Time-aware online reputation analysis

Peetz, M.-H.

Publication date
2015

Document Version
Final published version

9
Conclusions

This thesis introduced new algorithms to online reputation analysis focusing on the inherent temporal aspects of the underlying social media data. We studied how social media analysts perform online reputation analysis in Chapter 4. The findings motivated the research in the following chapters where we proceeded with the development of algorithms to make Online Reputation Analysis (ORA) easier and the data more accessible. In Chapter 5 introduced algorithms to estimate reputation polarity based on the findings of Chapter 4. We then moved to the finding and filtering of documents. Chapter 6 and Chapter 7 provided algorithms to find documents (in particular social media documents) based on salient time periods and recency, respectively. Chapter 8 combined the temporal ideas from Chapter 6 and Chapter 7 to filtering ever changing social media data with respect to an entity. Motivated by requirements from Chapter 4, we kept the analysts in the loop by using active learning methods.

Below we first answer the research questions introduced in Chapter 1. We continue with the final section where we share our view on the future work this thesis could influence.

9.1 Answers to Research Questions

In Chapter 4 we observed social media analysts annotate online media for the reputation of a company. We asked:

RQ1.1 What are the *procedures* of (social media) analysts in the analysis and annotation of reputation polarity?

We found that the procedures vary over different media types. For media types like Google result pages and Youtube videos, reading and processing the information is most important. We also found that the most important steps in the procedures are determining the topic, author, and reach of the tweet and that finding tweets should be automated. The analysts also use a lot of background information and filtering.

RQ1.2 On a per tweet level, what are the indicators that (social media) analysts use to annotate the tweet's impact on the reputation of a company?

We found that the indicators used by analysts to determine the reputation polarity of a tweet are based on the authority of the author of the tweet. This authority can be topical and based on both, online and offline data. The analysts also base their decision on the reach of a tweet.

We showed that successfully estimating the author's authority and determining who is exposed to the media in question is the key for an automatic estimation of reputation polarity.

In Chapter 5 we used some of the indicators found in Chapter 4 as features for automatically estimating reputation polarity on tweets for a specific entity. We made use of three feature groups based on the sender (the author of the tweet), the message itself, and the receiver. The latter tried to capture the reach in so far as it looked at *who* receives the message. Training of the reputation models was based on three settings: entity-independent, entity-dependent, and domain-dependent. We asked:

RQ2.1 For the task of estimating reputation polarity, can we improve the effectiveness of baseline sentiment classifiers by adding additional information based on the sender, message, and receiver communication model?

We found that adding additional features based on the communication model: sender, message, and receiver, we could reliably improve the results from the literature and the baseline sentiment classifiers.

RQ2.2 For the task of estimating reputation polarity, how do different groups of features perform when trained on entity-(in)dependent or domain-dependent training sets?

In general, the entity-dependent training scenario led to higher effectiveness than the entity-independent or domain-dependent scenario. Using features modeling the sender performed better in the domain-dependent scenario.

RQ2.3 What is the added value of features in terms of effectiveness in the task of estimating reputation polarity?

Most added value came from textual features of the messages. The impact feature, a feature based on the impact a message has on its recipients, was helpful in combination with features based on the message itself.

All in all, we found an effective approach to estimate the impact of a tweet on the reputation of an entity using very few, but focussed, training samples.

In Chapter 4 we also identified that finding and filtering tweets should be automated. A strong retrieval and filtering process improves the estimation of reputation polarity [230]. In the second part of the thesis, we analysed how to retrieve and filter documents using temporal knowledge. We used this temporal knowledge in Chapter 6 where we presented models that identify the temporal information need of queries. We asked:

RQ3.1 Are documents occurring within bursts more likely to be relevant than those outside of bursts?

and

RQ3.2 Can documents within bursts contribute more useful terms for query modeling than documents selected for relevance models?

Documents are not more relevant, but they were found to be different and they therefore introduce the right amount of variety into the topic models. The terms used for query modeling lead to significant improvements of effectiveness over non-temporal baselines.

