Automatic Synonym and Phrase Replacement Show Promise for Style Transformation

Foaad Khosmood
Department of Computer Science
University of California at Santa Cruz
Santa Cruz, California, USA
foaad@soe.ucsc.edu

Robert Levinson
Department of Computer Science
University of California at Santa Cruz
Santa Cruz, California, USA
levinson@soe.ucsc.edu

Abstract— **Style transformation refers to the process by which a piece of text written in a certain style of writing is transformed into another text exhibiting a distinctly different style of writing without significant change to the meaning of individual sentences. In this paper we continue investigation into the linguistic style transformation problem and demonstrate current achievements in transformation on sample texts from a standard authorship attribution corpus. Specifically, we use simple synonym and phrase replacement on the source text to strengthen the stylistic markers of a given target corpus. We validate our results using Java Graphical Authorship Attribution Program (JGAAP). We are able to demonstrate that simple replacements can alter the linguistic style of writing as detected by an independent process.**

Computational Stylistics, Style Processcing, Natural Langauge Processing, Machine Learning, Computational Linguistics

I. INTRODUCTION

"Written style" as a concept is notoriously difficult to define and therefore difficult to process computationally. Style, however, is integrally important in many computationally oriented fields. Researchers working in natural language processing (NLP), computational stylistics, authorship attribution, machine translation (MT) and corpus linguistics often focus on practical results-oriented tasks such as text classification which do not require in depth modeling of written style. Other researchers working in style-rich natural language generation (NLG) incorporate sociology and psychology to exhibit stylistic traits of pre-defined personality archetypes [1].

Style transformation parallels the larger problem of natural language processing in that it requires both understanding and generation of styles. This problem can be likened to machine translation [2] or statistical machine translation (SMT) since one text is effectively being translated into another form of the same text with the goal of preserving the meaning. Unlike MT/SMT, however, there are no significant corpora of human tagged equivilance classes between various styles that can be used for training. In addition style classifications themselves don't enjoy universal and hierarchichal taxonomic distinctiveness the way natural languages do.

We employ the concepts presented in [3][4] which link the problems of style classificaiton and style transformation. The authors argue that a robust classification algorithm can facilitate a measure of progress as well as a definable mark of success for transformation projects.

In the following sections, we briefly discuss the AAAC [5] reference corpus, construct our own style transforms, apply them to that corpus and discuss the results, implementaion and future work.

II. AAAC REFERENCE CORPUS AND JGAAP

Finding corpora for the purposes of stylistic transformation can be difficult. Ideally we would like sets of texts equivalent in most respects such as meaning, length, topic and viewpoint but with distinctly different writing styles. Furthermore, we need an unbiased process that is able to correctly classify unexamined texts of each style (corpus). The same process can be used to validate successful transformation by indicating a change in classification.

The corpus used for the 2004 Ad Hoc Authorship Attribution Contest [5] is probably the closest corpus to the ideal that we could obtain. We can assert that different authors have different authorial styles, and such stylistic differences are used in attribution algorithms.

The AAAC corpus has over a dozen problems labeled "A" through "M". Each of these problems consists of a set of training documents with known authors and a set of test documents with unknown authors. The task is to correctly match the test documents with their authors. These problems contain different types of text (essays, short stories, letters, plays and novels) mainly in English but also other languages. Problem A and K were generally considered the most difficult [6]. Problem K consists of excerpts from a book in Serbian-Slavonic. Problem A, which we concentrate on, consists of short essays on the same topic written by 13 US College students.

All the AAAC problems and some of the common algorithms used were later incorporated into JGAAP [7], a Java-based authorship attribution tool designed by Patrick Juola. JGAAP allows a particular attribution process to be easily repeatable.

We use JGAAP to validate our transformation results using one of the best performing algorithms, RN Cross Entropy (RNCE) from the AAAC problem A. "Cross Entropy" is "a measure of the unpredictability of a given event, given a specific (but not necessarily best) model of events and expectations" [8]. For more information about

cross entropy see [12] and see [13] for application to corpus linguistics. This algorithm, as implemented in JGAAP, correctly classifies 9 out of 13 sample texts in problem A.

