Acknowledging Discourse Function for Sentiment Analysis

Phillip Smith and Mark Lee

University of Birmingham
School of Computer Science
Edgbaston, B15 2TT
P.Smith.7@cs.bham.ac.uk, M.G.Lee@cs.bham.ac.uk

Abstract. In this paper, we observe the effects that discourse function attribute to the task of training learned classifiers for sentiment analysis. Experimental results from our study show that training on a corpus of primarily persuasive documents can have a negative effect on the performance of supervised sentiment classification. In addition we demonstrate that through use of the Multinomial Naïve Bayes classifier we can minimise the detrimental effects of discourse function during sentiment analysis.

1 Introduction

In discourse, sentiment is conveyed not only when a speaker is expressing a viewpoint, but also when they are attempting to persuade. In this study we examine the influence of these two functions of discourse on sentiment analysis. We hypothesise that training a supervised classifier upon a document set of a single discourse function will produce errors in classification if testing on a document set of a different discourse function.

Ideally, we would have tested this theory on a currently used resource in sentiment classification in order to examine and compare the behaviour of the expressive and persuasive discourse functions in the overall classification process. However, no such resource is appropriately annotated with the expressive and persuasive labels that are needed to test our hypothesis. We have therefore developed a document set from the clinical domain, annotated on the document level with discourse function information. The document set used for our experiments contains 3,000 short documents of patient feedback, which we have made available online.[1]

We investigate our hypothesis by testing four supervised classifiers that are commonly used in both the machine learning and sentiment analysis literature [1]. The classifiers that we use are the simple Naïve Bayes (NB), multinomial Naïve Bayes (MNB), logistic regression (LR) and linear support vector classifier (LSVC). We use both binary presence and term frequency features for each classifier. We investigate our hypothesis by running four sets of experiments,

[1] http://www.cs.bham.ac.uk/∼pxs697/datasets

varying the training and testing sets of each. We run two across the same discourse function, expressive to expressive and persuasive to persuasive, and we run two using the concept of transfer learning; expressive to persuasive and persuasive to expressive. Results of experiment exhibit decreases of up to 38.8% F_1 when training on the persuasive document set and testing on the expressive. We also show that the classifier with the least variability in macro-average F_1 is the MNB classifier, which suggests its robustness to the the effects of discourse function when performing supervised sentiment classification.

The remainder of this paper is structured as follows. Section 2 outlines the theory of discourse function, and describes the nature of expressive and persuasive utterances that we encountered. In Section 3 we describe the corpus used for experimentation. We describe how our experiments were set up in Section 4, and in Section 5 we discuss our results and their relative implications. Finally, we conclude and discuss avenues for future work in Section 6.

2 Discourse Function

Our study hinges on the premise that a difference in discourse function may be detrimental to the use of supervised machine learning classifiers trained for sentiment analysis. We base our definition of discourse function on that proposed by Kinneavy [2] who argues that the aim of discourse is to produce an effect in the average reader or listener for whom the communication is intended. This could be to share how one is feeling, or perhaps to persuade them. These two discourse functions fall into the *expressive* and *persuasive* categories, respectively. Kinneavy also includes two other discourse functions, informative and literary, in his theory of discourse [3].

To illustrate his theory, Kinneavy represents the components of the communication process as a triangle, with each vertex representing a different role in the theory. This is somewhat similar to the schematic diagram of a general communication system that is proposed by Shannon [4]. The three vertices in the triangle are labelled as the encoder, the decoder and the reality of communication. The signal, the linguistic product, is the medium of the communication triangle. The encoder is the writer or speaker of a communication, and the decoder is the reader or listener.

2.1 Expressive

In communication, when the language product is dominated by a clear design of the encoder to discharge his or her emotions, or to achieve his or her own individuality then it can be stated that the expressive discourse function is being utilised [3]. In this paper, we take expression to be communicated through text. Since the discourse function is in effect the personal state of the encoder, there is naturally an expressive component in any discourse. We however narrow this definition to only observe explicit examples of the expressive discourse function in text.

