Word Sense Subjectivity for Cross-lingual Lexical Substitution

Fangzhong Su
School of Computing
University of Leeds, UK
scsfs@leeds.ac.uk

Katja Markert
School of Computing
University of Leeds, UK
scskm@leeds.ac.uk

Abstract

We explore the relation between word sense subjectivity and cross-lingual lexical substitution, following the intuition that good substitutions will transfer a word's (contextual) sentiment from the source language into the target language. Experiments on English-Chinese lexical substitution show that taking a word's subjectivity into account can indeed improve performance. We also show that just using word sense subjectivity can perform as well as integrating fully-fledged fine-grained word sense disambiguation for words which have both subjective and objective senses.

1 Introduction

Cross-lingual lexical substitution has been proposed as a Task at SemEval-2010.[1] Given a target word and its context in a source language (like English), the goal is to provide correct translations for that word in a target language (like Chinese). The translations must fit the given context.

In this paper, we explore the relation between the sentiment of the used word in the source language and translation choice in the target language, focusing on English as the source and Chinese as the target language. Our work is motivated by the intuition that most good word translations will be *sentiment-invariant*, i.e. if a source word is used in a subjective (opinion-carrying) sense it will be often translated with a subjective sense in the target language whereas if it used in an objective sense, it will be

[1] http://lit.csci.unt.edu/index.php/Semeval_2010

translated with an objective sense. As an example, consider the two words *positive* and *collaborate* with example senses from WordNet 2.0 below.

(1) positive—greater than zero; "positive numbers" (*objective*)

(2) plus, positive—involving advantage or good; "a plus (or positive) factor" (*subjective*)

(3) collaborate, join forces, cooperate—work together on a common enterprise of project; "We joined forces with another research group"(*objective*)

(4) collaborate—cooperate as a traitor; (*subjective*)

In most cases, if the word *positive* is used in the sense *"greater than zero"* (**objective**) in an English context, the corresponding Chinese translation is "正的"; if *"involving advantage or good"*(**subjective**) is used, its Chinese translations are "积极的, 好的". Similarly, for the word *collaborate*, the sense *"work together on a common enterprise of project"* (**objective**) corresponds to "合作, 协作" in Chinese translation, and *"cooperate as a traitor"* (**subjective**) corresponds to "勾结, 狼狈为奸". Therefore, subjectivity information should be effective for improving lexical translation for what we previously (Su and Markert, 2008) termed *subjectivity-ambiguous* words, i.e. words with both subjective and objective senses such as *positive* and *collaborate* above.

We therefore incorporate subjectivity word sense disambiguation (SWSD) as defined in Akkaya et al. (2009) into lexical substitution. SWSD is a binary classification task that decides in context whether a word occurs with one of its subjective or one of its objective senses. In contrast to standard

multi-class Word Sense Disambiguation (WSD), it uses a coarse-grained sense inventory that allows to achieve higher accuracy than WSD and therefore introduces less noise when embedded in another task such as word translation. For example, the accuracy reported in Akkaya et al. (2009) for SWSD is over 20% higher than for standard WSD. Coarse-grained senses are also easier to annotate, so getting training data for learning is less arduous. On the minus side, SWSD can only be useful for subjectivity-ambiguous words. However, we showed (Su and Markert, 2008) that subjectivity-ambiguity is frequent (around 30% of common words).

2 Related Work

McCarthy and Navigli (2007) organized a monolingual English lexical substitution task in Semeval-2007, i.e finding English substitutions for an English target word. Mihalcea et al. organize an English-Spanish lexical substitution task in SemEval-2010. Approaches to lexical substitution in the past competitions did not use sentiment features.

Independent of these lexical substitution tasks, the connection between word senses and word translation has been explored in Chan et al. (2007) and Carpuat and Wu (2007), who predict the probabilities of a target word being translated as an item in a "sense inventory", where the sense inventory is a list of possible translations. They then incorporate these probabilities into machine translation. However, they do not consider sentiment explicitly.