We consider a sample unattributed text from the test data which JGAAP/RNCE can classify correctly. We then apply our transformation algorithm to it, and then add the transformed text as a new test document and re-run the same attribution program. If JGAAP shows a different classification, for the transformed text, we can conclude that some level of stylistic change has taken place, at least as it pertains to style markers relevant to the classification algorithm.

III. Style Transforms

Transforms are methods of manipulating and/or rewriting text. They can be defined at word, sentence and paragraph levels [3]. Transforms can vary from minute non-contextual regular expression based text replacements to sophisticated semantic rewriting and paraphrasing techniques using full parsing.

Transforms often correspond to style markers. Style markers are features whose presence transforms are designed to proliferate or diminish. In [3] a few different transforms are discussed which induce change in sentence lengths, paragraphs and acronyms. By far the most effective transforms are those based on vocabulary. This is partly because vocabulary usage has proven to be one of the more reliable and accurate markers in authorship attribution.

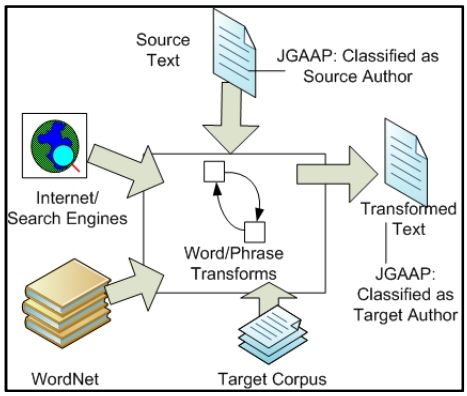

Figure 1. Overview of the transformation process

In this paper, we use two different vocabulary based transforms: synonym word replacement, and phrase replacement. We have implemented both of them in one complimentary algorithm which we describe below.

A. Phrase replacement transform algorithm

We present the phrase replacement transform in a series of steps below. Note that "sample" refers to the sample text being altered and "target" refers to the corpus that exhibits the desired style.

1. **Preprocessing:** All words and corresponding hit-rates (frequencies of occurrence in the corpus) are extracted from the target and made accessible in a file or database. Every word in the sample is tagged using a part-of-speech (POS) tagger.
2. **Phrasing:** Every sentence in sample is divided into a sequence of non-overlapping sentence fragments seperated by punctuation marks including commas, colons, simicolons and quotation marks. A phrase, for our purposes can be a one, two, three, four or five word sequence in the same sentence fragment. Therefore a word is considered a phrase of length one. Each phrase will be examined for replacements from WordNet [10]. In forming the phrase chunks, preference is given to the largest phrase that has a synonym in WordNet. Furthermore, the algorithm greedily selects phrases from left to right. Any phrase that is either tagged as a proper noun or quoted in quotation marks is skipped over as a candidate for replacement.
3. **Synset retrieval:** Each remaining candidate is looked up for one or more synonyms in WordNet where synonyms exists in groups called synsets. Single words are stemmed and have their POS tag also included in the WordNet lookup for further accuracy, i.e. only synsets with the identical tags are returned as candidates.
4. **Sense disambiguation:** From the sequence of word senses that are received from WordNet, the most common sense is chosen to extract synonym lemmas from.
5. **Synonym ranking:** Every synonym lemma of the chosen sense is scored based on the hit rate of its corresponding original lexeme in the target corpus, or based on statistical popularity derived from online databases, search engines or other sources. The latter choices are useful when there is not a specific target corpus to conform to and the desired outcome is a stylistic shift away from the original corpus, but not necessarily toward any other corpus.
6. **Replacement**: The lemma with the highest score is chosen to replace the original phrase. The replacement phrase is constructed by combining the replacement lemma with the original phrase residue (all the non-stemmed parts of the phrase usually constituting inflection). For example, the word "utilized" is stemmed as "utilize" with the "ed" as residue indicating past tense. We reference "utilize" as a verb in WordNet and the word "use" is chosen among the synonym lemmas. Combining "use" and "ed" will produce "used" which is the actual replacement phrase. Non-standard inflections such as "went" and "drove" are supported.

