

# Reputation Analysis with a Ranked Sentiment-Lexicon

Filipa Peleja  
Dep. Computer Science  
Universidade Nova Lisboa  
Portugal  
filipapeleja@gmail.com

João Santos  
Dep. Computer Science  
Universidade Nova Lisboa  
Portugal  
jme.santos@campus.fct.unl.pt

João Magalhães  
Dep. Computer Science  
Universidade Nova Lisboa  
Portugal  
jm.magalhaes@fct.unl.pt

## ABSTRACT

Reputation analysis is naturally linked to a sentiment analysis task of the targeted entities. This analysis leverages on a sentiment lexicon that includes general sentiment words and domain specific jargon. However, in most cases target entities are themselves part of the sentiment lexicon, creating a loop from which it is difficult to infer an entity reputation. Sometimes, the entity became a reference in the domain and is vastly cited as an example of a highly reputable entity. For example, in the movies domain it is not uncommon to see reviews citing *Batman* or *Anthony Hopkins* as esteemed references. In this paper we describe an unsupervised method for performing a simultaneous-analysis of the reputation of multiple named-entities. Our method jointly extracts named entities reputation and a domain specific sentiment lexicon. The objective is two-fold: (1) named-entities are naturally ranked by our method and (2) we can build a reputation graph of the domain's named entities. This framework has immediate applications in terms of visualization or search by reputation.

## Categories and Subject Descriptors

- *Information systems~Information extraction.*
- *Information systems~Sentiment analysis*

## Keywords

Reputation analysis, sentiment lexicons, LDA.

## 1. INTRODUCTION

When searching for opinions, an IR system must deal with the domain named-entities and with specific sentiment lexicons. In some cases, these named-entities (e.g., the actors or film titles in the movies domain), are so important that they become a synonym of high-quality (or low-quality). They become references in their area. It is not uncommon to find reviews where multiple citations to actors or movies occur. Thus, it becomes fundamental that an IR system identifies these named-entities and infers its reputation.

Reputation analysis for entities has been a topic of recent research, since the new trend of micro blogging has propelled an accumulation of data that reflects public opinion on various subjects. Go et al. [3] used various machine learning algorithms (Naïve Bayes, Maximum Entropy and SVM) to classify the overall sentiment of Twitter messages towards specific keywords, representing various distinct entities likes movies, famous people, locations and companies. Later, Chen et al. [1] proposed the extraction of sentiment polarity on tweets towards movies and people as a constrained optimization problem. Chen et al.'s

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGIR '14, July 6–11, 2014, Gold Coast, Queensland, Australia.  
Copyright © 2014 ACM 978-1-4503-2257-7/14/07...\$15.00.

approach used a lexicon containing both formal and slang words to accommodate Twitter vocabulary, built by collecting words from dictionaries such as SentiWordNet [2] and Urban Dictionary ([www.urbandictionary.com](http://www.urbandictionary.com)). Here, however, we argue that static-lexicons are too coarse-grained. As a consequence fail to capture relevant sentiment words that target numerous entities. Also, static-lexicons assign fixed weights to sentiment words. In this paper we capture the relevant sentiment words and weight them according to their sentiment relevance. The proposed method detects sentiment words fluctuations through an LDA generative model from users' sentences extracted from reviews of different ratings. Moreover, the method also weights the overall sentiment level associated to an entity, corresponding to the reputation of that entity.

## 2. RELATED WORK

A sentiment analysis challenge begins with the nature of text opinions: opinions are inherently subjective and written in natural language, which is also an ambiguous way of representing knowledge. One of the earliest approaches that contributed on identifying subjective sentences, hence opinionated sentences, was proposed by Hatzivassiloglou et al. [4]. In this work was used a seed of adjectives with a set of linguistic constraints to capture adjectives. Later, Turney et al. [11, 12] used a seed of adjectives from Hatzivassiloglou et al. [4] and the General Inquirer dictionary in a sentiment classification task.

Recent work on capturing relevant sentiment words has focused at identifying sentiment words by exploring the usage of slang or domain-specific sentiment words [1]. For instance, Urban Dictionary (UD) and Twittrat's ([twitter.com/twittrat](http://twitter.com/twittrat)) are dictionaries that aim at capturing sentiment words that more traditional dictionaries fail to capture (e.g. Multi-perspective Question Answering (MPQA) [14], General Inquirer ([www.wjh.harvard.edu/~inquirer/](http://www.wjh.harvard.edu/~inquirer/)) and SentiWordNet [2]). In the present work we stress on the need to capture relevant sentiment words which are strongly associated to a given context. Hence it is proposed a ranked based LDA to capture the most relevant sentiment words.