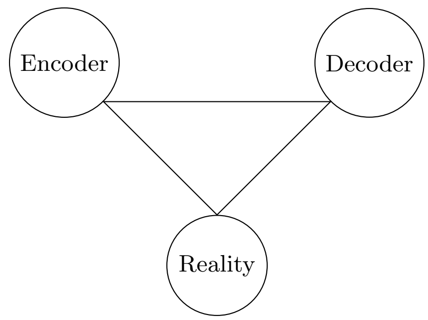

Fig. 1. Kinneavy's Communication Triangle [2]

We decompose the general notion of emotions that are conveyed to be valenced reactions, as either a positive or negative polarity based label. There is little consensus as to the set of emotions that humans exhibit, however methods have been put forward to extend these polarities into the realm of emotions [5, 6], so future work could extend this where needs be.

The components of expressive discourse when explicitly expressed are often trivial to identify. Utterances beginning with the personal pronoun *I* followed by an emotive verb often pertain to the expressive discourse function being utilised if they are succeeded by an additional emotion bearing component. Much research in sentiment analysis has observed the expressive discourse function [7–9].

2.2 Persuasive

Persuasion attempts to perform one or more of the following three actions: to change a decoder's belief or beliefs, to gain a change in a decoder's attitude, and to cause the decoder to perform a set of actions [10].

Sentiment can be viewed as a key component in persuasion, yet it is no trivial feat to define what a positive persuasive utterance is. We define what we shall call contextual and non-contextual persuasive utterances. First, let us observe the non-contextual persuasive utterances. An example of a positive persuasive utterance is: *You should give him a pay rise*. Taking this utterance alone, it is clear that the encoder of the signal is attempting to persuade the decoder to give someone more money for their work, which can be understood to be attempting to elicit a positive action from the decoder, for the benefit of the subject of the utterance.

To contrast this, we must demonstrate a non-contextual negative persuasive utterance. For example, take the utterance *Please fire him*. Here the encoder is attempting to stop the subject of the utterance from working, by persuading the decoder to ensure they cease working, which is typically seen as something negative (at least in Western societies).

Corpus	D_N	W	$D_{avglength}$	$W_{uniq.}$
Expressive				
Positive	750	47875	62	4869
Negative	750	50676	67	5411
Persuasive				
Positive	750	44527	59	4587
Negative	750	97408	129	7391

Table 1. Persuasive & expressive corpus statistics.

Now, we must also consider the class of persuasive utterances that we describe as 'contextual'persuasive utterances. An example of such an utterance is: Please give me a call. At first glance, this utterance lacks a clear sentiment. However if we precede this with the sentence *Great work!*, the above persuasive utterance becomes positive. However, if we precede out initial persuasive utterance with the sentence *You've messed up.* our seemingly emotionless persuasive utterance becomes negative. This agrees with the view of Hunston [11], that indicating an attitude towards something is important in socially significant speech acts such as persuasion and argumentation.

3 Corpus

The corpus which we use in our experiments is the NHS Choices Discourse Function Corpus, introduced in [12]. This is a corpus of patient feedback from the clinical domain. Patients were able to pass comments on hospitals, GPs, dentists and opticians through an online submission form. Whilst there were many fields to fill in, the fields that were of relevance to sentiment analysis were those labelled as 'Likes', 'Dislikes'and 'Advice'. These blanket labels help to define individual documents, and made the automatic extraction for experimentation a straightforward process. There was also no need to hand-label the likes and dislikes for sentiment, as the labels presupposed this. Annotation was required for the advice, as to whether a positive or negative sentiment was conveyed. This was undertaken by two annotators, and inter-annotator agreement was measured.

Typically, sentiment analysis concentrates on the positive and negative aspects of a review; the likes and dislikes. However, the literature [3] has shown that these expressive aspects of discourse function are not alone in communicating sentiment. As shown in earlier sections, the persuasive discourse function also conveys sentiment when it is employed. Advice comes under the umbrella term that is persuasion. When offering advice, the intention is often to persuade the decoder of the advice to act in a certain manner, or to acquire a certain belief set. This can be rephrased by saying that when we use the persuasive discourse function, we often use advice to successfully perform this action. Therefore, in this corpus, the comments of the likes and dislikes section of the corpus form the expressive subsection, and the comments that the patients submitted under the advice header form the persuasive subsection of the corpus.