B. Tools and implementation

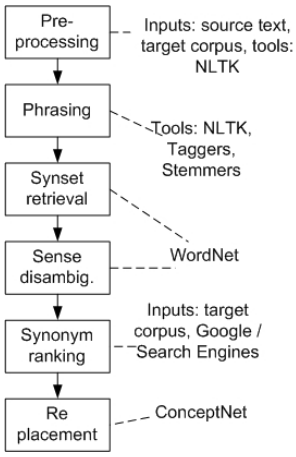

Figure 2. Summary of tools and inputs

In general we found the NLTK [8] most useful for its rich functionality and ease of integration with Python on Linux.

For the initial part-of-speech (POS) tagging, however, we used the Stanford log-linear POS tagger [9] with the left-3-words algorithm which has performed with 96.97% on the Wall Street Journal data set. For synset lookups, we used the WordNet interface from NLTK. For lemma operations including lemma/residue split and combine functions we used the natural language tools from MIT ConceptNet [11]. The transform itself was implemented as a python program.

The original and transformed samples were evaluated using JGAAP. The settings for running JGAAP (version 4.0) were the following:

- Load AAAC problem A defaults, adding the transformed sample as a new document with an unknown author
- Apply "Strip punctuation" to all documents in canonization step
- Use "words" as event set
- Use "RN Cross Entropy" as distance measure

IV. Results

A. Inducing stylistic shift

We demonstrate that a stylistic shift is possible using phrase replacements. We process all 13 documents of problem A with the phrase replacement transform above.

TABLE I. STYLISTIC SHIFT ON AAAC-04 PROBLEM A

Problem A Sample	True Attribution (Author #)	RNCE on original	RNCE on transformed	Stylistic shift
Sample1	3	12	12	no
Sample2	13	2	13	yes
Sample3	11	10	10	no
Sample4	7	4	4	no
Sample5	10	10	10	no
Sample6	12	12	13	yes
Sample7	8	8	2	yes
Sample8	1	1	9	yes
Sample9	5	5	5	no
Sample10	4	4	2	yes
Sample11	6	6	6	no
Sample12	2	2	9	yes
Sample13	9	9	2	yes

As can be seen in Table I, in seven of thirteen cases, the process resulted in a reclassification using the same algorithm that had classified the original text. This shows that simple phrase replacement does have a detectable effect on authorial styles, the features of which machine learning algorithms depend upon to make correct classifications.

B. Directed style transformation

We would like to examine directed, i.e. biased, stylistic transforms whereby the phrase replacement algorithm specifically tries to bring the style closer to a particular target. We select an arbitrary sample among the sample texts that JGAAP/RN Cross Entropy [6] can classify correctly. This was sample 6 which is written by Author 12. We also select another author, (Author 13) to serve as target corpus. The training documents labeled as author 13 were combined and word frequency statistics extracted as per step 1 of the phrase replacement transform. The Cross Entropy algorithm [6], which performed the best on problem A, compares every single training document with every test document and generates a distance with lower values indicating closer stylistic match.

TABLE II. SAMPLE #6 ATTRIBUTION AND RESULTS

	RNCE dist. with A12	RNCE dist. with A13	Attribution
Sample6-original	**468.742** 478.497 4160.666	492.493 492.653 3475.155	Author 12
Sample6-transformed	464.770 480.477 4088.701	**462.662** 472.736 3331.260	Author 13

The two samples are "sample6" (original test document 6 attributed correctly to Author 12) and "sample6-transformed". The 3 numbers in the cells indicate the RN Cross Entropy distances between the sample and each of the training documents. The algorithm selects the single lowest overall distance in all binary comparisons for attribution. These numbers are boldfaced in the table.

V. Evaluation

The evaluation for this kind of transformation problem is invariably two fold. First we must confirm that a shift in style has taken place, as detectable by some objective critic. Second, we must ensure that the transformed text is actually coherent and meaning-preserving.

The first evaluation is accomplished using JGAAP. We treat JGAAP and the RNCE algorithm as an objective critic and authorship classifier, in fact the single best performing classifier in JGAAP for the given problem. This classifier, for example, has determined that our transformed text is closer to Author 13, than to Author 12.

The second part of the evaluation is rather difficult and subjective. We do not have a precise standard for evaluating the accuracy of these sentences. The original student essays themselves contain some mistakes and awkward language. In addition, for issues of delineating between "style" and "meaning" there is no firm community standard and definition. Some researchers, for example, consider almost any possible rephrasing to be a change of meaning.