Reputation analysis have focused not only for summarizing the overall reputation but also to predict the reputation of other instances or events [5, 8]. Joshi et al. [5] explored the popularity of old movies among online critic reviews to predict opening weekend revenues for new movies, by comparing the similarity in metadata of old highly rated movies with new ones. More recently, Oghina et al. [8] predicted the IMDb movie ratings by analyzing their popularity on social media, namely Youtube and Twitter. Oghina et al. used textual tweets, comments and likes on videos related to specific movies to predict their reputation, then translated into a numeric rating scale from 1 to 10. In our proposal we aim at identifying the most relevant sentiment words that characterize the reputation of a given entity. Hence, movie

reviews were extracted from the IMDb ([www.imdb.com](http://www.imdb.com)) movie database and, the actors and characters (e.g. *batman*) names which represent the entities. The potential marketing usefulness of reputation analysis has led research on last years to focus extensively on monitoring and profiling relevant emerging topics for market brands and organizations on Twitter, such as Apple and Windows [6, 7, 10, 13]. Martín et al. [7] explored different approaches for identifying emerging topics that are relevant for an organization’s reputation, such as representing each tweet as a set of Wikipedia entries that are related to it and extending the LDA generative model to capture relevant topics on tweets. More recently, Spina et al. [10] tested a variety of different techniques with the same goals as Martín et al and others[13]. For obtaining aggregated sentiment regarding an organization Spina generated domain-specific semantic graphs to expand a sentiment lexicon. Also, for monitoring relevant topics, tested approaches included filtering relevant or irrelevant tweets to a certain company by discovering filter keywords and usage of wikified and LDA model in a similar way to Martín’s et al. approach.

### 3. RANKING BY REPUTATION

The problem we address in this paper aims at measuring an entity reputation by creating a sentiment lexicon and weighting its’ relevance towards the entity. The proposed sentiment lexicon is created with user sentences without human supervision. To identify the sentiment words it is proposed a multilevel generative model of users’ reviews. Therefore we propose a generative probabilistic model that ties words to different sentiment relevance levels and evaluate within each subjective sentence the sentiment word proximity to an entity.

**Problem definition.** Consider a set of  $M$  sentences  $\mathcal{D} = \{d_1, \dots, d_M\}$  containing user opinions towards a given movie. Each review  $d_i$  is represented by a tuple  $(w_i, s_i)$ , where  $w_i = (w_{i,1}, \dots, w_{i,N})$  is a vector of  $N$  word counts and  $s_i \in \{1, \dots, R\}$  is the associated sentiment value quantifying the user opinion about the product (it corresponds to the user rating). An entity is any movie, actor, character, director, etc of the domain. Entities are usually mentioned in reviews and are part of the domain taxonomy. Our goal is twofold: first we compute a fine-grain lexicon of sentiment words that best captures the varying level of user satisfaction, and second, we determine the reputation of an entity by measuring its impact in the domain, i.e., the most relevant sentiment words associated to an entity.

#### 3.1 Framework

The reputation analysis framework is divided into different steps, Figure 1 illustrates the process. First, to best capture complex sentiment expressions, we compute bigrams with a maximum of 3 words between every word pair. Stop words removal and lemmatization are also part of the initial step. Second, we compute the sentiment lexicons after removing the entities from the corpus to determine the influence of these entities in the domain sentiment characteristics. Finally, to ascertain the reputation of named entities, we propose two ways: the number of citations and the context in which they are cited.

#### 3.2 Ranked sentiment lexicon

LDA is a generative model that explores word co-occurrences at document-level and at the level of  $K$  latent topics. It samples a word distribution from a prior Dirichlet distribution for each latent topic. The probability of a sequence of words and its hidden topics

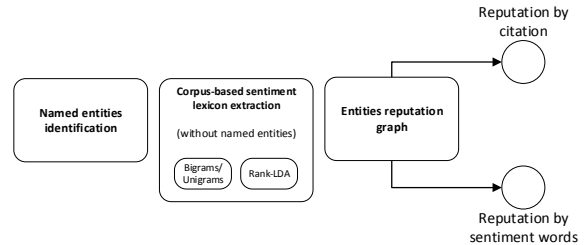


Figure 1: The Rank-LDA graphical model.

is given by  $p(w, z) = \int p(\theta) \cdot \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n) d\theta$  where  $\theta$  is the random parameter of a multinomial over topics. With ranked LDA the goal is to identify which words are used to express a sentiment. Figure 2 presents the graphical model of the proposed Rank-LDA method. At its core, the Rank-LDA links the latent topics to the sentiment relevance of each sentence. For each relevance there will be a set of hidden topics that will be activated.

