

# Personal Sense and Idiolect: Combining Authorship Attribution and Opinion Analysis

Polina Panicheva<sup>1</sup>, John Cardiff, Paolo Rosso<sup>2</sup>

<sup>1</sup>Social Media Research Group, Institute of Technology Tallaght, Dublin, Ireland  
ppolin@itnet.ie, John.Cardiff@ittdublin.ie

<sup>2</sup>Natural Language Engineering Lab, ELiRF, Universidad Politécnica de Valencia, Spain  
proso@dsic.upv.es

## Abstract

Subjectivity analysis and authorship attribution are very popular areas of research. However, work in these two areas has been done separately. Our conjecture is that by combining information about subjectivity in texts and authorship, the performance of both tasks can be improved. In the paper a personalized approach to opinion mining is presented, in which the notions of personal sense and idiolect are introduced; the approach is applied to the polarity classification task. It is assumed that different authors express their private states in text individually, and opinion mining results could be improved by analyzing texts by different authors separately. The hypothesis is tested on a corpus of movie reviews by ten authors. The results of applying the personalized approach to opinion mining are presented, confirming that the approach increases the performance of the opinion mining task. Automatic authorship attribution is further applied to model the personalized approach, classifying documents by their assumed authorship. Although the automatic authorship classification imposes a number of limitations on the dataset for further experiments, after overcoming these issues the authorship attribution technique modeling the personalized approach confirms the increase over the baseline with no authorship information used.

## 1. Introduction

Research on analysis of blogpost writings is an area attracting an increasing amount of attention. On one hand, one of the popular research domains is the analysis of documents containing the subjective opinion of the author. Much work is dedicated to the subject of eliciting the emotions and opinions of the author. A well-known task in subjectivity analysis is the polarity classification of documents: texts that contain product reviews for example, can be divided into two groups: positive and negative reviews. A broad description of the work done in the area of subjectivity analysis, and namely, polarity classification, can be found in (Pang, 2008). On the other hand, another research subject attracting considerable interest is the discovery of the identity or characteristics of an individual, based on their writings. Not only objective characteristics of a writer, like personal traits, age or gender, can be successfully detected (Schler, 2006), but the actual author of documents can also be identified, and cases of plagiarism resolved (Koppel, 2008).

Work in these two areas, subjectivity analysis and authorship attribution, has hitherto been carried out independently of the other. However we believe that not only have these areas a lot to contribute to each other, but also that research on authorship attribution and subjectivity analysis should be done in a joint way. The reason for this is that our personality traits, occupation, age and gender are very closely interrelated with what we think, feel and express, i.e., our ‘private states’ (Pang, 2008).

In this paper, we present our experiments designed to test our hypothesis that appraisal polarity is expressed in an individual way by different authors, and that by

harnessing knowledge of the writer's idiolect we can improve the results of the polarity classification task. We expect authorship information to be useful for sentiment analysis. However, such information is not always easily found in the World Wide Web. We describe a methodology to combine information about the author's personality with that of their private states, and discuss limitations.

## 2. Personal Sense and Idiolect

We believe that information on authorship and author's personality can be successfully utilised in the subjectivity analysis. Information about an author's private states – her/his feelings, likes and dislikes – can reveal aspects of their personality characteristics. All of these form a person's idiolect – ‘a language that can be characterized exhaustively in terms of ... properties of some single person at a time’ (Barber, 2009). An idiolect would represent a collection of personal characteristics at the same time, i.e., age, gender, social class, occupation, as well as personal traits and private states. Thus, idiolect can be seen as a combination of the so-called sociolect, genderlect, slang, jargon, etc.

Words have word-meaning that is common in the language and that is represented in their usage. We believe that another component of word-meaning – “personal sense” (Leontev, 1978) – that is not inherent in the language, but is different for each person, can carry personal information and combine the tasks. This component reflects a meaning of the word in terms of unique experience of a person, reflecting partly their private states, and partly their unique personal characteristics.

Word-meaning and personal sense are manifested implicitly in speech. Our hypothesis is that personal sense

will influence language use of an author, forming an idiolect. The goal of our work here is to make use of these individual idiolects in order to facilitate sentiment analysis, personality and authorship attribution tasks.

