Author Masking by Sentence Transformation

Notebook for PAN at CLEF 2017

Daniel Castro-Castro[1], Reynier Ortega Bueno[1], Rafael Muñoz[2]

[1]Desarrollo de Aplicaciones, Tecnología y Sistemas DATYS, Cuba

{daniel.castro, reynier.ortega}@cerpamid.co.cu

[2]Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, España

rafael@dlsi.ua.es

Abstract. Masking the writing style of an author has been useful and used by novelists for the purpose of passing unnoticed, as well as by people who aim to give information without being linked to it. Within the PAN evaluation framework, it is presented the task of paraphrasing or changing the writing style of a document, maintaining the topic that is being discussed. We propose a method that performs transformations in sentences, with an unsupervised approach, i.e., without previous data of the author or linguistic characteristics of a document collection. We make syntactic and semantic changes using dictionaries and semantic resources, as well as syntactic rules for sentence simplification. In the evaluation section, we will expose the observed strengths and weaknesses of the proposal.

Keywords: Author Masking, sentence transformation, sentence simplification

1 Introduction

In the past PAN 2016 evaluation framework, the task of Author Masking was presented, which consists in: "Given a document, paraphrase it so that its writing style does not match that of its original author, anymore". As described in the overview [4], several works focused on translation strategies [7][8], syntactic, lexical and semantic transformations [6][7][8].

For this year, the bases of the evaluation are maintained. The documents to be obfuscated are written in English and it is requested that the transformed or modified returned fragments do not exceed 50 terms [4][1].

In different publications reviewed, it can be seen that on average, sentences contain less than 50 words, with exceptions such as, legal documents or literary works. Generally, documents that contain facts, such as the news, are written with sentences that are not very extensive and with concrete ideas. Another interesting detail is that not all people write with very long sentences, because this is characteristic of those with high educational levels.

Taking into account these ideas, we consider that if we perform transformations in sentences, seeking to make them shorter, keeping the central idea, and replacing words with not used synonyms, we will achieve that the documents present significant changes at the syntactic level and that many of them are similar, making it difficult for Authorship Analysis algorithms to determine the true author of a document.

We will not use linguistic features previously calculated or extracted from collections of documents, because this scenario is not always available.

2 Implemented Proposal

An essential element for the analysis and transformation of a text corresponds to natural language analysis and processing tools that we use. In the present proposal we use the FreeLing 4.0[1] [1] Open Source Natural Language Processing (NLP) tool.

All the documents that must be transformed are written in English. The collection available for this task, is organized in a set of folders where in each one a document named "original.txt" is included and this is the one that we must mask.

In Figure 1 we show a summary of the architecture of the implemented method and we will describe with more details the principal stages involved. We will emphasize the sentence simplification stage, since for this, we adapted a simplification and decomposition tool developed in our center for the Spanish language.

As we mentioned earlier, it is important to be able to develop a stage of pre-processing and linguistic analysis of the document that we want to transform, for this we use the FreeLing tool. FreeLing allows us to perform the analysis of texts in UTF-8-code, which is an important characteristic for the output that is requested and to parse the original text. Initially, the text is segmented into lexical tokens and subsequently is performed the Part of Speech Tagging in each of the sentences obtained. The last two stages were the Named Entities Recognition (important for the simplification stage) and the Word Sense Disambiguation (WSD) to identify the sense of the words used in a text. For the WSD we used the UKB algorithm [3] available in FreeLing. We use the default parameters provided by FreeLing in each of the methods employed.

Having the text pre-processed, we proceed to perform the transformations. As we commented in the introduction, we included lexical, syntactic and semantic transformations, as well as substitution of contractions, substitution of synonyms and sentence simplification.

Contraction replacement: we use a dictionary of contractions (manually build searching in different sources in the web) and their expansions for the English language, manually constructed by linguistics specialists. The dictionary is loaded into a bimap key structure in order to perform efficient searches. In the segmented text we look if the author uses more contractions than his expansions or vice versa. If the author uses more contractions, then we replace all used contractions by their corresponding expansions. We proceed in a similar way if the expansions of the contractions are majority.

1 http://nlp.lsi.upc.edu/freeling/

In Table 1 we illustrate a part of the dictionary, which consists in 144 contractions and their expansions.