Subjectivity at the word sense level has been discussed by (Wiebe and Mihalcea, 2006; Su and Markert, 2008; Akkaya et al., 2009). Wiebe and Mihalcea (2006) and Su and Markert (2008) both show that this is a well-defined concept via human annotation as well as automatic recognition. Akkaya et al. (2009) show that subjectivity word sense disambiguation (SWSD) can boost the performance of a sentiment analysis system. None of these paper considers the impact of word sense subjectivity on cross-lingual lexical substitution.

3 Methodology

3.1 Task and Dataset

We constructed an English-Chinese lexical substitution gold standard by translating the English tar-

get words in the SENSEVAL 2 and SENSEVAL 3 lexical sample training and test sets into Chinese. We choose the SENSEVAL datasets as they are relatively domain-independent and also because we can use them for our SWSD/WSD subtasks as well. The translation is carried out by two native Chinese speakers with a good command of English. First, candidate Chinese translations (denoted by **T**) of the English target words are provided from the on-line English-Chinese dictionary *iciba*[2], which is composed of more than 150 different English-Chinese dictionaries. To reduce annotation bias, the order of the Senseval sentences is randomized. The annotators then independently assign the most fitting Chinese translation(s) (from **T**) for the English target words in the given Senseval sentences. For the agreement study, different Chinese translations (for example, "权威" and "泰斗" of the word *authority*) that are actually synonyms are merged. The observed agreement between the two annotators is 86.7%. Finally, the two annotators discuss the disagreed examples together, leading to a gold standard.

Since we evaluate how word sense subjectivity affects cross-lingual lexical substitution, we limited our study to the SENSEVAL words that are subjectivity-ambiguous. Therefore, following the annotation schemes in (Su and Markert, 2008; Wiebe and Mihalcea, 2006), all senses of all target words in SENSEVAL 2&3 are annotated by a near-native English speaker as *subjective* or *objective*. This annotator was not involved in the English to Chinese translation. We also discard subjectivity-ambiguous words if its *subjective* or *objective* senses do not appear in both training and test set. In total we collect 28 subjectivity-ambiguous words. Their English example sentences and translations yield 2890 training sentence pairs and 1444 test sentence pairs.

3.2 Algorithms

For the English-Chinese lexical substitution task, we first develop a basic system (called **B**) to assign Chinese translations to the target English words in context. This system uses only standard contextual features from the English sentences (see Section 3.3). We then add word sense subjectivity information to

---

[2]http://www.iciba.com

358

the basic system (see Section 3.4). We also compare including word sense subjectivity to the inclusion of full fine-grained sense information (Section 3.5).

All systems are supervised classifiers trained on the SENSEVAL training data and evaluated on the SENSEVAL test data for each of the 28 words. We employ an SVM classifier from the `libsvm` package[3] with a linear kernel.

3.3 Common Features

In the basic system **B**, we adopt features which are commonly used in WSD or lexical translation.

Surrounding Words: Lemmatized bag of words with stop word filtering.

Part-of-Speech (POS): The POS of the neighbouring words of the target word. We extract POS tag of the 3 words to the right and left together with position information.

Collocation: The neighbouring words of the target word. We extract 4 lemmatized words to the right and left, together with position information.

Syntactic Relations: We employ the MaltParser[4] for dependency parsing and extract 4 features: the head word of the target word, POS of the head word, the dependency relation between head word and target word, and the relative position (left or right) of the head word to the target word.

3.4 Subjectivity Features

We add a feature that incorporates whether the original English word is used subjectively or objectively. For an upper bound, we use the SENSEVAL gold standard sense annotation (**gold-subj**), mapped onto binary subjective/objective labels. For a more realistic assessment, we use SWSD to derive the subjectivity sense label automatically (**auto-subj**) using standard supervised binary SVMs and the features in Section 3.3 on the SENSEVAL data.

3.5 Sense Features

We compare using subjectivity information to using full fine-grained word sense information, incorporating a feature that specifies the exact word sense of the target word to be translated. This setting

also compares the SENSEVAL gold standard (**gold-senses**) and automatically predicted sense information (**auto-senses**), the latter via supervised multiclass learning on the SENSEVAL dataset.