In this paper we concentrate on a 3,000 document subset of the corpus. This is divided into two 1,500 document sets for the documents that primarily used the expressive and persuasive discourse functions respectively. The corpus can be further divided into two 750 document subsets with documents communicating a positive sentiment, and a negative sentiment. Table 1 outlines the token counts, average document length, and the number of unique tokens present in each section of the corpus that we used for experimentation. We should note that there were at least 750 contributors to this corpus, and with the data being mined from an online source, there were no stipulations as to the qualifications of the poster, so the language model that would be learnt by the classifier would have great linguistic variation.

4 Method

In our experiments, we wanted to explore how supervised machine learning algorithms are able to generalise across discourse function. In particular we examine the transferability of learned models trained on corpora of different discourse function. Our hypothesis is that differences in discourse function will detract from the transferability of learned models when detecting sentiment, if they are tested on datasets of differing discourse function. We experiment across all pairwise combinations of the training and testing document set. We initiate the experiments in this way so that the directionality of discourse function could be tested, along with the transferability of the learned models across discourse function. We use scikit-learn [13] and NLTK [14] Python packages for our classifiers.

When training our models (NB, MNB, LR and LSVC), we used the same data for each algorithm. The training set consisted of 1,000 training documents, 500 positive and 500 negative, for both of the respective discourse functions. The test set consisted of 500 documents, 250 positive and 250 negative, from each discourse function. These were randomly selected from the NHS Choices Discourse Function corpus [12], however ensuring that there is no overlap between the sets. For the expressive to expressive and persuasive to persuasive experiments, 10-fold cross validation of the machine learning methods was used.

5 Results & Discussion

Figure 2 shows the macro-average F_1 values for each experimental setup. Classifiers that are trained upon the expressive document set perform better than those trained on the persuasive document set, irrespective of classifier choice or feature set used. The NB classifier shows greatest variability in classifier performance, with a peak F_1 of 0.826 and a minima of 0.438. The LR and LSVC models also exhibit a degree of variation in F_1. The MNB classifier minimises the variability in performance, and is the most robust classifier that we tested.

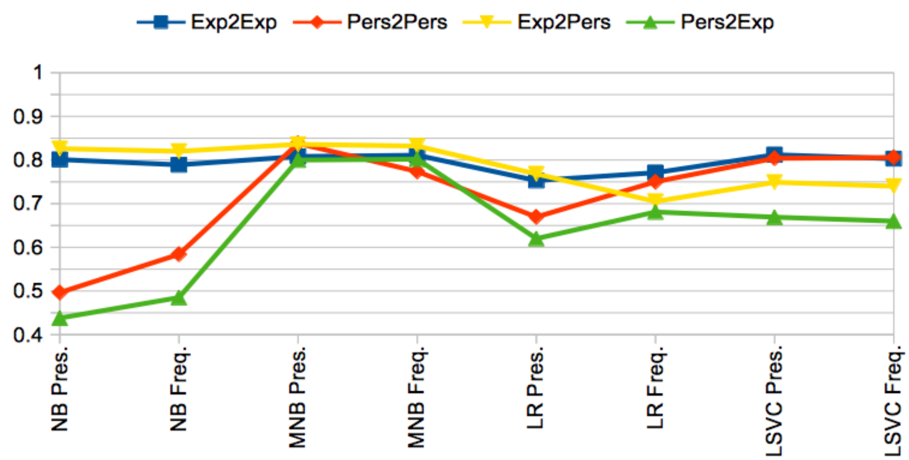

Fig. 2. Macro-averaged F_1 results for the cross-validation and transfer learning experiments.

Where other classifiers struggle when training on the persuasive document set, MNB achieves a macro-average F_1 of 0.802.

The results show the relative ease with which the expressive document set is able to create learned models of sentiment, and apply both to test sets of either an expressive or persuasive discourse function. When comparing cross-validation results to those of the transfer learning experiments, results exhibit minimal disturbance in macro-average F_1 score when models are trained on a corpus of expressive documents. This does not therefore support our hypothesis in the instance where we use the expressive document set to train our classifiers. However, this is only for the expressive function.