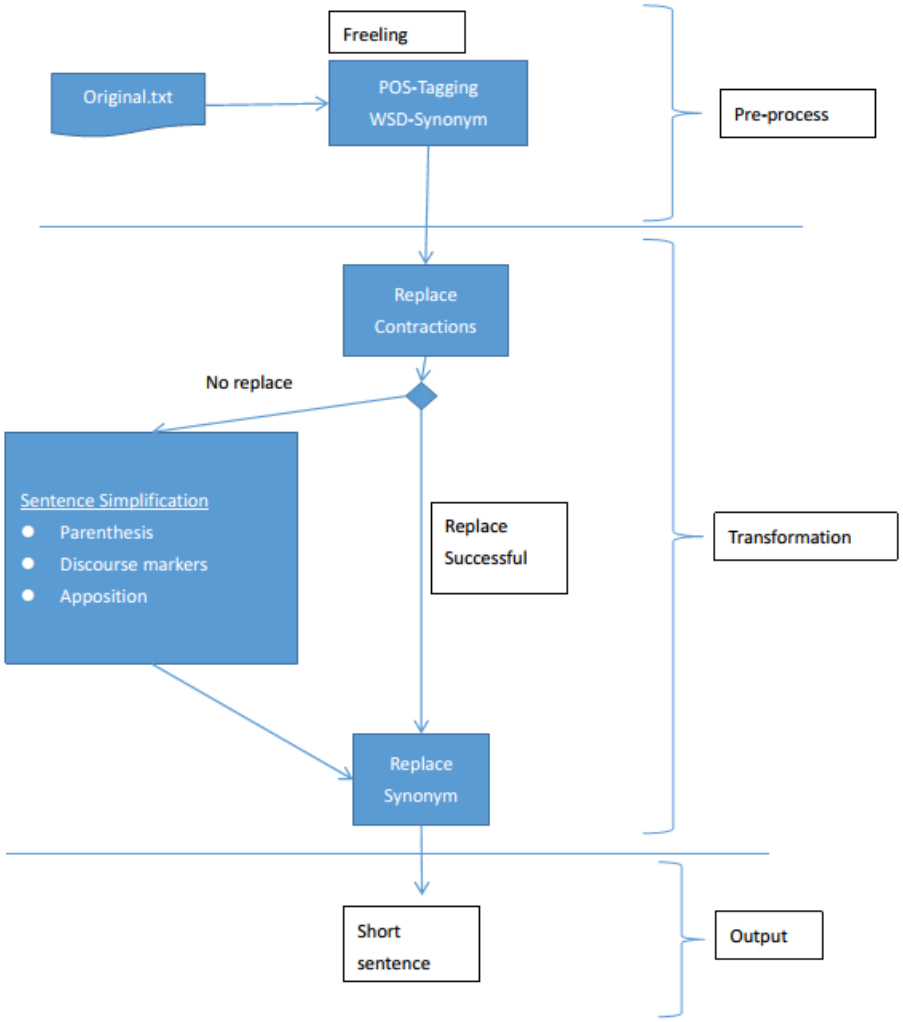

Fig. 1. Architecture of the Author Masking proposal. Simple sentence transformation.

Table 1. Examples of contractions and their corresponding expansions

Contractions	Expansion
Don´t	Do not
He´s	He is
I´d	I had

Synonym substitution: To accomplish this phase, we use the UKB sense disambiguation algorithm to obtain the words synset used in the sentence. Having for

each word the synset with which it is used in the context, we obtain from the semantic FreeLing resources the list of synonyms corresponding to that synset. We count the number of synonyms that have been used in the text for each word, and if there are synonyms in the list that were not used, then we substitute the word in analysis by this unused synonym. If there are more than one unused synonym, we selected the first in the list.

We include this synonym substitution, thinking in different scenarios. Some people do not care or are not aware that they use little lexical variety in their texts, so they tend to use the same words or very few known synonyms of a word. If we substitute a word by an unused synonym, then we would be varying its lexicon. It is important to note that this stage depends on the quality of the syntactic and semantic algorithms used and the richness and completeness of semantic resources.
Example of word substitution by synonyms:

- Original sentence: This is a rather complex topic to discuss, why I have chosen three main subjects to concentrate on: Would the curiosity decrease and lead to less use of narcotics, if the drugs were legalized?
- Simplified sentence: This is a rather complex issue to have out, why I sustain to pick out three main issues to center on: Would the wonder diminish and extend to less employment of narcotics, if the drugs were decriminalized?

Sentence simplification: In our research center2, we implemented with the support of linguists, a library for Spanish sentence simplification and decomposition. Based on this work, we adapted the rules defined for Spanish to be used considering the English grammar, and did not use the sentence decomposition part, because for that it is necessary the shallow parsing tree and these are more complex rules. We did not use rules supported by a superficial or deep syntactic analysis, only simple rules defined by the grammatical POS-Taggin labels and the lexical units present in the sentence.

We simplified elements in parentheses, where no Named Entities appeared inside the parentheses; eliminated discourse markers in the sentence, using for that a dictionary (manually build searching in different sources in the web). Another idea was the elimination of appositions, we divided the appositions into two classes and we called them apposition “pattern 1” and apposition “pattern 2”. Appositions are the phrases in which a Named Entity (NE) appears and also it´s explanation (this explanation is the apposition). Pattern 1 will correspond to the phrase or fragment of text where a NE appears first and next it´s explanation after a comma; Apposition pattern 2 is when the explanation of the NE appears first and then the NE. Within the appositions we eliminate the explanation.

