Managing Diversity through Social Media

Tadej Štajner, Inna Novalija

Artificial Intelligence Laboratory, Jožef Stefan Institute, Ljubljana, Slovenia
{tadej.stajner, inna.koval}@ijs.si

Abstract. This paper describes the approach for interactive diversity analysis based on social media data. The presented approach is applicable for a number of common value management activities, such as reputation management, competition analysis, market analysis, sentiment analysis and similar tasks, related to measuring community activity. The design of the developed diversity analysis component allows achieving the quickest possible turnaround time required to train a multi-class model for a particular task. The evaluation is performed in the telecommunication domain on the dataset of Twitter posts.

Keywords: Natural language processing, Diversity analysis, Sentiment analysis, Social networks.

1 Introduction

This paper describes the approach for diversity analysis based on social media data. **Diversity analysis** represents a collection of related tasks that segments posts and users into meaningful groups, based on differing viewpoints that depend on the topic, location and language, providing sentiment and opinion modeling given these groups [1]. Diversity analysis can be applied for a number of brand management activities, such as **reputation management**, **competition analysis**, **market analysis**, and **sentiment analysis**. These tasks have a goal of establishing, monitoring and improving the status of subject (for instance, a product or a particular organization) in the community.

We motivate our work with the case study on customer relationship management through social media, especially various forms of characterization of the retrieved relevant posts. Since social media is consistently becoming an important communication channel for brand and customer relationship management, new technical issues are coming up when we try to apply previously known techniques to communication through social media. Imagine that our user, working in customer relationship management or marketing, is trying to measure the impact and opinions on a new campaign. In this scenario, we emphasize the use case of making sense of the social media responses by having the user retrieve and segment tweets relevant to the campaign in question. More specifically, the scenario discussed in this paper is the following: out of a stream of social media posts, assign classes to the posts that are relevant to the brand management task at hand, where the task can be unknown in advance. Furthermore, by classifying these posts we are annotating them with concepts from a lightweight ontology [6], a use case that has already successfully demonstrated using active learning techniques for a related task, namely ontology construction.

In this research we present an approach that can be used to train one of several ad-hoc model-based queries that can appear in such an environment: language, topic or sentiment classification. Since brand management requires constant maintenance, it's difficult to foresee in advance what exact problem we are tackling. Therefore, most such problems need to be solved ad-hoc. In order for that to be practical, they also need to be trained quickly. The key contribution of this research is the demonstrated ability to interactively train such a model in order to answer these types of questions, providing quick analyst response. For the purpose of this paper, we frame the problems as multi-class classification tasks, where we also allow for unlabeled examples. There are several relevant problems that fit this framework. While some of these problems can be defined in advance, such as coarse topic classification, language detection or general sentiment detection, some can be predictable, for instance the launch of a new product, a new competitive product or a community-driven action. While we can initially approach this problem using information retrieval techniques, formulating a query may be difficult and ambiguous, which is why we can also consider this problem in terms of multi-class classification.

We focus on the telecommunication domain, the tweets about different competitors – telecommunication companies are labeled with different topics (*Vodafone, O2, Simobil*), and we wish to evaluate the overall sentiment in social media with respect to these competitors. The sentiment classification contributes to the activities of reputation management and sentiment analysis. However, due to topic-sensitivity of sentiment analysis, the there is a need to produce specialized sentiment models.

The design of the developed diversity analysis software components allows achieving the quickest possible turnaround time required to train a multi-class model for a particular task. Some tasks originate from a particular marketing or public relations project on a particular new offering, whose public opinion needs to be monitored, which call for ad-hoc definition of tasks. We also need the changes are immediately apparent in the result set.

This paper also describes this approach in detail, the respective design decisions and demonstrates an example software component. The approach and software evaluation is performed in the telecommunication domain on a dataset of Twitter posts.

2 Related Work

There are a number of approaches discussing diversity analysis. In their work Hasan et al. [1] explained the baselines of the knowledge diversity model and presented the glossary of diversity relevant terms. They present three motivating business scenarios for formalization of the knowledge diversity model – mining diversity in Wikipedia, diversified news and customer relationship management.

