Computational Identification of Conceptual Metaphors in Communities of Blogs

Eric Baumer
Department of Informatics
University of California, Irvine
Irvine, CA 92697-3440
ebaumer@ics.uci.edu

Bill Tomlinson
Department of Informatics
University of California, Irvine
Irvine, CA 92697-3440
wmt@uci.edu

Abstract

This poster presents a computational analysis of conceptual metaphors in a community of political blogs. Like sentiment analysis or opinion extraction, computational metaphor identification can provide a means of understanding the particular framings or conceptualizations used in a community. This poster includes an overview of the implementation and a summary of results.

Keywords

Blogs, metaphor, blog communities, computational metaphor identification.

1. Introduction

As blogs become a more important part of internet culture, news media, and people's daily lives, there is an increasing interest in methods of making sense of this vast and growing body of content. A variety of computational methods have been applied in an attempt to automatically determine topics of conversation, e.g., [1], or opinions expressed, e.g., [9]. Another potential perspective of a community can be provided by examining the conceptual metaphors that it uses. Consider, for example, the language used to describe having an argument. "I *attacked* his *position*." "Your claims are *indefensible*." "She *obliterated* her *opponent*." The words we use to describe an argument evince images of physical combat or war. Lakoff and Johnson [5] argue that this is evidence for the conceptual metaphor that ARGUMENT is WAR. The goal of this poster is to explore the extent to which the set of metaphors used by a community of blogs can be informative about that community's particular views or values.

In previous research, a number of computational approaches have been taken to processing textual metaphor, e.g., [4,6]. These approaches have focused on discerning metaphorical language from literal, and then performing extra or different processing on the metaphorical language in order to compute its literal, "true" meaning. In the *computational metaphor identification* (CMI) approached used here, the goal is not to look at individual text as metaphorical or literal, but to identify underlying conceptual metaphors that permeate a body of text.

2. Implementation Overview

The techniques used in this implementation extend those from CorMet [7]. The main difference from previous work is that CorMet was designed to extract known metaphorical mappings between corpora of pre-determined domains, whereas the techniques presented here are used to identify potential conceptual metaphors in novel target corpora.

This implementation hinges largely on selectional preference learning [8]. For example, the English verbs "eat" or "drink" tend to have a *human*[1] or *animal* as the subject and either *food* or *potable liquid* as the direction object, respectively. The selectional preference strength for a given verb-case slot pair is calculated by taking the relative entropy of the prior distribution over word classes and the posterior distribution conditioned on a given verb-case slot:

$$S_R(p) = \sum P(c|p) \log \frac{P(c|p)}{P(c)}$$

Grammatical relations used to calculate selectional preferences are generated using a typed dependency parser [3]. Selectional preference learning requires classes of words, but the corpus contains word tokens. WordNet synsets are used for word classes, where a single word token counts as a partial observation of any of the synsets it might represent. Selectional preference is the overall choosiness of a given verb-case slot, while selectional association is defined as the strength with which a given verb-case slot select for a given synset (see [8] for details). Selectional associations are calculated for those verbs in a corpus with the highest frequency relative to general English, as derived from the British National Corpus (BNC).

The synsets for which those relatively frequent verbs select are then clustered. Each synset is represented as a vector of selectional associations, where the nth element of the vector is the degree to which the nth verb-case slot selects for that synset. The result is a set of conceptually coherent clusters of synsets. For example, in the domain of a chemistry or biology LAB, the verbs "pour," "flow," "freeze," and "evaporate" all select for *liquids* and *fluids*, so the corpus of documents from this domain would result in a cluster of *fluids* and *liquids*.

The mapping process involves a corpus of documents in a specified source domain and a target corpus in which to identify metaphors. Corpora for source domains are drawn from Wikipedia. Every Wikipedia article belongs to at least one category, and categories are organized into a directed acyclic

---

[1] ALL CAPS are domains, Capital first letters are Wikipedia categories, SMALL CAPS are concepts in a metaphor, and *italics* are specific words that are part of an instance of a metaphor.

