label will be 1. On the contrary, if v is relevant to q but the emotion m_1 is absent, or there is a more dominant presence of m_2, then the video relevance label will be 0. For those cases on which v is irrelevant to the query, the relevance label is 0. The main idea behind our labeling is to train models which capture the *affective context* C of the users, that is, the subset of emotions of interest when performing a video search. This would be useful in scenarios where a user is eager to watch a video that is not only relevant to a given query, but will also satisfy her current need of experiencing an emotion m_1, that is, within her affective context $C = \{m_1\}$.

3. EXPERIMENTAL EVALUATION

Dataset. For a set of around 7,000 queries, we collected from YouTube the top-300 result videos, corresponding metadata and up to 10,000 most recent comments posted for each video. We also created video uploaders profiles, consisting of the number of uploaded videos, total views for these videos, and subscribers. Our first goal is to define the ground truth for our task of *learning to rank for joy*, that is, each video should have a label that represents its relevance with respect to a given query as well as the presence of the emotion m_1, which is in the affective context, $m_1 \in C$, in a *higher degree* than the presence of its opposite emotion m_2. To this end, we proceed in two stages: *Stage 1: Elicit relevance judgments* and *Stage 2: Elicit affective judgments*. In *Stage 1*, we first need to know how relevant is each video v to its corresponding query q. We perform this labeling as described in [1]. In *Stage 2*, we carry out a second annotation phase, we ask 4 users to annotate all the videos in our collection (after Stage 1) according to the emotions they experience while watching the videos. We presented a video at a time, and asked the user *What emotion would you associate this video with?* The possible answers were: (i) *joy*, (ii) *sadness*, (iii) *is neutral*, and (iv) *the video cannot be accessed*. With the resulting annotations, and after defining the affective context as $C = \{m_1\}$, where m_1 = joy, towards which we will focus the ranking task, we define the videos' relevance labels as follows: if the video is relevant *and* associated to m_1 more than to m_2 = sadness, then the label is 1, otherwise is 0. That is, we require the video to be relevant *and* associated to the emotion *joy*.

In our experiments, we use RankSVM [3] to learn the ranking function because it exhibits the highest ranking performance for NDCG@10, when using only basic features for video retrieval [1]. We learn *three different ranking functions*: **Basic.** Uses only **basic** features for training and corresponds to our baseline. **Social and Sentic.** Trained using **social and sentic** features. **All.** Uses **all** the features (basic, social, sentic) for training. Using grid search, we found that setting the value of SVM cost constant to 10 ($c = 10$), led to the best results in all the models evaluated in our experiments.

Results. For a set of 50 queries, we conducted 10 rounds of cross validation experiments, each one consisted of 45 queries for training and 5 queries for testing. As *evaluation metrics* we used P@10, MAP, NDCG at 5 and 10, as well as Mean NDCG. We report the average ranking performance over the ten rounds, for each of the models. Figure 1 shows the results of our experiments. In this sentic ranking scenario, the ranking function trained on social and sentic features clearly outperforms the one trained using basic features, in particular, for NDCG@5, NDCG@10 and Mean-NDCG, which shows that if we exploit social feedback, we can successfully address a specific sentic task. The ranking model trained on the set of *all* features, while outperforming the baseline, does not reach the performance of the model corresponding to the *social and sentic*. Overall, we can see that *basic* features, which show a good performance retrieving relevant videos, are not necessarily good for the task of affective learning to rank.

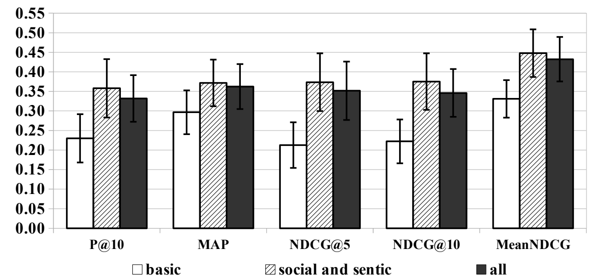

Figure 1: Average Ranking Performance.

4. CONCLUSIONS AND FUTURE WORK

Our goal in this paper was to explore if features extracted from social feedback perform better than basic features (video title and tags) when the retrieval task requires an affective understanding of YouTube videos. To this end, we went beyond a general relevance ranking model for video retrieval, which ignores the emotions associated to the videos, towards an affective setting. Our results show that the ranking functions learned based on social and sentic features, outperform the ones learned based on basic features. We focus on a particular pair of emotions, joy and sadness, which can be useful in many scenarios, e.g., when users are looking for videos that will help them change their mood from sad to happy. However, our approach can easily adapt to a broader affective context that includes more or different emotions. As future work, we plan to leverage our affective approach for recommender systems, where the emotions extracted from social feedback will enhance personalization engines with better sentic capabilities.

Acknowledgments. This work is partially supported by EU FP7 Project CUBRIK (contract no. 287704) and The Scientific and Technical Research Council of Turkey (TUBITAK) under the grant no. 113E065. Ismail Sengor Altingovde acknowledges the Yahoo! Faculty Research and Engagement Program.

5. REFERENCES

[1] S. Chelaru, C. Orellana-Rodriguez, and I. S. Altingovde. How useful is social feedback for learning to rank youtube videos? *World Wide Web*, pages 1–29, In press.

[2] E. Diaz-Aviles, C. Orellana-Rodriguez, and W. Nejdl. Taking the pulse of political emotions in latin america based on social web streams. In *LA-WEB*, pages 40 –47, oct. 2012.

[3] T. Joachims. Optimizing search engines using clickthrough data. KDD '02, pages 133–142, NY, USA, 2002. ACM.

[4] S. M. Mohammad and P. D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 2011.

[5] R. Plutchik. *A General Psychoevolutionary Theory of Emotion*, pages 3–33. Academic press, New York, 1980.