The diversity in opinions has been discussed by Kim and Hovy [2] in their paper on opinion sentiment. They presented a system that, given a topic, automatically identifies people with opinions (and opinion sentiments) on that topic.

Turney [3] developed an unsupervised algorithm for classifying reviews as recommended or not recommended based on the average semantic orientation of the phrases in the reviews.

At the same time, social media have been widely used for opinion determination. For instance, Bizău et al. [4] presented an approach to creating domain dependent opinion vocabulary based on Twitter comment, acknowledging the diversity that heavily affects that particular problem.

In our research, social media data is used as a corpus for common brand management activities - reputation management, competition analysis, market analysis and sentiment analysis. We are not only combining different brand management activities in one tool – we are also minimizing the turnaround time required to train a multi-class model for a particular task, while letting the user to track the quality of the obtained model in order to estimate whether more training is necessary,

3 Interactive Diversity Analysis

To accommodate the functional requirements for brand positioning in social media, we have developed an approach that enables interactive ad-hoc training of tasks such as topic, language and sentiment detection as semi-supervised multi-class classification tasks. The developed software component is intended to be the analytics back-end for a social media analytics dashboard.

The situation we are dealing with is providing social media analytics to answer ad-hoc model-based queries on social media streams. The main problem is that the model may not be known in advance, so it needs to be trained quickly.

For instance, for a telecommunication domain, the model should answer the following queries:

What kinds of sentiment do people express on O2's iPhone offer?
Which operator has the most social media talk about Android devices?

In order to train a specified model, we use a machine learning technique called a**ctive learning**, a technique that intelligently picks unlabelled examples which need to be labeled in order to minimize training time.

At the same time, we encounter a number of the technical design constraints:

- We have multiple classes to classify in;
- We can't realistically expect the user to label all examples;
- We should make use of feature distribution using all available (even unlabelled) data;

- The model changes should propagate in real-time with respect to labeling.

The design of the software component was driven to achieve the quickest possible turnaround time required to train a multi-class model for a particular task. While some tasks are pre-defined, many of them come from a particular marketing or public relations project on a particular new offering. Therefore, we allow the ad-hoc definition of tasks. We also focus on providing real-time training, so that new labels are quickly incorporated in the model.

3.1 Internal domain model

First, we describe the concepts of the diversity analysis software service. In order to accommodate several different tasks under the same umbrella (i.e. language, topic and sentiment detection), we operate with the following concepts:

- A **task** is a concept that describes the goal that we wish to achieve. Each task has multiple target classes. We allow examples also to not belong to any class. For example, language detection has each language as a target class, topic classification has topics and sentiment classification has positive and negative, along with neutral.
- A **model** is a particular approach configured to execute a task. A single task may have multiple models. Each model can have a single model algorithm, which defines the concrete approach for the classification (for instance SVM or Naïve Bayes).
- An **item** is a single data point, consisting of a set of features. This may be a single tweet or a survey record. Each item may have an associated source, language, author and geographic location.
- A **label** indicates that a particular item is associated with a particular target class in a particular task. The labels are used for training and evaluating the models.

For example, imagine that we have a sentiment detection task, which is a three-way classification problem (positive, negative, and neutral). Then, we declare a model using Naïve Bayes. In order to train the model, we use the search functionality to retrieve examples that we consider to be relevant for the model. Given a query "blackberry", we obtain the following tweet (among others):

I LOVE THE NEW BLACKBERRY TV COMMERCIAL. COMING AFTER BLACKBERRY BOYS CAMPAIGN BY VODAFONE, IMAGE MAKE OVER GUARANTEED FOR BLACKBERRY

When the model is still fresh and untrained, none of the examples are labeled. Since we wish to train the model, we assign a positive label to this example. That label gets added to the task and is subsequently used to retrain the model.

