Bioorganic Wiki

Study on Creative Language

Jack Miller, Dr. Brian Goess, Dr. Andrea Tartaro

FURMAN COMPUTER SCIENCE

INTRODUCTION

The Bioorganic wiki is a student-created online textbook written by students in an intermediate-level, college chemistry class. Previous research[1] identified creative linguistic forms in the text such as metaphor, anaphora, and understatement. While this past work looked broadly at creative language within the wiki, this project looks closely at usages of personification. This research investigates personification and how it relates to student understanding of scientific topics.

[1] Tartaro, Andrea, Brian C. Goess, and Mike Winiski. "Creative Language in a Student-generated Bioorganic Chemistry Wiki Textbook." *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, 2015.

PERSONIFICATION

Personification is a specific type of metaphor in which non-human nouns are attributed human-like qualities.

"Of course we can't have **thrombin going around ripping apart every arginine or lysine it sees**."
-author1, Furman Bioorganic Wiki Chapter 9

In this example the serine protease 'thrombin' is attributed with the human verb 'ripping'.

Figure 1: Graphs detailing the total figure of speech additions per author of two wiki chapters.

Figure 1 illustrates the use of different linguistic forms in the wiki. Of the various figure of speech categories we tracked, personification greatly outnumbered the other fourteen categories. Personification usages also survived years of editing rounds. This suggests personification could contribute an important role in student learning.

Sentence Type	Relationship
Type I	Subject IS-A Object
Type II	Verb acting on Noun
Type III	Adjective acting on Noun

Detecting Type I Metaphors[2]

1. Parse sentences to get all R ← {sub, be, object} relations
2. For each R
   1. If Hyponym(sub,obj) = true, R is non metaphoric
   2. Else R is a metaphoric relation
3. All sentences with at least one true relation are metaphoric

Detecting Type II and III Metaphors[2]

1. Parse sentences to get all R ← {word*, noun} relations
2. For each R
   1. Sort nouns *w* by decreasing *p*(*w*|*word*)**. Take the top *k* nouns with probability higher than threshold *T*
   2. If ∃*w* such that *w* is related to *noun* through the hyponym relation in WordNet then *R* is non metaphoric
   3. Else R is a Type II metaphoric relation
3. All sentences with at least one true relation are metaphoric

**Word* refers to the interchangeability of Verb with Type II metaphors and Adjective with Type III metaphors
** *p(w|word)* is the ratio between count of noun used with *word* over the total counts of the noun in the bigram

Determining Parts of Speech

A parts of speech tagger is used to resolve the issues of determining sentence relations.

WordNet Database[3]

WordNet is a database of lexical relations grouped into sets of cognitive synonyms. It gives a hierarchical structure to nouns, verbs, and adjectives expressed in a tree.

Web 1T Corpus[4]

The Web 1T Corpus is a database of n-gram and frequency counts derived from over 1 trillion word tokens from web pages.

References

[2] Krishnakumaran, Saisuresh, and Xiaojin Zhu. "Hunting elusive metaphors using lexical resources." *Proceedings of the Workshop on Computational approaches to Figurative Language*. Association for Computational Linguistics, 2007.

[3] Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

[4] Thomas Brants and Alex Franz. 2006. *Web 1T 5-gram Version 1*. Linguistics Data Consortium, Philadelphia.

Figure 2: Algorithms designed by Krishnakumaran and Zhu along with formal descriptions of their POS tagger, WordNet databases, and Web 1T corpus.