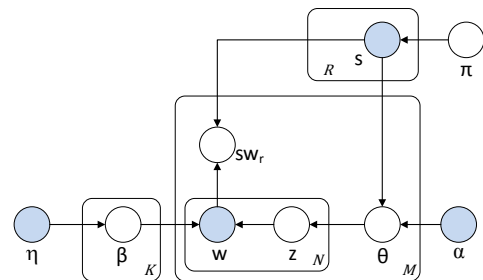


Figure 2: The Rank-LDA graphical model.

Rank-LDA is structured as follows:  $\beta$  is the per-corpus topic Dirichlet( $\cdot | \eta$ ) distribution,  $\theta$  is the per-sentence topic Dirichlet( $\cdot | \alpha$ ) distribution,  $z$  is the per-word topic assignment following a Multinomial( $\cdot | \theta^{(d)}$ ) distribution, and  $w$  correspond to the set of words observed on each sentence. Finally,  $s_i \in \{1, \dots, R\}$  is the per-sentence sentiment relevance and  $sw$  is the per-word random variable corresponding to its sentiment distributions across the different sentiment levels of relevance. The random variables  $\alpha, \eta$  and  $\pi$  are the distribution priors. The sentiment word distributions are given by the density distribution

$$p(sw | s) = \int p(\theta) \cdot \prod_{n=1}^N p(z_n | \theta, s) p(sw | z_n) d\theta + \tau \quad (1)$$

where we compute the marginal distribution of a word given a sentiment level, over the  $K$  latent topics of the Rank-LDA model. The variable  $\tau$  is a smoothing parameter that we set to 0.01.

The sentiment word distribution function can also be used to rank words by its positive/negative weight or to calculate a word’s cross-sentiment occurrences. To achieve such conversion is through a normalization function such as

$$RLDA(sw) = \frac{p(sw|s=i) - p(sw|s=j)}{\min(p(sw|s=i), p(sw|s=j))} \quad (2)$$

where  $p(sw|s=i)$  and  $p(sw|s=j)$  contain the word  $w$  relevance values in rating  $i$  and  $j$ .

#### 3.3 Entity reputation graph

IMDb movie reviews are in a rating scale from 1 to 10, thus, Table I presents the words distributions  $p$  (see equation 1)

obtained with the lower and higher rating, 1 and 10. The words distributions for the opposite ratings is clearly depicted for the words *horror*, *garbage* and *excellent*, as for the words *television* and *pilot* the weights do not differ.

The word *batman* represents a word highly related to a specific set of movies and intuitively would not be observed as a sentiment word. Nonetheless, in Table I the word *batman* depicts a high relevance value. Reasoning on this observation, in reviews from rating 1 and 10 the word *batman* is mentioned with a frequency of 1,286 and 5,107, in 212 and 251 different movies respectively. For example,

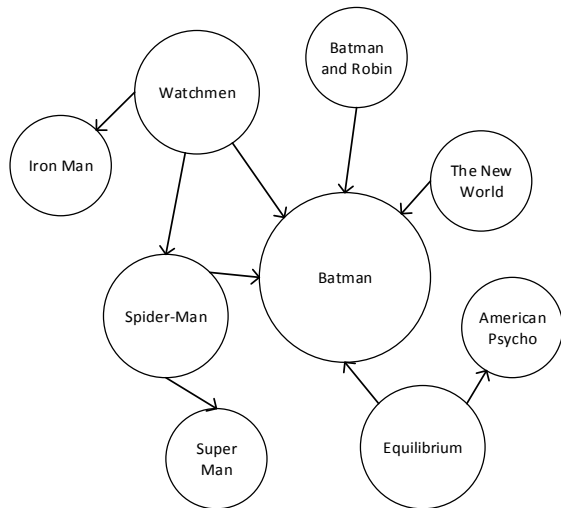
1. (...) it took a non-batman movie to finally make a decent-cool looking-realistic Batman costume? (Watchmen);
2. Go watch crash, Capote, walk the line, sideways even batman begins for modern Hollywood, this is dreadful. (The New World);

hence, words like *batman* enclose a sentiment weight. The proposed ranked LDA sentiment lexicon captures these relevant sentiment words.