3.2 External domain model

While the system is general enough to map to any multi-concept annotation task, it is primarily meant to be applied to fit into the Knowledge Diversity Ontology (KDO) [11], which is designed to represent the diversity of opinions and mentions on the web, a domain which is relevant for describing situations in customer relationship and brand management. The classes in task examples that we are demonstrating can all be mapped to instances of KDO concepts.

Specifically for the tasks that we are demonstrating, we can map them to the following RDF statements: the language classification asserts an `xml:lang` property of the post. Topic classification is a task of annotating posts by asserting that they have a `sioc:topic` relation with a certain topic. In more specific classifications where we are considering detecting concrete entities, we recommend using the `kdo:mentions` predicate. In the sentiment detection case, we are annotating the post with the `kdo:hasSentiment` predicate, asserting that the post has a sentiment (`kdo:Sentiment`) of a given polarity (`kdo:hasPolarity`, `kdo:Polarity`), which can be one of `kdo:positivePolarity`, `kdo:negativePolarity` or `kdo:neutralPolarity`.

3.3 Description of approach

As stated above, in order to achieve good turn-around time, we employ active learning, which allows data instances to be labeled for training by an oracle (usually a human annotator). It is now a well-established finding that such methods can achieve higher accuracy with fewer training examples than passive learning [5].

However, these techniques often suffer from the cold-start problem: when no or few examples are labeled, it is difficult to get quality queries from the active learning mechanism. Therefore, we supply an alternative method for retrieving examples: besides offering to label uncertain examples, we also allow the user to supply a textual query that retrieves relevant items using a search index, a method which was also used in the OntoGen system in order to quickly define new ontology concepts [6]. If a model is available, we also use it to label the results of the search query.

The key hypothesis of active learning is that if the algorithm is allowed to choose the data from which it learns, it can perform better with less learning. Then, the chosen data points are submitted to the oracle as queries, which the oracle then labels. Active learning algorithms therefore need to have the ability to rank the unlabelled examples with regard to the expected utility they can bring to the model. In practice, there are several approaches to estimate that.

When using Support Vector Machines to train the classifier, one possible method to select the examples for labeling is to pick the examples which lie the closest to the hyperplane in the feature space that separates the classes [7]. In a multi-class setting, we can use the sum of distances to the separating hyperplanes. Intuitively, the examples that are close to the hyperplane that separates classes tend to be the ones which are less certain to belong in one or the other class. Therefore, it makes more sense to label those data points than labeling them randomly.

When using Naïve Bayes to train classifier, we can employ the uncertainty sampling approach [8]. In this framework, the active learner recommends the instances about which it is the least certain on how to label, which is straightforward for probabilistic learning models, such as Naïve Bayes. For multi-class learning, this can be generalized to recommend the instance whose prediction is least confident over all classes.

We employ basic feature extraction: the word tokens and bigrams are extracted from the content, stripped of case and used in a bag-of-words model.

3.4 Architecture

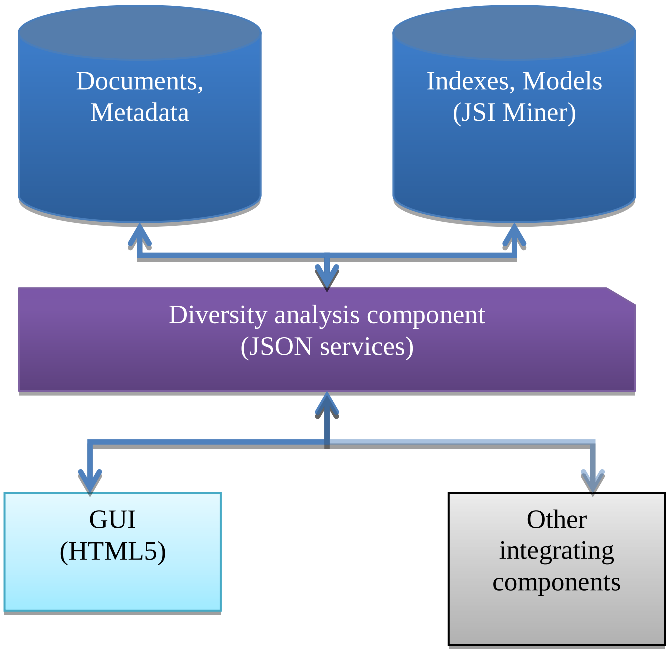

Fig. 1. Architecture of the interactive diversity analysis service.

