Monolingual Distributional Similarity for Texty Text Generation

Juri Ganitkevitch, Benjamin Van Durme and **Chris Callison-Burch**
Department of Computer Science, Johns Hopkins University
Baltimore, MD 21218, USA

Abstract

Previous work obtained collections of paraphrases by either relying on sentence-aligned parallel datasets, or by using distributional similarity metrics over large text corpora. Our approach combines these two orthogonal sources of information by directly integrating them into the decoding algorithm. We hope to report significant improvements in output quality on an array of text-to-text generation tasks.

1 Introduction

A wide variety of tasks in text-to-text generation can straightforwardly be cast as paraphrasing problems. Paraphrasing techniques have successfully been applied to applications such as document summarization (Barzilay et al., 1999; Barzilay, 2003), text simplification, sentence compression, as well as information retrieval tasks like query expansion and question generation (McKeown, 1979; Anick and Tipirneni, 1999; Ravichandran and Hovy, 2002; Riezler et al., 2007).

A major challenge in building a paraphrase-based text-to-text generation system is the *extraction* (and scoring) of a set of paraphrases (a *paraphrase table* or *grammar*) from data. In the past, each paraphrase extraction approach has focussed on a single type of dataset. Methods based on sentence-aligned parallel corpora, both monolingual and bilingual, successfully leveraged the equivalence in meaning implied by the parallelism of the data (Barzilay and McKeown, 2001; Pang et al., 2003; Bannard and Callison-Burch, 2005; Madnani et al., 2007; Cohn and Lapata, 2008; Zhao et al., 2008).

Approaches drawing from plain monolingual text, on the other hand, rely on distributional similarity metrics as their semantic equivalency cue. These methods use the vast amounts of data available to them to make up for the noise in the resulting signal (Lin and Pantel, 2001; Bhagat and Ravichandran, 2008).

However, alignments and distributional similarity are based on orthogonal information. In fact, previous work on both machine translation and paraphrasing has shown that adding features based on distributional similarity yields significant improvements over their alignment-based baseline systems (Chan et al., 2011; Klementiev et al., 2012).

Expanding on these impressive improvements over purely alignment-based baselines, we present a the deeper integration of distributional similarity measures into our paraphrasing system. More precisely, we present the following contributions:

- We define the notion of distributional similarity for paraphrase patterns that contain multi-word gaps. This generalizes over previous approaches that defined the notion for pairs of contiguous phrases (Chan et al., 2011), and single-word gaps (Lin and Pantel, 2001; Bhagat and Ravichandran, 2008). We extend a previously presented rich SCFG-based paraphrase extraction method (Ganitkevitch et al., 2011) to compute distributional signatures for the paraphrases.

- We present a text-to-text generation approach that integrates similarity metrics directly into the decoding procedure. Contextual similarity is evaluated for both the paraphrase patterns applied in the text-to-text derivation and contiguous target-side phrases generated when combining rules in our paraphrastic SCFG.

- Finally, we compare the effectiveness of out method against a variety of baselines on an example text-to-text generation task, sentence compression. We show improvements in quality over both a purely alignment-based paraphrasing system and an ILP-based compression model.

2 Notes

What is our claim? Monolingual distributional similarity and pivot-based methods are orthogonal sources of information. Combining them yields improvements.

How do we implement it? Need to conceptually separate decoding from grammar extraction (to me both are "from MT", but here they are separate things:

How do we back it up? Show results wherein the combination improves over both.

3 Todo

Write up what I've done so far:

- Similarity features, scores, setup
- Subtleties and "design decisions" in scoring the rules

4 Model

Outer and inner context matching? Think about how differently defined contexts can combine in SCFG derivations, go from there.

Sparsity is an issue.

5 Outline

Main contribution: tight integration of monolingual similarity into both paraphrase extraction and generation (decoding).

Challenges: stateful feature, efficient extraction of signatures, compare (?) to phrase-to-phrase similarity as a feature.

Comparisons: no MonoDS, reranking approaches.

Statistic: how many pairs of phrases do we see in decoding?

Review paraphrase-based text-to-text generation.

Highlight the ability to incorporate feature functions.

Previously: pivot-derived feature functions and a language model.

Papers and stuff:

(Langkilde and Knight, 1998b)

(Langkilde and Knight, 1998a)

References

Peter G. Anick and Suresh Tipirneni. 1999. The paraphrase search assistant: terminological feedback for iterative information seeking. In *Proceedings of SIGIR*.

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of ACL*.

Regina Barzilay and Kathleen McKeown. 2001. Extracting paraphrases from a parallel corpus. In *Proceedings of ACL*.

Regina Barzilay, Kathleen R. McKeown, and Michael Elhadad. 1999. Information fusion in the context of multi-document summarization. In *Proceedings of ACL*.

Regina Barzilay. 2003. *Information Fusion for Mutli-document Summarization: Paraphrasing and Generation*. Ph.D. thesis, Columbia University, New York.

Rahul Bhagat and Deepak Ravichandran. 2008. Large scale acquisition of paraphrases for learning surface patterns. In *Proceedings of ACL/HLT*.

T.P. Chan, C. Callison-Burch, and B. Van Durme. 2011. Reranking bilingually extracted paraphrases using monolingual distributional similarity. In *EMNLP Workshop on GEMS*.

Trevor Cohn and Mirella Lapata. 2008. Sentence compression beyond word deletion. In *Proceedings of the COLING*.

Juri Ganitkevitch, Chris Callison-Burch, Courtney Napoles, and Benjamin Van Durme. 2011. Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation. In *Proceedings of EMNLP*.

Alexandre Klementiev, Ann Irvine, Chris Callison-Burch, and David Yarowsky. 2012. Secret paper title. In *Secret Conference*.

I. Langkilde and K. Knight. 1998a. Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 704–710. Association for Computational Linguistics.

I. Langkilde and K. Knight. 1998b. The practical value of n-grams in generation.

Dekang Lin and Patrick Pantel. 2001. Discovery of inference rules from text. *Natural Language Engineering*.

Nitin Madnani, Necip Fazil Ayan, Philip Resnik, and Bonnie Dorr. 2007. Using paraphrases for parameter tuning in statistical machine translation. In *Proceedings of WMT07*.

Kathleen R. McKeown. 1979. Paraphrasing using given and new information in a question-answer system. In *Proceedings of ACL*.

Bo Pang, Kevin Knight, and Daniel Marcu. 2003. Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences. In *Proceedings of HLT/NAACL*.

Deepak Ravichandran and Eduard Hovy. 2002. Learning sufrace text patterns for a question answering system. In *Proceedings of ACL*.

Stefan Riezler, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu Mittal, and Yi Liu. 2007. Statistical machine translation for query expansion in answer retrieval. In *Proceedings of ACL*.

Shiqi Zhao, Haifeng Wang, Ting Liu, and Sheng Li. 2008. Pivot approach for extracting paraphrase patterns from bilingual corpora. In *Proceedings of ACL/HLT*.