**Table I: RLDA word sentiment relevance.**

Word	$p_1$	$p_{10}$	RLDA
horror	0.069	0.016	3.286
garbage	0.036	0.010	2.585
excellent	0.010	0.034	2.425
batman	0.010	0.132	12.192
television	0.010	0.012	0.188
pilot	0.010	0.012	0.188

Considering *batman* as a sentiment word the reputation of entities as *watchmen*, *equilibrium* and others is weighted with the sentiment word weight of the entity *batman*, as in this case, is also viewed as relevant sentiment word (Figure 3).



**Figure 3: Reputation given by the sentiment word *batman* towards other movies (entities).**

## 4. EVALUATION

### 4.1 Data

**IMDb-Extracted:** This dataset contains 1,007,926 million movie reviews, corresponding to a total of 10,651,052 million sentences. Reviews are rated in a scale of 1 to 10. For evaluation purposes the dataset was evenly split into three disjoint subsets (A, B and C). Following Pang et al. [9] approach the subset A was used to build a subjective classifier, thus, sentences from movie plots are labeled as objective and sentences from users reviews as subjective. Moreover, to perform a balanced subjective classification the number of subjective sentences was reduced to match the number of objective sentences. For the subset B and C, 693,349 and 1,449,546 were classified as objective sentences respectively. The ranked LDA lexicon is built with subjective sentences from split B. The subjective sentences from split C are used for evaluation purposes. Table II presents the detailed information of the IMDb-Extracted. This dataset is available <https://novasearch.org/datasets/>.

**Table II: Detailed information of IMDb-Extracted.**

Split	#reviews	#sentences	#subjective sentences
A	335,975	167,074	83,537
B	335,950	2,981,996	2,288,647
C	335,976	3,953,522	2,503,976

### 4.2 Experiments and results

To evaluate the quality of the ranked sentiment lexicon for entity reputation two crowdsourcing ([www.crowdfunder.com](http://www.crowdfunder.com)) tasks are performed. First, given a sentence it is asked for the annotator to judge if a specified sentiment word is relevant to characterize the entity reputation (Table IV); and secondly, given 5 sentiment words it is asked for the annotator to judge if the entity described by those words has a positive, negative or neutral reputation. Therefore the first task evaluates if the captured sentiment words are relevant to measure the entity reputation and the second task evaluates the proposed method ability to correctly weight the sentiment word polarity. For the first task it was used 3,000 sentences, the sentences were obtained randomly from the split C. And, for the second task 3,000 combinations of sentiment words in which roughly one third were bigrams. For both experiments, it was created a gold standard by selecting the units where workers had an agreement of 75% or more, resulting in 2036 gold units for the first task and 943 gold units for the second task. Table III compares the obtained annotations with methods based on our lexicon. Task REL refers to the first task, POL-UNI and POL-BI refers to the second task for sentiment words obtained from unigrams and bigrams, respectively. The obtained results for the relevance task suggests that a very high percentage of the captured sentiment words with our lexicon are relevant for reputation analysis of entities. In parallel, results for the polarity task shows that the associated weights for the sentiment words perform well on standard binary polarity evaluation.

**Table III: Crowdsourcing for Entities Reputation measured with ranked RLDA sentiment words.**

Task	Precision	Recall	F-1
REL	84.5%	94.0%	89.0%
POL-UNI	80.2%	85.2%	82.6%
POL-BI	81.4%	82.0%	81.7%

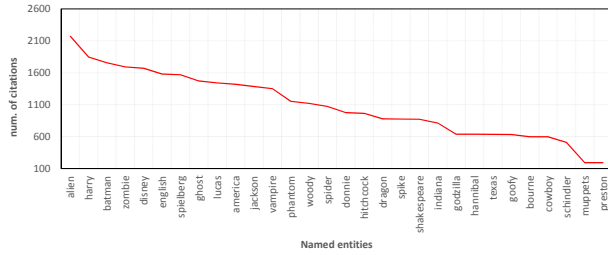


Figure 4: Reputation given by other entities (e.g. through citation).

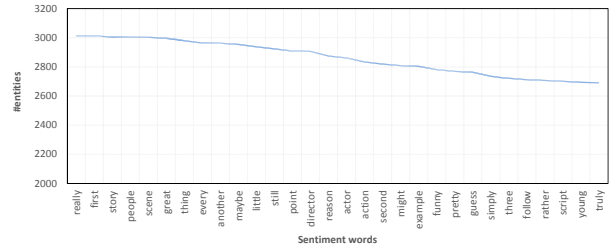


Figure 5: Sentiment words used as named entities reputation qualifiers.

Table IV: Evaluation sentences for crowdsourcing.