The architecture of the component is displayed in Figure 1. While the diagram also includes a graphical user interface built for the purpose of demonstrating the system, the primary method of interface is through the Interactive diversity analysis API. In this work, we demonstrate the functionality of the component with a HTML5 client that directly connects to the API and is in no way privileged than other potential consumers of the component.

The back-end is driven by an embedded relational database to store the items and metadata. The search indexing and feature construction part is driven by JSI's Miner infrastructure to ensure handling of datasets bigger than available main memory. The indexing and retrieval back-end is the same that has also been used in other use cases, such as real-time news recommendation [9]. The models themselves are stored in-memory, since they need to be re-trained often. This does not pose significant challenges for scalability, since the required storage from the model is usually proportional to the size of the feature space and not item count. The system also utilized caching of generated feature vectors for items to enable efficient re-training of models when new labels arrive.

The main controller of the system is taking care of the maintenance of models. Since we require real-time data updates and immediate effect of labeling without blocking the user's requests, we have implemented concurrent model training. For instance: when the system receives a new label, it first checks if the model is currently in the process of re-training. If it is, the label is put in a queue for the next pass. This way, adding a label is a non-blocking process with regard to model training. When the model finishes training, the new re-trained model immediately replaces the old model, so that all subsequent classifications are executed with the new model. This ensures that also the search operation is non-blocking with regard to model training.

Usually, the training time is on the order of a second for several thousand examples – just long enough to be a perceptible delay in blocking mode. In non-blocking mode, this delay is barely noticeable, since adding a label and issuing a new search query with classification are two separate interactions, and the re-training usually finishes in the time before the user starts with search and classification.

3.5 User interface

Figure 2 displays the graphical user interface that demonstrates the functionality of the diversity analysis service, built in HTML5 and using jQuery for display logic. It has four main sections: the query configuration box, the result set, the uncertain examples and at the bottom, the configuration panel. The example shows the state after the user has issued the query 'blackberry' in the sentiment classification context, which caused all results to be labeled according to the model.

The main use case is the following: after the user defines the task and the model, she then issues a query that is considered to be most informative for that particular task. After the result set is displayed, each result is equipped with a set of buttons that act as labeling triggers: pressing one of them acts as adding a label into the model. If the user repeats the query after labeling some examples, he may notice that the results may be re-grouped and re-ordered. This is the consequence of online updates of the model and can also be used to correct the algorithm by labeling misclassified examples.

However, there are several strategies by which the models are trained. While the first strategy requires issuing a query and labeling the results as described in the previous paragraph, the second one is by looking at the uncertain examples that are provided in on the right-hand side of the UI. These examples are provided by the active learning mechanism and they are sampled with proportion to their classification uncertainty given the model. In probabilistic models, such as multi-class Naïve Bayes which is used in this implementation, the uncertainty is equal to the total uncertainty over all classes.

However, due to large class imbalance, which is often an issue in brand management tasks that we encountered, the uncertainty sampling approach suggested examples that were often unrelated to the task at hand. Therefore, we devised another strategy to mitigate that. The third strategy is a hybrid of both: it samples the most uncertain examples that satisfy the search query.

The behavior of the user interface is therefore the following: if the user issues a query without a model, the center of the page displays search results without any classification. If a user is working in the context of a particular active learning model, try to classify the search results in real-time and also provide the uncertainty samples that match the query. Thirdly, if the user issues a blank query, the system just provides uncertain examples in the right-hand blue box without any restrictions on a query.