4 Experiments and Evaluation

For the English-Chinese lexical substitution task, we evaluate 6 different methods: **Baseline** (assign the most frequent translation to all examples), **B** (use common features), **B+gold subj** (incorporate gold standard word sense subjectivity), **B+gold sense** (incorporate gold standard sense), **B+auto subj** (incorporate automatically predicted word sense subjectivity), and **B+auto sense** (incorporate automatically predicted fine-grained senses). We measure lexical substitution accuracy on the SENSEVAL test data by comparing to the human gold standard annotation (see Section 3.1). Results are listed in Table 1.

Results. Table 1 shows that our standard lexical substitution system **B** improves strongly (near 11% average accuracy gain) over the most frequent translation baseline. Incorporating sense subjectivity as in **B+gold subj** leads to a further strong improvement, confirming our hypothesis that word sense subjectivity can improve lexical substitution. Incorporating fine-grained senses **B+gold senses** yields only a slightly higher gain, showing that a coarse-grained subjective/objective classification might be sufficient for subjectivity-ambiguous words for aiding translation. In addition, the small gain using fine-grained senses might disappear in practice as automatic WSD is a more challenging task than SWSD: in our experiment, **B+auto sense** performs worse than **B+auto subj**. The current improvement of **B+auto subj** over **B** is significant (McNemar test at the 5% level). The difference between the actual performance of word sense subjectivity and its potential as exemplified in **B+gold subj** is, obviously, caused by imperfect performance of the SWSD component, mostly due to a distributional bias in the SENSEVAL training data, with few examples for rarer senses of the target words.

For some words (such as *authority* and *stress*), the additional sense subjectivity feature does not improve lexical substitution, even when gold standard labels are used. There are two main reasons for this. First, one candidate Chinese translation might cover

Table 1: Accuracy of lexical substitution with different different feature settings

Word	Subjectivity of Senses	Baseline	Basic (B)	B+gold subj	B+gold senses	B+auto subj	B+auto senses
authority	3-S 4-O	50.5%	70.3%	70.3%	84.6%	70.3%	79.1%
blind	2-S 1-O	87.0%	88.9%	94.4%	94.4%	88.9%	88.9%
cool	3-S 3-O	46.0%	46.0%	68.0%	68.0%	58.0%	48.0%
dyke	1-S 1-O	89.3%	89.3%	92.9%	92.9%	89.3%	89.3%
fatigue	1-S 2-O 1-B	80.0%	80.0%	82.5%	85.0%	82.5%	82.5%
fine	5-S 4-O	78.5%	78.5%	90.8%	80.0%	80.0%	78.5%
nature	1-S 3-O 1-B	53.3%	62.2%	73.3%	71.1%	64.4%	62.2%
oblique	1-S 1-O	65.5%	75.9%	86.2%	89.7%	79.3%	79.3%
sense	3-S 2-O	47.5%	67.5%	77.5%	77.5%	75.0%	72.5%
simple	2-S 2-O 1-B	71.2%	71.2%	75.8%	74.2%	72.7%	71.2%
stress	3-S 2-O	92.1%	92.1%	92.1%	92.1%	92.1%	92.1%
collaborate	1-S 1-O	90.0%	90.0%	93.3%	93.3%	93.3%	90.0%
drive	3-S 5-O 1-B	51.4%	78.4%	89.2%	86.5%	83.8%	78.4%
play	4-S 13-O 1-B	23.3%	40.0%	48.3%	56.7%	41.7%	43.3%
see	7-S 11-O	30.9%	36.8%	58.8%	61.8%	42.6%	38.2%
strike	3-S 10-O 1-B	20.5%	27.3%	43.2%	45.5%	29.5%	38.6%
treat	2-S 4-O	36.4%	61.4%	65.9%	81.8%	56.8%	65.9%
wander	1-S 2-O 1-B	79.2%	81.3%	83.3%	83.3%	81.3%	81.3%
work	2-S 9-O 2-B	56.8%	56.8%	75.0%	75.0%	63.6%	61.4%
appear	1-S 2-O	42.7%	63.4%	80.2%	90.8%	65.6%	66.4%
express	2-S 2-O	81.5%	81.5%	90.7%	88.9%	83.3%	81.5%
hot	3-S 4-O 1-B	85.0%	85.0%	85.0%	85.0%	85.0%	85.0%
image	3-S 4-O	56.7%	83.6%	94.0%	92.5%	85.1%	79.1%
interest	2-S 4-O 1-B	38.7%	73.1%	84.9%	88.2%	74.2%	71.0%
judgment	4-S 3-O	46.9%	65.6%	78.1%	75.0%	68.8%	62.5%
miss	3-S 5-O	50.0%	63.3%	70.0%	66.7%	63.3%	60.0%
solid	4-S 10-O	40.0%	40.0%	44.0%	48.0%	44.0%	44.0%
watch	3-S 4-O	86.3%	86.3%	90.2%	88.2%	86.3%	86.3%
AVERAGE		**57.4%**	**68.5%**	**77.9%**	**80.2%**	**70.7%**	**70.1%**