We assume a dualistic approach to language and consider a single message to be communicable in a variety of styles, similar to the definition presented in [14] which considers style an "option". We adopt a standard for evaluating the resulting sentences inspired by automatic translation literature. Accordingly we find that the transformed sentences fall into three categories: "correct," "passable," and "not correct". Our own examination indicates the majority of sentences are passable, but future work must involve human critics to make this determination.

For example the following sentence from the sample 6 text discussed above, is considered a good transformation. (The "=>" delineates before and after text).

Work provides more than mere [sustenance => nourishment] however , more importantly it provides an individual with [purpose => aim] in life .

This one indicates a bad transformation, introducing some false or ungrammatical substitutions such as "water line" for "watermark" and "in one case" for "once".

These [tasks => projects] have a [relatively => comparatively] low [watermark => water line] and are clearly delineated [once => in one case] we [achieve => accomplish] them , so [basically => essentially] [once => in one case] we have reached a certain [level => degree] in society.

Given space limitations we cannot reproduce entire texts, but full documents can be shared electronically.

VI. Conclusions and Future Work

Manual style transformation as a scientific process has been attempted [15], but it is painstakingly slow and time consuming. Automatic style transformation, while not rising to the same quality is nevertheless possible and our work suggests it is promising. Given the wide variety of applications from digital forensics, to interactive entertainment, plagiarism evaluation and intelligent writing assistants, there is reason to be optimistic about any progress.

Our goal is to use robust style classification techniques to aid in style transformation, more precisely to discriminate in favor of more effective, directed transformations and to validate results. One such validation is demonstrated in this paper whereby a document attributed to one author by a standard classifier was transformed into a similar document attributed to a different author using the same classifier.

Future work will advance along two tracks. First to utilize more powerful style transforms resulting in desired classifications. Secondly, we want to concentrate on fine-tuning transformation techniques such that the resulting sentences are more accurate and more natural sounding. Popularity of phrases (how common is it to say something in a certain way?) can be a factor in synonym ranking. Such popularity can be measured using large corpora or Internet search engines.

References

[1] F. Mairesse and M. Walker, "A personality-based framework for utterance generation in dialogue applications," Proceedings of the AAAI Spring Symposium on Emotion, Personality, and Social Behavior, Palo Alto, CA, 2008.

[2] C. DiMarco, "Stylistic Choice in Machine Translation," AMAT, 1994.

[3] F. Khosmood and R. Levinson, "Automatic Natural Language Style Classification and Transformation", BCS Corpus Profiling Workshop, London, UK, 2008.

[4] F. Khosmood and R. Levinson, "Toward automated stylistic transformation of natural language text," Digital Humanities, Washington, D.C. , 2009.

[5] P. Juola, "Ad-hoc Authorship Attribution Contest" ACH/ALLC 2004, Gothenberg, Sweden, 2004.

[6] P. Juola, Authorship Attribution, NOW Publishers, 2009.

[7] P. Juola, et. al. JGAAP, a Java-based,modular, program for textual analysis, text categorization, and authorship attribution. http://www.mathcs.duq.edu/~fa05ryan/wiki/index.php/Main_Page , 2009.

[8] NLTK, The Natural Language Tool Kit, project home page: http://www.nltk.org/, 2009.

[9] Kristina Toutanova and Christopher D. Manning, "Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger". In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000), pp. 63-70., 2000.

[10] WordNet at Princeton University Cognitive Science Library, http://wordnet.princeton.edu , 2009.

[11] ConceptNet, MIT Common Sense Computing initiative's ConceptNet (v 3) API. http://csc.media.mit.edu/docs/conceptnet/, 2009.

[12] A. J. Wyner, "Entropy estimation and patterns," in Proceedings of the 1996 Workshop on Information Theory, 1996.

[13] P. Juola, "What can we do with small corpora? Document categorization via cross-entropy," in Proceedings of an Interdisciplinary Workshop on Similarity and Categorization, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, UK, 1997.

[14] J. Walpole "Style as Option," College Composition and Communication, vol. 31, No. 2, Recent Work in Rhetoric: Discourse Theory, Invention, Arrangement, Style, Audience, (May, 1980), pp. 205-212, 1980.

[15] D. L. Hoover, Language and style in The inheritors, University Press of America in Lanham, Md., 1999.