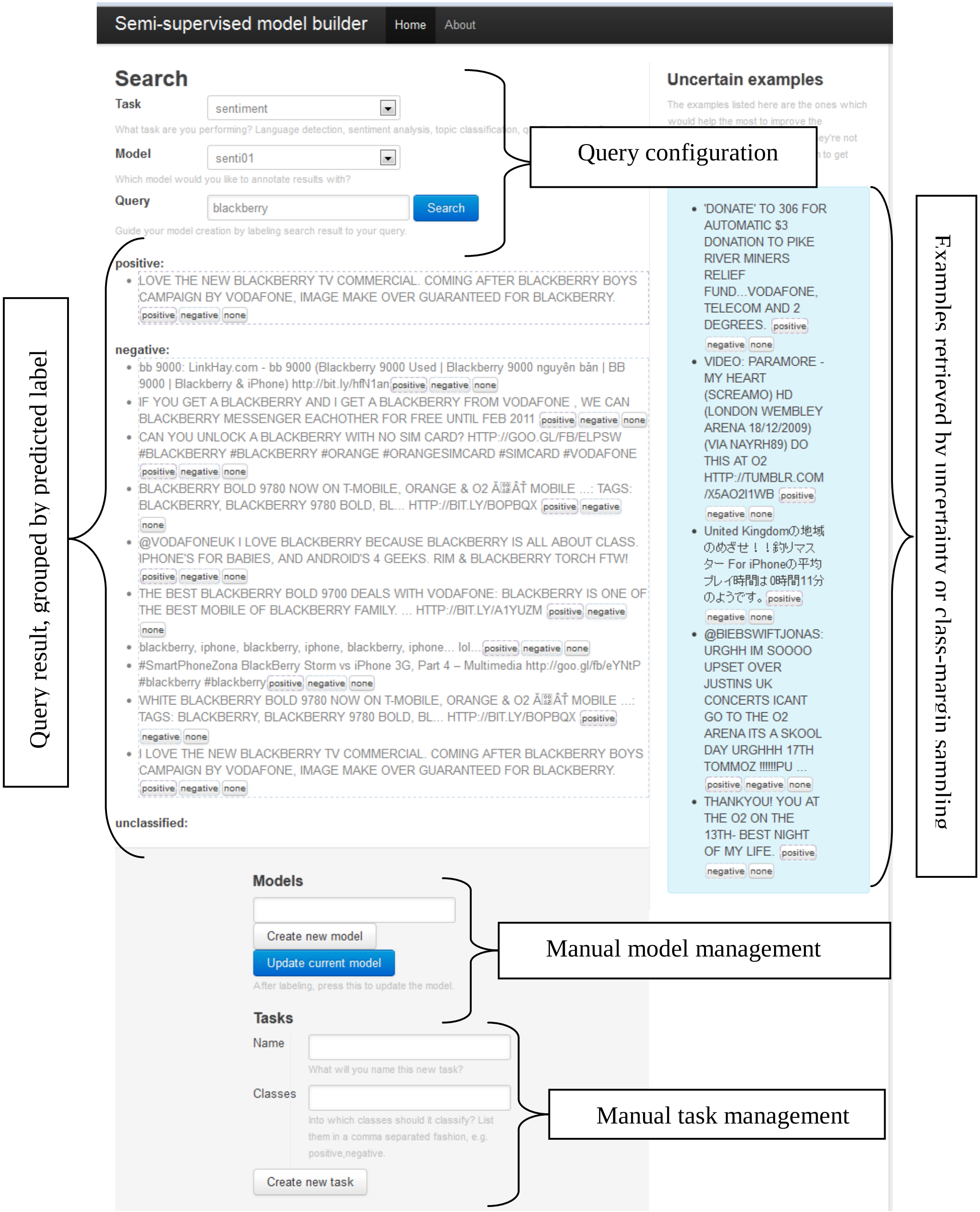

Fig. 2. HTML5 front-end that demonstrates the functionality of the diversity analysis service

4 Evaluation

The evaluation of the proposed approach and developed software is performed in the telecommunication domain on the dataset of Twitter posts. We are using a dataset of 14000 tweets gathered during one day.