In the documents of the Masking evaluation task, there occur few of these appositions. As a generalization, we include the deletion of text fragments that appear between commas and after another fragment of text. These fragments of text between commas could correspond to some list, explanation of a previous idea, etc.
Example of parentheses simplification:

2 http://www.cerpamid.co.cu

- Original sentence: Use a pointed stick (a pencil with the lead point broken off works well) or a similar tool.
- Simplified sentence: Use a pointed stick or a similar tool.

Example of discourse markers elimination:

- Original sentence: Basically, my job involves computer skills.
- Simplified sentence: My job involves computer skills.

Example apposition pattern 1 elimination:

- Original sentence: Athena, goddess of wisdom, helped the Greeks in the battle.
- Simplified sentence: Athena helped the Greeks in the battle.

In order to not incorporate abundant transformations, we decided that the sentence simplification will only be performed when no contractions are substituted in the sentence of analysis.

In the output file we included only transformed sentences, which had less than 50 terms and included at least one change.

3 Experiments and results

In the overview [1], each of the participants proposal are evaluated taking into account three dimensions [5], *safe, sound and sensible*. The first one of these dimensions is measured using automatic authorship verifiers and the last two by a peer-review process [1].

We will analyze some manually extracted examples from the results of the analysis of our method using the documents of the training set, exposing cases of transformations that were syntactically and semantically correct and other examples that we considered errors. These examples are extracted from the analysis and obfuscation of the provided texts of the collection of the year 2016 that can be freely downloaded.

Example of deletion the phrases between commas:

- Original sentence: He will never become a mature, responsible adult, even if he would succeed in getting rid of his addiction to drugs (one of the effects of hashish is that it stops the psychological development if used during the adolescence).
- Simplified sentence (correct): He will never become mature even if he would succeed in getting rid of his addiction to drugs.
- Original sentence: Well, it's the rules, you know.
- Simplified sentence (mistakes): Well you know.

Example of word substitution by synonyms:

- Original sentence: Oh--you--yes!
- Simplified sentence (mistakes): buckeye state -- you -- yes!

We presented some errors in the NLP pre-processing stage (segmentation and POS-Tagging) using the FreeLing tool, obtaining very long text fragments of more than one sentence, which we did not analyze and caused, that for some documents the number of transformations made, were low or none.

4 Conclusions and future work

We proved through the manual analysis performed to the execution of our method, that it is feasible to use lexical transformations, substitution of contractions or their expansions and the elimination of discourse markers and fragments of text in parentheses. The simplification by apposition only considering the grammatical labels of the words is not sufficient and some errors are generated, obtaining extremely short sentences. Replacing words with not used synonyms can introduce very elaborate elements, and it can also transfer errors caused by using NLP tools, although this is an idea to continue working on since it is a phenomenon present in practical scenarios.

We propose to consider strategies for the decomposition of coordinated, juxtaposed and subordinated sentences; evaluating semantically related words by hyperonymy and hyponymy; as well as strategies for mixing sentences.

5 References

1. Hagen, M., Potthast, M., Stein, B.: Overview of the Author Obfuscation Task at PAN 2017: Safety Evaluation Revisited. In: Working Notes Papers of the CLEF 2017 Evaluation Labs. CEUR Workshop Proceedings, CLEF and CEUR-WS.org (Sep 2017)
2. Lluís Padró, Evgeny Stanilovsky. FreeLing 3.0: Towards Wider Multilinguality Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA. Istanbul, Turkey. May, 2012.
3. Lluís Padró, Samuel Reese, Eneko Agirre, Aitor Soroa. Semantic Services in FreeLing 2.1: WordNet and UKB In Pushpak Bhattacharyya AND Christiane Fellbaum AND Piek Vossen (ed.) Principles, Construction, and Application of Multilingual Wordnets pg. 99--105. Narosa Publishing House. Global Wordnet Conference 2010. Mumbai, India. February, 2010.
4. Martin Potthast, Matthias Hagen, Benno Stein: Author Obfuscation: Attacking the State of the Art in Authorship Verification. CLEF (Working Notes) 2016: 716-749
5. Matthias Liebeck, Pashutan Modaresi, Stefan Conrad: Evaluating Safety, Soundness and Sensibleness of Obfuscation Systems. CLEF (Working Notes) 2016: 920-928
6. Muharram Mansoorizadeh, Taher Rahgooy, Mohammad Aminian, Mehdy Eskandari: Author Obfuscation using WordNet and Language Models. CLEF (Working Notes) 2016: 939-946
7. Tsvetomila Mihaylova, Georgi Karadjov, Yasen Kiprov, Georgi Georgiev, Ivan Koychev, Preslav Nakov: SU@PAN'2016: Author Obfuscation. CLEF (Working Notes) 2016: 956-969
8. Yashwant Keswani, Harsh Trivedi, Parth Mehta, Prasenjit Majumder: Author Masking through Translation. CLEF (Working Notes) 2016: 890-894