These results suggest that if there were a hierarchy of discourse functions, then persuasion is perhaps a subset of expression, and it inherits elements of the expressive vocabulary in order to carry out its role. We base this on the results of classification from the expressive to persuasive, and the poor adaptation of any classifiers trained on the persuasive document set. Consequently, we are inclined to believe that the persuasive discourse function cannot fully function without expressive elements. Examples of this are appeals to emotional elements, such as in congressional debates [15], where persuasion through fact alone are not the sole tactics used to sway the voters.

There is a clear drop in classifier performance when training on the persuasive corpus, and performing transfer learning. This supports our hypothesis for all classifiers, where each classifier trained in this way underperforms, sometimes to a considerable degree. We believe that this could be due to the implicit nature of sentiment that the persuasive discourse function conveys, and could be attributed to the structure of a text, in particular the interface between syntax and lexical semantics [16]. Further work is required to examine the differences in structure between documents of the respective discourse functions in order to confirm this assumption.

One interesting classifier is the MNB classifier. This performed consistently well during our study, and was even able to cope with the effects of cross-discourse classification to a high degree, performing well on the difficult persuasive to expressive classification experiments. We believe that this is due to the minimization in error rate that it has previously been shown to achieve, as it is able to deal with overlapping vocabularies and variable document length [17]. This performs considerably better than the simple NB classifier, and we believe that this is due to the difference in feature distribution that is observed in the models.

6 Conclusion

This paper has observed the effects of discourse function on supervised machine learning approaches to sentiment analysis. The effects of classification across the expressive and persuasive discourse function were recorded, and we found that despite both discourse functions conveying sentiment, the corpus with documents primarily utilising the expressive discourse function was preferable to train learned models upon, in comparison to a document set of primarily persuasive documents. In empirical results on a corpus of patient feedback containing documents of both discourse function testing across discourse, we found that there was an average improvement in accuracy of up to 38.8% when using the expressive subcorpus instead of the persuasive as a training set. We also find that the MNB classifier is preferable to others in order to minimise the effects of discourse function on sentiment classification.

In future work we will investigate further the effects of discourse function on other learned classifiers in order to determine if any others are able to minimize its effects on supervised machine learning models.

References

1. Liu, B.: Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing **2** (2010) 568
2. Kinneavy, J.E.: The Basic Aims of Discourse. College Composition and Communication **20** (1969) 297–304
3. Kinneavy, J.L.: A Theory of Discourse: The Aims of Discourse. Norton (1971)
4. Shannon, C.E.: A Mathematical Theory of Communication. Bell Systems Technical Journal **27** (1948) 379–423
5. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
6. Smith, P., Lee, M.: A CCG-based Approach to Fine-Grained Sentiment Analysis. In: Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology, The COLING 2012 Organizing Committee (2012) 3–16

7. Mullen, T., Collier, N.: Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In Lin, D., Wu, D., eds.: Proceedings of EMNLP 2004, Association for Computational Linguistics (2004) 412–418
8. Bloom, K., Garg, N., Argamon, S.: Extracting Appraisal Expressions. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, Association for Computational Linguistics (2007) 308–315
9. Dermouche, M., Khouas, L., Velcin, J., Loudcher, S.: AMI&ERIC: How to Learn with Naive Bayes and Prior Knowledge: an Application to Sentiment Analysis. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Association for Computational Linguistics (2013) 364–368
10. Miller, G.R.: 1. In: The Persuasion Handbook: Developments in Theory and Practice. Sage (2002) 3–17
11. Hunston, S.: Corpus Approaches to Evaluation. Routledge (2011)
12. Smith, P., Lee, M.: Cross-discourse Development of Supervised Sentiment Analysis in the Clinical Domain. In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, Association for Computational Linguistics (2012) 79–83
13. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
14. Bird, S., Loper, E., Klein, E.: Natural Language Processing with Python. OReilly Media Inc (2009)
15. Guerini, M., Strapparava, C., Stock, O.: Resources for Persuasion. In: Proceedings of LREC 2008, European Language Resources Association (2008) 235–242
16. Greene, S., Resnik, P.: More than Words: Syntactic Packaging and Implicit Sentiment. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics (2009) 503–511
17. McCallum, A., Nigam, K., et al.: A Comparison of Event Models for Naive Bayes Text Classification. In: AAAI-98 Workshop on Learning for Text Categorization. Volume 752., AAAI (1998) 41–48