RQ3.3 What is the impact on the retrieval effectiveness when we use a query model that rewards documents closer to the center of the bursts?

While the blog datasets had narrow and noisy bursts and feature better effectiveness using documents closer to the center, the opposite was the case for news data.

RQ3.4 Does the number of pseudo-relevant documents used for burst detection matter and how many documents should be considered for sampling terms? How many terms should each burst contribute?

As long as the number of pseudo-relevant documents was high enough to avoid spurious bursts, the number of documents did not matter: the results were stable. Selecting few documents from the bursts sampled more useful terms for query modeling. The effectiveness was stable with respect to the number of terms contributed.

RQ3.5 Is retrieval effectiveness influenced by query-independent factors, such as the quality of a document contained in the burst or size of a burst?

We did not find normalisation to have an effect on the retrieval effectiveness.

The overall findings here were that sampling terms from bursts in pseudo-relevant documents raises effectiveness on three temporal datasets.

We also identified a second type of temporal information need in this thesis: recency. In Chapter 8 we explained how recency is related to our memory and retention models from the psychology literature. We used the retention models as temporal priors in a retrieval setting and asked:

RQ4.1 Does a prior based on exponential decay outperform other priors using cognitive retention functions with respect to effectiveness?

While the effectiveness of exponential decay as a prior on news data was found to be on par with the best performing priors, using exponential decay on microblog data disappoints in particular with respect to precision.

RQ4.2 In how far do the proposed recency priors meet requirements, such as efficiency, performance, and plausibility?

The recency priors were found to meet the requirements to different degrees. The Weibull function follows the requirements best and compromises well between them. The exponential decay used in the literature was not found to follow the requirements adequately.

We showed that using priors with a cognitive motivation did indeed perform better on data with a recency information need, in particular on microblog data.

In Chapter 4 social media analysts identified the filtering process as a task that should be automated. For entity filtering, we proposed a semi-automated approach based on active learning for a streaming scenario. We asked:

RQ5.1 For the entity filtering task, does margin sampling improve effectiveness over random sampling, i.e., is it a strong baseline?

Margin sampling improved the effectiveness over random sampling significantly. In general, it proved to be a successful strategy to improve effectiveness with few annotations.

Since we were following a streaming scenario, it stood to reason that temporal approaches improve margin sampling. Based on work in Chapter 7 we incorporated recency priors into margin sampling and asked:

RQ5.2 For the entity filtering task, does a sampling based on recency priors and margin sampling together, outperform margin sampling with respect to F-score?

We found that for some entities margin sampling with temporal priors works better than vanilla margin sampling. However, using the temporal prior also harmed the effectiveness for other entities.

Chapter 6 showed that burst-based approaches work well for retrieving social media documents which lead to looking at temporal reranking in Chapter 8. Our first reranking approach reranked based on bursts in the candidate set, the second approach reranked based on arrival date of the tweet. We asked:

RQ5.3 For the entity filtering task, does temporal reranking of margin sampled results based on bursts or recency, outperform margin sampling with respect to F-score?

Margin sampling was found to be the best approach to use for very weak and very strong models, but burst based reranking outperformed vanilla margin sampling for strong-enough models.

We found that active learning is a feasible approach improve effectiveness of entity filtering models in a streaming scenario. The flavour of margin sampling to use, however, strongly depends on the entity.

In this thesis we discussed different aspects of online reputation analysis and when appropriate, merged them with temporal approaches. Much of the experimental work is motivated by our user study of reputation analysts in from Chapter 4: be it the use of indicators that take the reach of a tweet into account for the estimation of reputation polarity, or be it an improvement of the retrieval of social media documents to avoid expensive manual annotation work, or keeping the reputation analysts in the loop with active learning approaches. Temporal approaches played a key role in each of the technical chapters. We proposed temporal streaming scenarios in Chapter 5 and Chapter 8. Chapter 6 took into account salient time periods (bursts) when sampling terms for documents, while Chapter 7 investigated temporal priors.

