

ỨNG DỤNG LIÊN KẾT TỪ VÀO VIỆC KHẢO SÁT TRẬT TỰ TỪ ANH-VIỆT

Phạm Phú Hội, Ngô Quốc Hưng, Nguyễn Đăng Hải
Khoa CNTT, Trường Đại học Khoa học Tự Nhiên - ĐHQG tp.HCM

Tóm tắt:

Trong hệ dịch tự động, khi ta xác định được nghĩa của các từ trong câu ngôn ngữ và ghép các nghĩa đó lại với nhau thì chưa hẳn ta được câu dịch hoàn chỉnh. Do đó việc chuyển đổi trật tự giữa các từ trong ngôn ngữ nguồn sao cho khi gắn nghĩa vào các từ đó để được câu dịch hoàn chỉnh là việc làm cần thiết và không thể tránh khỏi. Từ trước đến nay để chuyển đổi người ta chủ yếu dựa vào các quy luật được rút ra một cách thủ công từ các câu song ngữ của hai ngôn ngữ đó. Tuy nhiên, công việc này chắc chắn phải mất nhiều thời gian và công sức mà lại không thể tìm ra hết các quy luật chuyển đổi giữa hai ngôn ngữ đó. Đặc biệt là trong song ngữ Anh-Việt vốn đã khác nhau về loại hình.

Trong bài báo này, chúng tôi sẽ trình bày một mô hình để chuyển đổi giữa các từ trong câu tiếng Anh sao cho phù hợp với câu tiếng Việt một cách tự động. Mô hình được thực hiện trên một ngữ liệu lớn đã được liên kết từ nhằm rút ra các quy luật dịch chuyển trật tự từ khi chuyển ngữ từ tiếng Anh sang tiếng Việt.

Để thực hiện được công việc này, chúng tôi tạo ra các khuôn luật và nhờ quá trình học trên tập ngữ liệu lớn đã được liên kết sẵn để rút ra các luật chuyển đổi một cách tự động dựa vào các khuôn luật đó. Các luật này tuân tự được áp dụng lên câu ngôn ngữ nguồn để chuyển đổi trật tự các từ trong chúng. Ngữ liệu song ngữ để học ở đây là một cơ sở dữ liệu được nhập từ các sách song ngữ về khoa học kỹ thuật gồm khoảng 100.000 câu song ngữ Anh-Việt và đã được liên kết và gán nhãn từ loại nhờ chương trình liên kết từ (Word-Alignment) nhằm xác định mối quan hệ giữa các từ để chuyển đổi.

Chúng tôi đã tiến hành trong ngữ liệu nói trên và kết quả là hệ thống đã đổi trật tự từ chính xác khoảng 94%. Chúng tôi đã áp dụng kết quả này vào việc gắn nghĩa tiếng Việt cho từ tiếng Anh nhằm tạo ra câu dịch tiếng Việt hoàn chỉnh.

APPLYING WORD ALIGNMENT INTO INVESTIGATING ENGLISH-VIETNAMESE WORD ORDER

Pham Phu Hoi, Ngo Quoc Hung, Nguyen Dang Hai

Department of Information Technology, University of Natural Sciences -

VNU.HCM

Abstract:

In the automatically translation system, as we identified the word meanings in bilingual sentence pairs it doesn't mean that we will have a fully done translated sentence. Thus, word order transfer among words in source sentence is essential and indispensable work. As we know so far, to transfer word orders people usually base on rules which are retrieved by hand from bilingual sentence pairs. However, it is sure that this work will take a long time to be done and maybe we can't find all transfer rules between two languages. Especially in English-Vietnamese languages which are different in morphology.

In this paper we will present a word order transfer model to transfer words in English sentence so that it corresponds to Vietnamese sentence automatically. This model is applied to a large corpus which is aligned in word level to retrieve word order transfer rules from English to Vietnamese.

To do so, we made templates and use them to learn on large corpus aligned to retrieve transfer rules automatically. These rules will be applied into source sentences in order to transfer words in them. Bilingual corpus which is used to learn is a database input from bilingual technological science books consisted about 100,000 bilingual sentences in English-Vietnamese. Every bilingual sentence pair is aligned and labeled POS by Word Alignment Project to identify the relation among words to transfer.

We processed on this corpus and our model performs the precision of about 94 percent. We also applied this result into labeling Vietnamese meaning for English words to obtain fully done Vietnamese sentence