### 3. Experiment: the Personalized Approach to Opinion Mining

#### 3.1 Knowing the Author

In opinion mining and polarity classification tasks the goal is usually to attach a document, or a piece of text, to one class or another. In polarity classification, documents are classified as having positive or negative opinion towards a product. From a formal point of view there should be a big difference between the documents expressing positive and negative opinion polarity, in order for the classification to be possible. Strong results in opinion analysis confirm this to be the case, see for example, (Pang, 2008).

However, the authorship of the documents is not normally taken into account. Given two sets of documents by Author1 (A1) and Author2 (A2), we find some features that A1 and A2 will have in common in positive appraisal documents, and other features that will work for both A1 and A2 in negative appraisal documents. Our consideration is that features that distinguish positive and negative opinion for A1 and A2 may be different. In other words, A1 and A2 might express and describe their appraisal in an individual way using their idiolects, so different from one another, that the overall polarity classification results might be improved, by classifying the documents written by A1 and A2 separately; using the A1 and the A2 set of documents as two different datasets for the experiment.

The aim of the experiment is to observe the hypothesis that in the idiolects of the different authors their appraisal is expressed in an individual way. In order to prove this, we constructed a corpus, in a similar format to the one described in (Pang, 2002). The corpus consisted of 300 short movie reviews, 30 reviews for each of 10 authors. For each of these 30, 15 were positive and 15 were negative reviews. To investigate different corpus volumes and to achieve higher statistically significant results, we doubled the corpus for each author, and repeated the experiment with 600 documents, 30 positive and 30 negative reviews for each author out of 10. We used the unigram features with the Linear Support Vector Machine<sup>1</sup> algorithm for the polarity classification.

As a baseline, we divided the corpus into 10 groups, each having 15 positive and 15 negative (30 positive and 30 negative for the doubled corpus) reviews, in a random way, so that every group consisted of documents by different authors. We performed the 10-fold cross-validation experiment for each of these groups separately. The mean accuracy result for the baseline experiment with the 10 shuffled groups was **56.47%** for the smaller corpus, **64%** for the doubled corpus.

For the next experiment, we used the same reviews, but we organized the 10 groups of documents so that each group corresponded to a single author, i.e., it consisted of 30 positive and 30 negative reviews by the same author. The authors were different for every different group. The number of documents and the settings of the classification experiment stayed the same. The mean accuracy result for the 10 groups was **69.67%**, for the bigger corpus the mean value reached **74.97%**. The t-test showed that for the experiment with 15 reviews written by a single author, the result was better than for the one with 15 reviews by random authors, with 75% significance; whereas for 30 reviews by a single author yielded better results than for 30 reviews by different random authors with 89% significance. Using the entire corpus as a dataset for the classification, for the 300 reviews the accuracy result was **73.17%**, while for the 600 reviews it was **78.35%**. We summarize the results in Table 1.

	Corpus 1 300 texts	Corpus 2 600 texts
Overall corpus classification	<b>73.17</b>	<b>78.35</b>
10 random groups	<b>56.47</b>	<b>64.00</b>
10 author groups	<b>69.67</b>	<b>74.97</b>

Table 1. The polarity classification results for different datasets (%)

The results for the both datasets confirmed our assumption that in different authors' idiolects the appraisal polarity is expressed in an individual way. Moreover, the mean result for the 60 documents by every author was slightly better than the result for all the 300 documents. This suggest the intuition that in terms of the volume of the datasets for polarity classification in the web, it is more useful to double the corpus by the same single author, than to increase it 5 fold using texts by different authors.

#### 3.2 Getting to Know the Author: Applying Authorship Attribution

Knowing the authorship of the reviews, we can use such information and increase the performance of polarity classification. However, this is not a very realistic state of affairs, when we use the ever-changing world wide web as a corpus. A very popular way of handling this issue is automatic authorship attribution. In our next experiment we applied an authorship attribution algorithm to the existing document corpus, investigated if the resulting authorship information increases the performance of polarity classification; and observe the drawbacks and limitations of the approach.

We used the Java Graphical Authorship Attribution Program (JGAAP), described in (Juola, 2006), for supervised authorship attribution task. The tool allows for the choice of the classification features, including lexical, character, phonetic, grammatical features; and the choice

<sup>1</sup> For the experiments we used a machine learning tool, Weka 3.6.1, that is available for download at <http://www.cs.waikato.ac.nz/ml/weka/>.

of the classifying algorithm: the traditionally used Naïve Bayes and Support Vector Machine and a number of others. For an authorship classification experiment using JGAAP it is necessary to have at least one training example per each author: it starts with learning authorship classes from a trial set of documents by known authors, and proceeds to classifying every document with unknown authorship against the resulting authorship classes.