Chapter 4 points out that there are very few steps in the annotation process that reputation analysts deem possible to automate. This thesis aimed to bridge the divide by building a trust relationship between algorithm designers and reputation analysts. In Chapter 8 we saw that using the knowledge of reputation analysts leads to very stable results for filtering tweets for specific entities. To achieve this, reputation analysts must be willing to help algorithms and algorithm designers by recording and sharing their (background) knowledge—and do so with confidence, since their knowledge is invaluable. Finally, reputation analysts can also take home that relying on simple sentiment approaches will not be as successful in estimating the reputation polarity as well as dedicated reputation polarity approaches.

To conclude, this thesis contributed several new approaches to make online reputation analysis on social media more accessible and less tedious for reputation analysts.

9.2 Future Research Directions

This thesis has resulted in several lessons for online reputation analysis and temporal information retrieval. In the following we lay out future research directions, in particular on how to study annotation behavior, interaction with social media experts, temporal information retrieval and online reputation analysis.

Study annotation behavior To the best of our knowledge no study looked at the actual behavior and intentions of experts doing annotations of online media. While there is an abundance of data-oriented approaches for automatic classification, future work should also follow expert-oriented approaches similar to our approach in Chapter 4. Understanding how the human mind solves classification problems can give a unique viewpoint towards automatic classification. This requires different, non-invasive, methodological approaches tailored to studying experts, in particular to understand how and when experts make use of background information.

Interaction with (social media) experts In Chapter 7 we used active learning to address an ORA task. This gave a peek into the potential of active learning for ORA. Social media analysts insist on accuracy and distrust automatic approaches (see Chapter 4). Future work respecting this insistence should go into two directions. For one, we should develop active learning algorithms for all kinds of ORA problems like topic modeling, monitoring and subsequent alerting or filtering. Algorithms should not only make use of highly accurate expert input, but also interactions: a system that can ask questions like "is this really a burst?" or "did a new topic evolve around the entity, and are those fitting keywords for the topics?" and can successfully incorporate this additional information in the model will outperform current passive approaches.

The second direction of future work is the actual interaction with experts. Communication between system and expert done right is just as important as the improved algorithms. Finally, the active learning approaches should not be limited to ORA but can easily be generalised to digital humanities. Again, experts in digital humanities are willing to spend valuable time to interact with the system [174] and we should make good use of this.

Temporal information retrieval An obvious extension to our work in temporal information retrieval would combine recency-based and burst-based approaches. One way to achieve this is to classify queries into different temporal information needs, similar to [125]. Temporal information needs do not necessarily only have to be based on recent or salient time periods, but also on traumatic time periods (9/11) or culturally pervasive time periods (1969): events with an emotional semantic frame. Using more fundamental understandings from psychology to model information needs can lead to personalised temporal models for search.

Additionally, recency priors can also function as an indicator for reliability of benchmarks: we saw that for the Blog06 benchmark (see Chapter 3), the queries and the temporal information need were broader the later the queries were created with respect to the collection. The recency priors, being based on how people remember, can normalise for this behavior in evaluation.

Finally, most studies using temporal algorithms are based on static document collections and queries [62, 75, 144]. Future work should bring temporal information retrieval benchmarks with a varying set of temporal information needs underlying queries. Integrating active learning and its evaluation with those benchmarks, will lead to a rise of strong temporal active learning algorithms.

Online reputation analysis The algorithms presented in this thesis to support ORA are by no means perfect and often function more as a starting point. Future work should focus on implementing stable, easy to understand by lay(wo)men, and effective algorithms for the estimation of reputation polarity and filtering. More work should be put into implementing the indicators we found in Chapter 4, in particular how we can model offline authorities with only online data. Somehow, this thesis touches on all aspects of the *backbone* of ORA: we retrieve, we filter, and we estimate the reputation polarity. We would love to see similar temporal ideas applied to semantically more complex aspects such as topic modelling and alerting, and author profiling. Finally, we hope for an interface that combines and incorporates all algorithms presented in this thesis to ease online reputation management.