

Multi Language Support for Virtual Assistants

Prise en charge multilingue pour les assistants virtuels

对虚拟助手的多语言支持

پشتیبانی چند زبانه برای دستیاران مجازی

Supporto multilingue per assistenti virtuali

仮想アシスタントの多言語サポート

Soporte multilinguaje para asistentes virtuales

वर्चुअल असिस्टेंट के लिए मल्टी लॅंग्वेज सपोर्ट

가상 어시스턴트를위한 다국어 지원

Suporte em vários idiomas para assistentes virtuais

Sep 2020

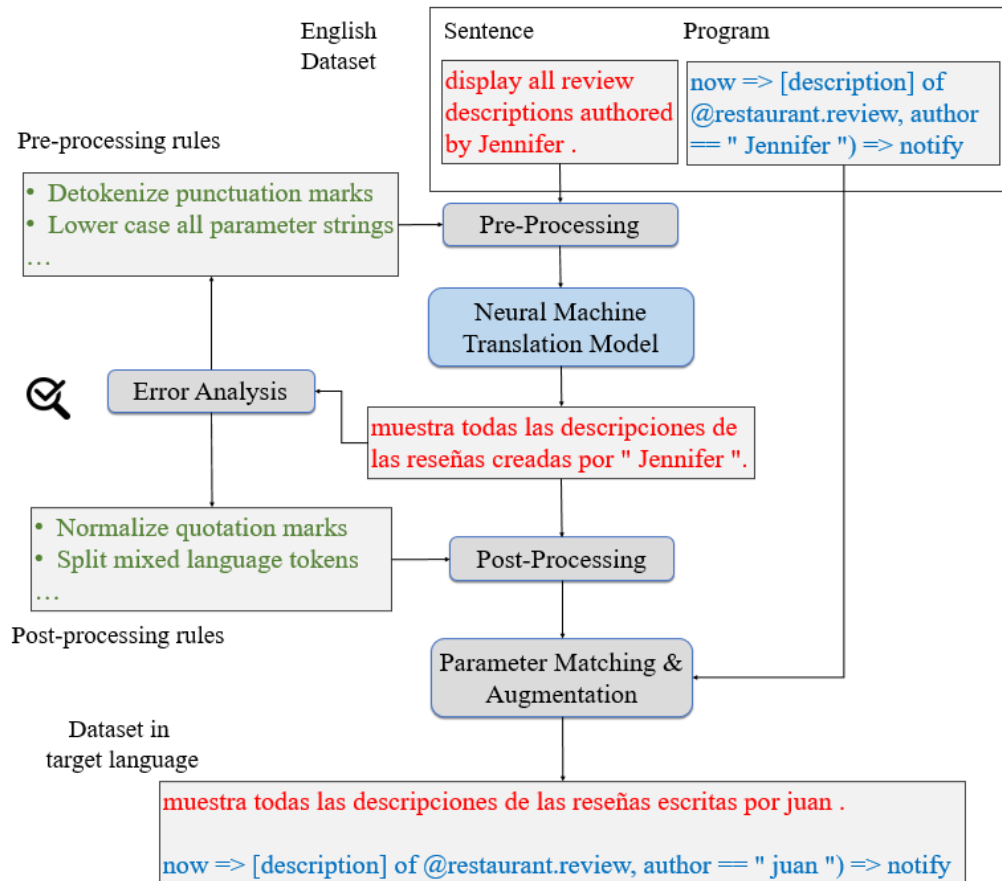
Goals:

- Extending the current capabilities of Virtual Assistants to other languages in a cost and time efficient manner
- Eliminating manual annotation and template-development for every new language



Semantic Parser Localizer(SPL):

- Translate English sentences to target languages
- Use neural network's encoder-decoder cross-attention to extract entity alignment
- Replace target parameters with local named entities
- Improve pre/post processing informed by error analysis



Advantages:

- Requires no manually annotated data
- Allows for augmentation of entity values in target language which is necessary to handle open domain data
- Achieves 15-25% improvement for several languages over previous SOTA
- Human translations are only needed for test data

Limitations:

- Difficult to control quality and variety in generated sentences

Project goal: Improve parser accuracy to match (and potentially surpass) English accuracy

Project Requirements: AI/ ML Knowledge, PyTorch

Project plan:

1. Familiarize yourself with our genie repos (toolkit, k8s) and SPL codebase
2. Choose a domain (e.g. Restaurants, Movies, etc.) and language you know well
3. Follow SPL steps and get baseline results
4. Choose one possible approach from “our list of ideas” (or propose yours!)
5. Implement and test it
6. Iterate through steps 4 and 5 until outperforming baseline accuracy
7. Celebrate!

Project goal: Beat MultiWOZ Chinese/ CrossWOZ benchmark

Project Requirements: Chinese speaker, AI/ ML Knowledge, PyTorch

Project plan:

1. Familiarize yourself with our genie repos and ConvLab2 codebase
2. Use SPL and genie to generate and auto-annotate a new Chinese dataset
3. Work on closing the gap between our results and SOTA
4. Improve the contextual neural network informed by error analysis
5. If your gap becomes “negative ” in step 3 you deserve a paper!