In our corpus we had collected the 600 documents by 10 authors and a smaller 300-documents corpus, both balanced in terms of polarity and authorship. To perform authorship attribution, we used the smaller corpus as a reference group of documents with known authorship: for each author we had a learning set of 15 positive and 15 negative documents. We used the rest, i.e. the second half of the bigger corpus, as a test set.

After testing a number of features, the Character Trigrams yielded the best result, confirming our expectation based on (Stamatatos, 2009). The Cosine distance was experimentally used as the classifier.

The authorship classification accuracy results for different authors ranged very considerably from 0.3 to 0.93, with the standard deviation of 0.26, and the mean accuracy rate among the 10 authors reaching 0.64. We consider this a successful result, being very close to some of the mean accuracy results reported in (Juola, 2006) for the “Ad-hoc Authorship Attribution Competition” (AAAC), namely 0.65. On the one hand, our reference group contained a very big number of documents comparing to the AAAC competition tasks, which made the classification task easier. On another hand, most of the tasks in the competition only included 3 or less known author classes, whereas in our case their number was 10, which made the task harder and significantly decreased the task baseline.

The correlation coefficient for polarity classification and authorship attribution results for the 10 authors reached a small but positive number of 0.176, indicating an insignificant trend that the style of the authors that is distinctive in terms of idiolect, it bears a lot of idiolect features that distinguish it from other author’s idiolects, allows also for easier polarity classification.

### 3.3 Polarity Classification of Automatically Attributed Documents

We used the authorship attribution results described in section 3.2 in order to modify our corpus. We applied the results of the classification, so that for every author their document collection contained the documents, whose authorship was considered unknown and was identified automatically by the classifier. Thus, we got 20 collections, a positive and a negative one for each author. Our goal was to proceed with the polarity classification experiment on the new, automatically attributed dataset, to find out if authorship attribution algorithms can aid sentiment analysis, the same way as knowing authorship did.

According to our presuppositions, the application of the authorship attribution algorithm raises real-life issues crucial for the polarity classification task.

First of all, the resulting dataset was not balanced in terms of authors. With 30 documents per author at the

start, the resulting collections ranged from 8 to 46 texts. Secondly, and more importantly, the collections were not balanced against polarity anymore (see Table 2).

Author id	1	2	3	4	5	6	7	8	9	10
Positive	20	15	1	20	15	20	15	6	11	27
Negative	26	6	7	23	14	23	12	4	14	21

Table 2. Numbers of files for each of the 10 authors in the authorship attribution results

In order to overcome these issues, we supplemented the resulting datasets with documents used as authorship attribution learning examples, so that for each author:

- the volumes of the positive and negative datasets were the same;

- the volume of the set was exactly 30 documents.

For example, for the smallest set, consisting of 1 negative and 7 positive reviews, we added 14 randomly selected negative and 8 positive reviews from the learning set of the same author in the authorship attribution experiment. For the biggest set, containing 27 negative and 21 positive files, we had to eliminate 12 random negative and 6 random positive documents.

Obviously, this modified the dataset considerably and did not allow obtaining pure results, making the task easier for the smaller collections supplemented with documents by the same author, and harder for the bigger collections, from which documents had to be eliminated. This imbalance is reflected in the big correlation coefficient value of -0.320 between the polarity classification results and the authorship accuracy results for each author: the collections attributed modestly were easier to classify in terms of polarity than the collections attributed well, because the former were supplemented with files with native authorship, whereas from the latter some files, most of the by the native author, were eliminated, in order to balance the dataset.

The resulting collection of documents was used to perform the polarity classification experiment similar to one described in Section 3.1. The resulting mean accuracy of **57.67%**, according to our expectation, showed a statistically insignificant increase over the randomly grouped baseline result of **56.47%**. The accuracies for the 10 authors correlated positively with these from the experiment described in Section 3.1 for the 10 separate author-groups, with the correlation coefficient being 0.727.

