SemEval-2 Task 15: Infrequent Sense Identification for Mandarin Text to Speech Systems

Peng Jin[1] and Yunfang Wu[2]
[1]Laboratory of Intelligent Information Processing and Application, Leshan Normal University, Leshan China
[2]Institute of Computational Linguistics Peking University, Beijing China
{jandp, wuyf}@pku.edu.cn

1 Introduction

There are seven cases of grapheme to phoneme in a text to speech system (Yarowsky, 1997). Among them, the most difficult task is disambiguating the homograph word, which has the same POS but different pronunciation. In this case, different pronunciations of the same word always correspond to different word senses. Once the word senses are disambiguated, the problem of GTP is resolved.

There is a little different from traditional WSD, in this task two or more senses may correspond to one pronunciation. That is, the sense granularity is coarser than WSD. For example, the preposition "为" has three senses: sense1 and sense2 have the same pronunciation {wei 4}, while sense3 corresponds to {wei 2}. In this task, to the target word, not only the pronunciations but also the sense labels are provided for training; but for test, only the pronunciations are evaluated. The challenge of this task is the much skewed distribution in real text: the most frequent pronunciation occupies usually over 80%.

In this task, we will provide a large volume of training data (each homograph word has at least 300 instances) accordance with the truly distribution in real text. In the test data, we will provide at least 100 instances for each target word. The senses distribution in test data is the same as in training data.All instances come from People Daily newspaper (the most popular newspaper in Mandarin). Double blind annotations are executed manually, and a third annotator checks the annotation.

2 Participating Systems

Two kinds of precisions are evaluated. One is micro-average:

$$P_{mir} = \sum_{i=1}^{N} m_i / \sum_{i=1}^{N} n_i$$

N is the number of all target word-types. m_i is the number of labeled correctly to one specific target word-type and n_i is the number of all test instances for this word-type. The other is macro-average:

$$P_{mar} = \sum_{i=1}^{N} p_i / N\ ,\ \ p_i = m_i / n_i$$

There are two teams participated in and submitted nine systems. Table 1 shows the results, all systems are better than baseline (Baseline is using the most frequent sense to tag all the tokens).

System	Micro-average	Macro-average
156-419	0.974432	0.951696
205-332	0.97028	0.938844
205-417	0.97028	0.938844
205-423	0.97028	0.938844
205-425	0.97028	0.938844
205-424	0.968531	0.938871
156-420	0.965472	0.942086
156-421	0.965472	0.94146
156-422	0.965472	0.942086
baseline	0.923514	0.895368

Table 1: The scores of all participating systems

References

Yarowsky, David. 1997. "Homograph disambiguation in text-to-speech synthesis." In van Santen, Jan T. H.; Sproat, Richard; Olive, Joseph P.; and Hirschberg, Julia. Progress in Speech Synthesis. Springer-Verlag, New York, 157-172.

Proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010, page 87,
Uppsala, Sweden, 15-16 July 2010. ©2010 Association for Computational Linguistics