Metaphor	Target	Source	verb-case slot	Conf	Example Sentence
TERRORISM is an EVIL ATTACK	terrorism	evil, Satan	fight - prep_against	0.104	"The USA Today report... sparked a renewed debate over government intrusion into Americans' civil liberties in the *fight against terrorism*."
		hit, collision	protect - prep_against	0.036	
		war, hostility, campaign	violate - nsubj	0.002	
		battle, struggle, conflict	violate - nsubj	0.001	
TERRORISM is an EVIL ATTACKER	terrorism	evil, Satan	fight - prep_against	0.104	"US... efforts to *combat terrorism*."
		bandit, brigand	combat - dobj	0.007	
		intruder, stranger	violate - nsubj	0.007	
RELIGION is an ATTACK	religion, faith	intruder, stranger	violate - nsubj	0.018	"His 'faith based' initiatives violate the first [amendment]."
		war, hostility, campaign	violate - nsubj	0.013	
		battle, struggle, conflict	violate - nsubj	0.012	

Table 1 – Metaphors and supporting mappings in the Think Progress community. The first column lists the manual assigned metaphor mapping. The next four columns describe the computationally identified mappings supporting that metaphor: salient term(s) from the target cluster, salient term(s) from the source cluster, the mediation verb-case slot, and the confidence for the mapping. The last column is example sentence from a blog post (emphases added).

graph. The corpus for a domain is defined as all of the Wikipedia pages in a given category, such as Military, and all the pages in all its subcategories, such as Military history.

Mappings are found using selectional associations for clusters across different corpora. The selectional association of a verb-case slot for a cluster is the average of that verb-case slot's selectional association for each member of the cluster. For cluster x in domain A and cluster y in domain B, the polarity of a mapping between them is:

$$\sum_\alpha \Lambda(\alpha, y, B) * 0.75 + \Lambda(\alpha, x, A) * 0.25$$

where α ranges over the verb-case slots that select for x in A, and $\Lambda(\alpha, x, A)$ is the selectional association strength of the verb-case slot α for cluster x in domain A. The overall confidence for a mapping M is:

$$\frac{(vcs(M))}{tot\_vcs} * 0.25 + \frac{pol(M)}{max\_pol} * 0.75$$

where $vcs(M)$ is the number of verb-case slots that mediate a given mapping, max_vcs is the total number of verb-case slots selecting for the two clusters being mapped, $pol(M)$ is the polarity of the given mapping, and max_pol is the maximum polarity of all mappings between the two domains in question.

3. Test Evaluation

This CMI technique was applied to a community of blogs from the 2007 ICWSM dataset [http://www.icwsm.org/data.html]. Starting with Think Progress [http://thinkprogress.org], a community of blogs was identified using [2]. The resulting community from this data set contained 23 blogs with 1950 posts, yielding a corpus of 645,421 words. Mapping were then sought from the MILITARY domain, a corpus of 1867 articles containing 507,589 words.

The first two metaphors shown in table 1, framing TERRORISM as an EVIL ATTACK or an EVIL ATTACKER, are not entirely surprising, given the rhetoric surrounding terrorism in US politics of the

time, especially the "war on terror." The metaphor RELIGION is an ATTACK also makes sense, given that Think Progress is a liberal blog, and most liberals negatively view the conservative value of religion in political leaders. For example, religion is framed as an attack the "violates the first [amendment]." For someone not familiar with the Think Progress community, these metaphors provide a quick introduction to the conceptual framings used in the community's discussions. The results demonstrate that CMI can effectively identify sensible and meaningful conceptual metaphors.

4. References

[1] E. Baumer and D. Fisher, "Smarter Blogroll: An Exploration of Social Topic Extraction for Manageable Blogrolls," *Hawai'i Int'l Conf on System Sciences*, 2008.

[2] J. Bulters and M. de Rijke, "Discovering Weblog Communities: A Content- and Topology-Based Approach," *Int'l Conf on Weblogs and Social Media*, March 2007.

[3] M. de Marneffe, B. MacCartney, and C. Manning, "Generating Typed Dependency Parses from Phrase Structure Parses," *Language Resources and Evaluation (LREC 2006)*, 2006.

[4] D. Fass, "Met*: A method for discriminating metonymy and metaphor by computer," *Computational Linguistics*, vol. 17, pp. 49-90, 1991.

[5] G. Lakoff and M. Johnson, *Metaphors We Live By*. Chicago, IL: University of Chicago Press, 1980.

[6] J. Martin, "Metabank: A knowledge base of metaphoric language conventions," *Computational Intelligence*, vol. 10, pp. 134-149, 1994.

[7] Z. Mason, "CorMet: a computational, corpus-based conventional metaphor extraction system," *Computational Linguistics*, vol. 30, pp. 23-44, March 2004.

[8] P. Resnik, *Selection and Information: A Class-Based Approach to Lexical Relationships*. Univ of Pennsylvania, Department of Computer and Information Science, 1993.

[9] K. Yang et al., "Fusion Approach to Finding Opinions in Blogosphere," *Int'l Conf on Weblogs and Social Media*, March 2007.