However, from Table 2 it is obvious that not only for different authors the authorship attribution algorithm worked with various success rate, but there is also a strong tendency of the algorithm towards selecting a small number of ‘greedy’ classes and assigning most of the documents to them, while leaving the rest with almost no units. This demands a different evaluation framework, which is outside of this work.

In our case there were four authors, id 1, 4, 6 and 10, representing the ‘greedy’ classes: starting with 15 files, each class gained at least 20 at the end. Initial analysis of the results for these classes shows that when performing authorship attribution to aid sentiment analysis, it is these

'greedy' groups that should be aimed at and evaluated, despite the fact that they do not always represent the actual authorship of the documents.

#### 4. Conclusions and Further Work

The experiment described in this paper confirmed the hypothesis that the appraisal polarity is expressed in an individual way by different authors; moreover, the differences are so considerable that in order to investigate the polarity of documents automatically, a subsequent amount of documents by the same author gives more useful information than a much bigger sample of documents written by other authors.

The personalized approach has improved the results of the polarity classification task. This leads to the intuition that any opinion mining task could be improved if considered in terms of idiolect. We applied an authorship attribution algorithm, to test whether, and to what extent, the personalized approach with authors known could be substituted with automatic authorship attribution. A simple authorship attribution algorithm with medium performance proved to supply useful information for polarity classification task, increasing the performance. As expected, authorship attribution imposed limitations on the dataset in terms of its volume and balance, making the subsequent polarity classification results harder to evaluate.

Thus, we conclude that taking into account authorship, whether known or classified automatically, is a useful direction to take in sentiment analysis. However, the former is not particularly realistic provided that the corpus is extracted automatically from the web, and the latter imposes limitations, especially when applied to a small dataset. This is why we consider investigating features of idiolect representing broader groups of authors in our future work, namely groups of authors sharing the same occupation.

Our conjecture is that personal sense relates to occupation, or profession, in a particularly strong way. Occupation influences the everyday experience of a person, forming a sociolect, common among individuals of the same profession, but differentiating them from those working in another field. Thus, the next step of the personal sense and idiolect research is to find out, in what way and degree occupation actually forms a sociolect in a person's language use. Our hypothesis is that occupation plays an important role in speech, and our goal here is to identify the features that reflect the professional differences in language use. The experiments are aimed at harnessing differences in perspective for authors belonging to different professions.

To summarize, our initial results have proven that by acquiring knowledge of the writer's idiolect the results of the polarity classification task can be improved. In our future experiments, we will attempt to further reinforce the value of exploiting the concept and role of idiolect in subjectivity analysis tasks.

#### Acknowledgement

The work of the third author is supported by the TEXT-ENTERPRISE 2.0 TIN2009-13391-C04-03 research project.

#### References

- Barber, A. (2009). "Idiolects", *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/spr2009/entries/idiolects/>, accessed 5 November 2009
- Juola, P., Sofko, J., Brennan, P. (2006). A prototype for authorship attribution studies, *Literary and Linguistic Computing*, vol. 21, no. 2, pp. 169-178
- Koppel M., Schler J., Argamon S. (2008). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology*, Volume 60 Issue 1, pp. 9-26
- Leontev, A.N. (1978). *Activity, Consciousness, and Personality*. Hillsdale, Prentice-Hall.
- Mitrofanova, O., Lashevskaya, O., Panicheva, P. (2008) Statistical Word Sense Disambiguation in Contexts for Russian Nouns Denoting Physical Objects., *Proc. TSD 2008*, pp. 153-159
- Pang B., Lee L. (2008). *Opinion Mining and Sentiment Analysis*. N.Y.:Now Publishers Inc., 2008.
- Pang, B., Lee, L., Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; pp. 79-86
- Panicheva P., Cardiff J., Rosso P. (2009) A Co-occurrence Based Personal Sense Approach to Opinion Mining. In: *Proc. 1st Workshop on on Opinion Mining and Sentiment Analysis (WOMSA), CAEPIA-TTIA Conference*, Seville, Spain, November 13, pp. 205-212
- Schler, J., Koppel, M., Argamon, Shl., and Pennebaker, J. (2006) Effects of Age and Gender on Blogging. In *Proc. 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*
- Stamatatos, E. (2009). Intrinsic Plagiarism Detection Using Character n-gram Profiles. In *Proc. of the 3rd Int. Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse*