Resolving Translation Ambiguity and Target Polysemy in Cross-Language Information Retrieval+

Hsin-Hsi Chen* Guo-Wei Bian* and Wen-Cheng Lin*

Abstract

This paper deals with translation ambiguity and target polysemy problems together. Two monolingual balanced corpora are employed to learn word co-occurrence for the purpose of translation ambiguity resolution and augmented translation restrictions for that of target polysemy resolution. Experiments show that the model achieves 62.92% monolingual information retrieval, which is 40.80% better than that of the select-all model. When target polysemy resolution is added, the retrieval performance represents approximately a 10.11% increase over that of the model which resolves translation ambiguity only.

Keywords: Cross-language information retrieval, Query translation, Translation ambiguity, Target polysemy, Augmented translation restriction

1. Introduction

Cross language information retrieval (CLIR) [Oard and Dorr, 1996; Oard, 1997]deals with the use of queries in one language to access documents in another. Due to differences between the source and target languages, query translation is usually employed to unify the languages in queries and documents. In query translation, translation ambiguity is a basic problem to be resolved. A word in a source query may have more than one sense. Word sense disambiguation identifies the correct sense of each source word, and lexical selection translates it into the corresponding target word. The above procedure is similar to the lexical choice operation in a traditional machine translation (MT) system. However, there is a significant difference between the applications of MT and CLIR. In MT, readers interpret the translated results. If the target word has more than one sense, readers can disambiguate its meaning automatically. In CLIR, however, the translated

---