both subjective and objective uses of the word. For example, both the objective sense (*"physics force that produces strain on a physical body"*) and subjective senses (*"difficulty that causes worry or emotional emotional tension"* and *" a state of mental or emotional strain or suspense"*) of *stress* are often translated as "压力" in Chinese. Second, in some cases, subjectivity word sense disambiguation is too coarse-grained and finer-grained WSD is actually necessary. For example, the subjective usages of *authority* in SENSEVAL examples are often translated as "专家, 权威", "自信" or "可信" (called **List-S**), and objective usages are often translated as "局, 部","当局","权力, 职权" or "授权, 批准" (called **List-O**). In this case, word sense subjectivity might help to distinguish **List-S** from **List-O**, but not among the candidate translations within a single list.

5   Discussion

We tackle cross-lingual lexical substitution as a supervised task, using sets of manual translations for a target word as training data even for baseline system **B**. However, we do not necessarily need dedicated human translated data as we could also use existing parallel texts in which the target word occurs. Therefore, we think that a supervised approach to lexical

substitution is feasible. However, we do need additional monolingual sense-tagged data in the source language for incorporating our word sense subjectivity features.[5] Although a disadvantage, more and more sense-tagged data does become available (such as OntoNotes). We also only need tagging at a coarse-grained sense level, which is much easier to create than fine-grained data.

6   Conclusion and Future Work

We investigate the relation between word sense subjectivity and cross-lingual lexical substitution. The experimental results show that incorporating word sense subjectivity into a standard supervised classification model yields a significantly better performance for an English-Chinese lexical substitution task. We also compare the effect of sense subjectivity to the effect of fine-grained sense information on lexical substitution. The differences between the two methods turn out to be small, making a case for the "easier", coarse-grained SWSD over WSD for subjectivity-ambiguous words. Future work will widen the study by (i) looking at a wider range of words and languages, (ii) improving automatic SWSD results for better application and (iii) integrating unsupervised subjectivity features into cross-lingual lexical substitution.

References

Cem Akkaya, Janyce Wiebe, and Rada Mihalcea. 2009. Subjectivity Word Sense Disambiguation. *Proceedings of EMNLP'09*.

Marine Carpuat and Dekai Wu. 2007. Improving Statistical Machine Translation Using Word Sense Disambiguation. *Proceedings of EMNLP'07*.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word Sense Disambiguation Improves Statistical Machine Translation. *Proceedings of ACL'07*.

Diana McCarthy and Roberto Navigli. 2007. SemEval-2007 Task 10: English Lexical Substitution Task. *Proceedings of SemEval-2007*.

Fangzhong Su and Katja Markert. 2008. From Words to Senses: A Case Study in Subjectivity Recognition. *Proceedings of COLING'08*.

Janyce Wiebe and Rada Micalcea. 2006. Word Sense and Subjectivity. *Proceedings of ACL'06*.

---

[5]In our case, this is the same data as the data the lexical substitution algorithms are trained on, but this is not mandatory.