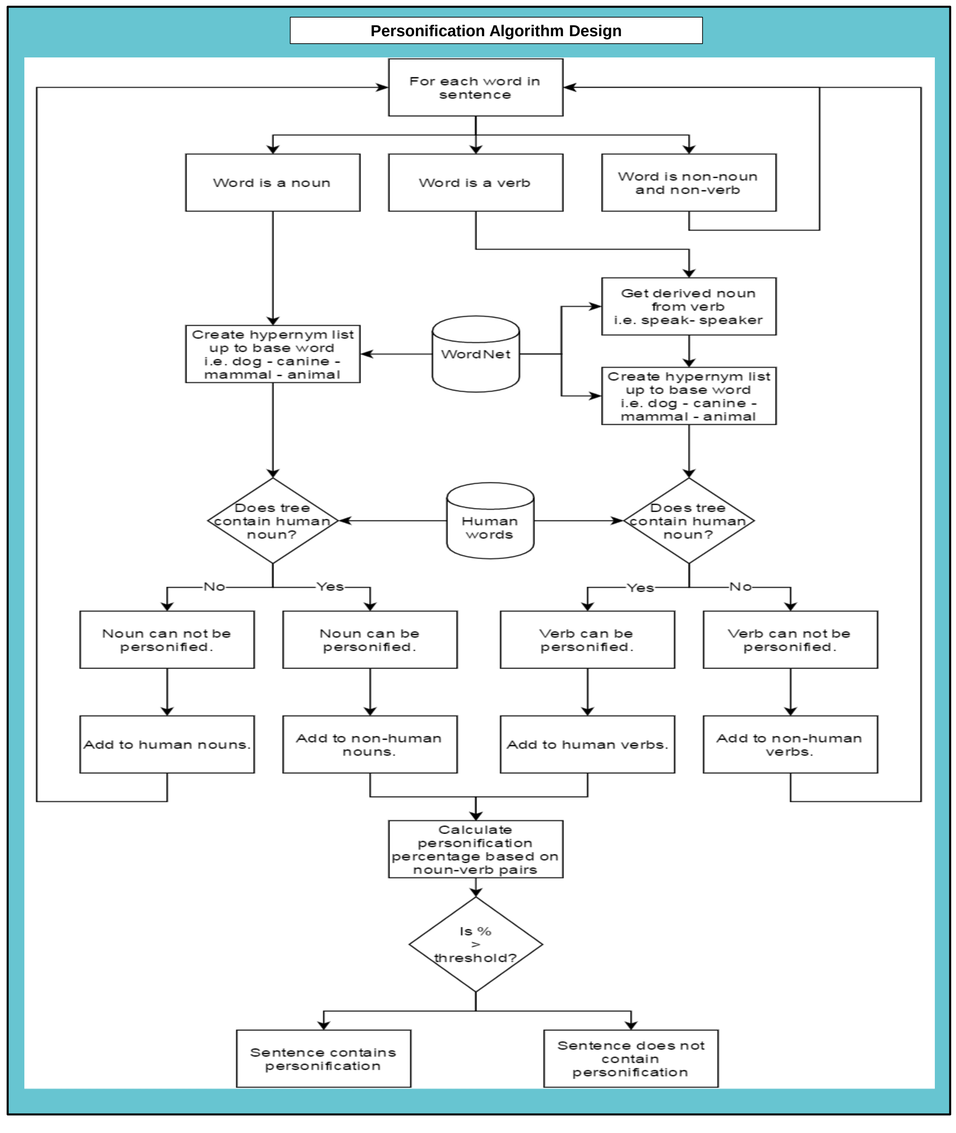

Figure 3: Algorithm designed to detect personification. Various changes to Krishnakumaran and Zhu's algorithm were made such as the removal of the Web1T Corpus.

PROJECT AIMS

1. Develop algorithm to automatically detect personification
2. Determine accuracy of algorithm by testing against known personification, non-personification, and metaphor corpuses

Results

In order to automatically detect personification, this project expanded on the research previously done by Krishnakumaran and Zhu. Their algorithm (Figure 2) for detecting metaphors using WordNet and bigram counts was redesigned specifically for detecting personification. The personification algorithm (Figure 3) results are listed below (Figure 4).

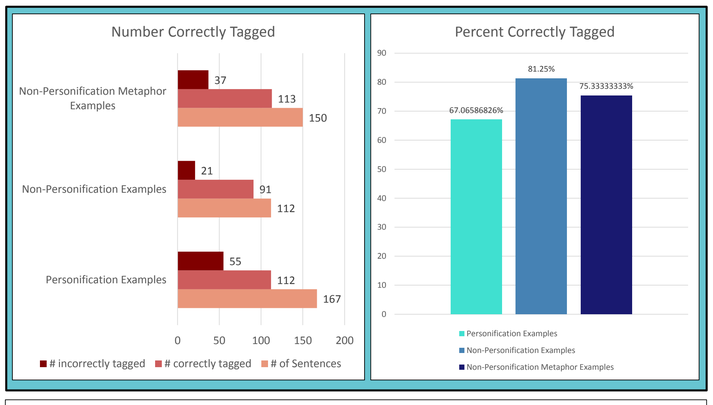

Figure 4: Results table recording the accuracy of the personification algorithm against three personification corpuses and the

ANALYSIS

The algorithm can reliably determine non-personification and metaphoric non-personification but is less accurate at determining whether a sentence contains personification. This stems mostly from the NLP problem of sense disambiguation within the WordNet database.

CONCLUSIONS

Personification's heavy usage within the wiki suggests its importance for student learning. The personification algorithm designed through this research is useful not only in analyzing the most commonly used literary device in the Furman Bioorganic Wiki but also in the general field of natural language processing.

ACKOWLEDGEMENTS

Furman Computer Science Department
Furman Advantage and Hipp Fellowship Programs
Associated Colleges of the South
Furman Chemistry Department