Theme: Natural Language Processing

Topic: Cross-lingual and cross-culture low resource transfer learning for multilingual dialogue system

Background

Dialogue system has been widely adopted in various domains such as e-commerce, finance, IT services and public services. Especially along with the maturity of translation and multilingual NLP technologies, multilingual dialogue system has gained more and more attention.

However, multilingual dialogue system suffers from the resource imbalance of different languages. Compared to rich resource languages like English, Chinese, Spanish, French, low resource language may be ten times and even hundred times smaller in both unsupervised and supervised data. Hence, in order to fast expand dialogue system to various low resource languages, it would be difficult to collect and annotate enough data like rich resource languages. How to efficiently reuse resources accumulated in rich resource languages to improve performance of other languages is a key problem to setup and operate a multilingual system.

Besides, dealing with colloquial and cultural specific expression has long been a well-known problem for dialogue system. It has huge impact on the generalization and robustness of dialogue system especially in multilingual setting. It has been shown that Singlish and Taglish can differ a lot from American English on phrase and word order. Overcoming the problem of cross-culture transfer would significantly improve the experience of local users.

Target

- An approach of transfer learning from rich resource language to low resource language. Achieve SOTA results on public dataset XNLI and XTREME.
- Establish dataset of cross-culture transfer learning, target languages are

Singlish, Taglish, traditional Chinese Taiwan, Cantonese. And a transfer learning approach to significantly improve performance over baseline methods.

- Publish paper on top NLP conference

Related Research Topics

- Transfer learning
- Few-shot learning
- Multilinguality
- Dialogue system
- Cross-lingual
- Cross-culture

Suggested Collaboration Method

AIR (Alibaba Innovative Research), one-year collaboration project.