Sentence	Entity	rlda
Having seen a few <u>Hitchcock</u> movies in my day,I <u>cannot believe</u> Zemeckis thought this script qualified.	Hitchcock	cannot##believe
<u>Seagal</u> is the only man standing between blah and blah and blah de <u>blah blah</u> .	Seagal	blah##blah
If there was an <u>excellent</u> <u>Batman</u> , this is the real deal.	Batman	excellent
<u>Anthony Hopkins</u> did a <u>great</u> job as Diego de la Vega/Zorro.	Anthony Hopkins	great

### 4.3 Discussion

In Figure 4 is examined the number of different entities associated to a sentiment word. For instance, the sentiment word *batman* is related to 1,557 entities and *indiana* to 812 entities. In Figure 5 is shown the entities association to domain related sentiment words (e.g. *drama*, *trailer* and *oscar*). Moreover, Figure 6 presents the top positive and negative sentiment words. This illustrates how rank rlda captures both general and domain specific sentiment words. Also, characters and actor names are frequently used as positive, or negative, reference (Figure 3), thus, the relevance of using these sentiment words when measuring an entity reputation.

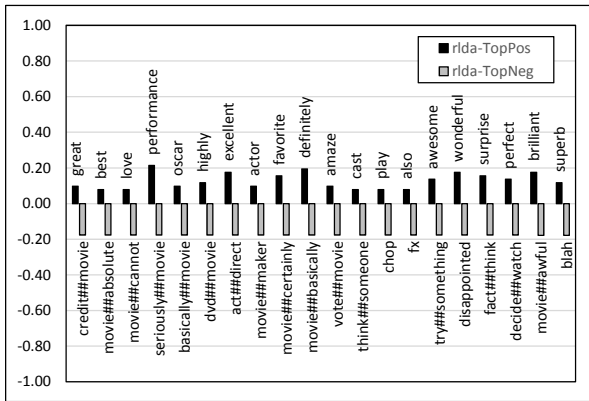


Figure 6: Top positive and negative rlda sentiment words.

### 5. CONCLUSIONS

In this paper we have proposed a method to measure entities reputation. This was performed by selecting a set of sentiment words that best represent the opinions targeting a given entity. More, specifically, by exploring the words distributions in a generative ranked based LDA model, it is built a sentiment lexicon. The obtained lexicon was able to capture domain specific sentiment words that traditional static sentiment lexicons fail to capture or infer a generic sentiment weight. However, we stress that sentiment words that are usually overlooked as relevant sentiment words by traditional methods (e.g. *batman* or *hannibal*) enclose a relevant sentiment weight and we have shown that for several entities these sentiment words are frequently used. Furthermore, we have successfully evaluated the proposed approach in two crowdsourcing tasks.

**Acknowledgements.** This paper was partially funded by the Portuguese Science Foundation (projects UTA-Est/MAI/0010/2009 and PTDC/EIA-EIA/111518/2009).

### 6. REFERENCES

- Chen, L. et al. Extracting Diverse Sentiment Expressions with Target-Dependent Polarity from Twitter. *ICWSM*,2012.
- Esuli, A. and Sebastiani, F. Sentiwordnet: A publicly available lexical resource for opinion mining. *In LREC '06*,2006.
- Go, A. et al. Twitter Sentiment Classification using Distant Supervision. Technical report, Stanford, 2009.
- Hatzivassiloglou, V. and McKeown, K.R.. Predicting the semantic orientation of adjectives. *ACL*,1997.
- Joshi, M. et al. 2010. Movie Reviews and Revenues: An Experiment in Text Regression. *HLT '10: NAACL*, 2010.
- Martin-Wanton, T. et al. An unsupervised transfer learning approach to discover topics for online reputation management. *CIKM'13*,2013.
- Martin-Wanton, T. et al. UNED at RepLab 2012: Monitoring Task. *CLEF*, 2012.
- Oghina, A. et al.. Predicting IMDB Movie Ratings Using Social Media. *ECIR'2012*,2012.
- Pang, B. and Lee, L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *ACL*,2005.
- Spina, D. et al. UNED Online Reputation Monitoring Team at RepLab 2013. *CLEF*, 2013.
- Turney, P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *ACL '02*,2002.
- Turney, P.D. and Littman, M.L. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, *TOIS*,2003.
- Villena-Román, J. et al. DAEDALUS at RepLab 2012: Polarity Classification and Filtering on Twitter Data. *CLEF '2012*.
- Wiebe, J. and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*. *ACH*, 2005.