We have created 3 scenarios, connected to common value management activities:

Scenario 1: Telecommunication company reputation management. For a particular telecommunication company ("*Vodafone"*) we have created a model, which displays the tweets related to company activities. Using active learning techniques, we have labeled a number of posts with positive, negative and neutral sentiment and then checked company reputation based on the created model.

The examples of positive and negative tweets about "Vodafone" are presented below:

Positive tweet:

VODAFONE INTRODUCES CUSTOMERS TO THE BLACKBERRY TORCH: IT'S AN EXCITING DAY FOR VODAFONE MALTA A... HTTP://BIT.LY/CIK0P6 #VODAFONE

Negative tweet:

VODAFONE: YOUR WIFE'S PHONE MAY NOT BE COVERED SIR. ME: WHY? VODAFONE: COZ SHE DROPPED IT. ME: WHAT DO YOU COVER? VODAFONE: NOT DROPPING IT.

Fig. 3. Examples of positive and negative tweets for Vodafone company

After labeling 50 posts, we have obtained the precision of 42 % and the recall of 50 %. Increasing labeling to 300 posts, we got the precision of 69 % and a recall of 63 %. In such way, active learning technique contributes to better results for our reputation management scenario.

Scenario 2: Competition analysis in the telecommunication domain. For a number of different telecommunication companies ("*Vodafone*","*O2"*) we have created a model, displaying the variety of opinions from Twitter users about the competitors.
Using active learning techniques, we have labeled twitter posts with relative posts for a particular competitor.

For a query "customer" we have observed how user opinions vary among the competitors. Fig. 4 provides the examples of different Twitter posts for competitors in order to spot weaknesses in their customer relationship management.

O2:

O2, BEEN A CUSTOMER OF THEIRS FOR EIGHT YEARS AND YET I PAY MORE THAN A NEW CUSTOMER. TOSSERS

JUST HAD THE RUDEST O2 CUSTOMER SERVICE MANAGER EVER ON THE PHONE. THEY NEED TO WORK ON THEIR CUSTOMER CARE COZ I AM SO LEAVING THEM.

Vodafone:

OF FRUSTRATION: DEALING WITH #VODAFONE CUSTOMER SERVICE. I HOPE YOU ARE LISTENING, #VODAFONE - YOUR CUSTOMER SERVICE SUCKS

A NICE CUSTOMER SERVICES MAN, HE'S GONNA SIT THERE ALL DAY TO GET ME CONNECTED TO UPGRADE SERVICES N1 #VODAFONE 'S CUSTOMER SERVICE.

Fig. 4. Examples of Twitter posts for different competitors

Scenario 3: Market analysis in the telecommunication domain. For a selected topic ("*BlackBerry"*), we have created a model, displaying user opinions about this topic in different languages (English, Spanish, and German). Using active learning techniques, we have labeled 20 posts as relative for the selected topic in a particular language. In such way, we were able to obtain user opinions in several markets – Spanish speaking market, English speaking market and German speaking market. Fig. 5 demonstrates the examples of Twitter posts for English and Spanish languages.

English:

LOVE THE NEW BLACKBERRY TV COMMERCIAL. COMING AFTER BLACKBERRY BOYS CAMPAIGN BY VODAFONE, IMAGE MAKE OVER GUARANTEED FOR BLACKBERRY.

GET FREE BLACKBERRY INTERNET WITH VODAFONE PAY & GO UNTIL 30 JUNE 2011 WE HAVE SIMS AVALIBLE SO DROP IN FOR FREE BLACKBERRY INTERNET

Spanish:

@SAJIID_GAGA HACIENDO BERRINCHE POR BLACKBERRY MSN VENTE A MOVISTAR UNIDOS HACEMOS MAS JAJA... Y EL COMERCIAL? LO SIENTO AMO MI TRABAJO JAJA

MOVISTAR SOLO TARDO MINUTOS PARA ACTIVARME EL PLAN BLACKBERRY.. DIDITEL TARDO 3 DIAS :/

Fig. 5. Examples of Twitter posts for different languages

With having more posts in English, after labeling 20 Twitter posts, the precision for English language constituted 53%. The recall for English language was 76 %. While state-of-the-art language detection techniques achieve higher scores (with more training data), our purpose was to demonstrate that even low training numbers can achieve reasonable performance.

5 Conclusion

In this paper we described the approach for diversity analysis based on social media data. We have shown how diversity analysis can be applied for a number of common value management activities, such as reputation management, competition analysis, market analysis, sentiment analysis. The presented approach can be used to train a language, topic or sentiment classifier.

This paper also described an example client of the software component. The created software allows working not only with pre-defined tasks, but also with ad-hoc definitions of tasks. The approach and software evaluation is performed in the telecommunication domain on the dataset of Twitter posts.

There are still several requirement points that need to be addressed. Most notably, we need to support composition of models via search facets, for example supporting queries like "search only for English O2 tweets and obtain their sentiment", as well as drill-down search. Another required feature is keyword suggestion, which can be achieved via query expansion and keyword extraction techniques. Besides that, we plan to also include support of structure extraction in opinions, a task appropriate for relational clustering and visualization methods.

When observing the model performance, we noticed that there is room for improvement in several directions for our future work. Firstly, the current implementation still does not take into account the potentially vast quantity of unlabeled data using semi-supervised algorithms, possibly further improving the convergence of the active learning workflow. Secondly, we can also improve performance by applying a transfer learning technique called multi-task learning [10], where one can train models for multiple related tasks simultaneously, having them improve each other.

Furthermore, there are several situations where we need to answer model-based queries which are special cases of existing queries. While many of these can be made via simple composition of model classifications and attribute-based criteria, there are some cases where model specialization is necessary. For example, we could use a general sentiment classifier in order to bootstrap another sentiment classifier for a very specific topic. For instance, training a sentiment classifier on the topic "pre-paid service for *Movistar* in Spanish" can be bootstrapped with the general Spanish sentiment model and specialized with additional annotations.

There are still open issues in the design of the classifier: for the sake of demonstration simplicity, the feature construction is currently the same for all of the tasks. Having a possibility to add new feature sets at runtime would definitely improve performance on several of the mentioned tasks.

Furthermore, since the algorithm is designed to work on streams of data, along with streams of correction inputs, it is also a potential scenario for application of online learning algorithms. Another interesting application of these techniques can also focus on providing global and aggregate results, not only samples that match queries.

6 Acknowledgements

This work has been partially supported by the project RENDER under the grant number FP7257790 funded by the Seventh Framework Programme of the European Comission and the Slovenian Research Agency.

7 References

1. Hasan, R., Siorpaes, K., Krummenacher, R., Flöck, F.: Towards a Knowledge Diversity Model, In: DiversiWeb-2011, Knowledge Diversity on the Web (2011).
2. Kim, S-M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of COLING (2004).
3. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on ACL (2002).
4. Bizău, A., Rusu, D., Mladenić, D.: Expressing Opinion Diversity. In: DiversiWeb-2011, Knowledge Diversity on the Web (2011).
5. Settles, B.: Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison.
6. Fortuna, B., Grobelnik, M., Mladenić, D.: Ontogen: Semi-automatic ontology editor. Human Interface and the Management of Information. Interacting in Information Environments, pp. 309-318, Springer (2007).
7. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. The Journal of Machine Learning Research, vol 2., pp. 45-66, JMLR.org (2002).
8. Lewis, D.D., Gale, W.A.: A sequential algorithm for training text classifiers. In: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval (1994).
9. Fortuna, B. and Fortuna, C. and Mladenić, D.: Real-time news recommender system, Machine Learning and Knowledge Discovery in Databases, pp. 583-586, Springer, 2010
10. Pan, S.J. and Yang, Q.: A survey on transfer learning, Knowledge and Data Engineering, IEEE Transactions on, vol 22, no. 10, pp. 1345-1359, IEEE, 2010
11. A. Thalhammer, I. Toma, R. Hasan, E. Simperl, and D. Vrandecic, How to Represent Knowledge Diversity, Poster at 10th International Semantic Web Conference, 2011.