Classification of Written Customer Requests: Dealing with Noisy Text and Labels

Viljami Laurmaa, Mostafa Ajallooeian
Swisscom AG, Bern, Switzerland
viljami.laurmaa@swisscom.com, mostafa.ajallooeian@swisscom.com

1 Setting / Background

Swisscom, the biggest Telco in Switzerland, receives more than half a million written requests per year. These requests should be processed and addressed individually with utmost care to give the best and fastest service to the customer. We have recently developed and deployed a machine learning based system which receives each written customer request plus some additional metadata (e.g. customer segment, communication channel, etc.) and classifies the intent of the request. This helps routing the request to the correct team of agents with relevant skills to address it.

2 Data

We had access to about a million requests including metadata which had been labeled by support agents while treating the requests. Text comes in 4 languages (DE, FR, EN, IT) and some of the German words are Swiss-German dialect. Various artifacts induce noise in the content, such as e-mail headers, footers, signatures. Most requests come from various contact forms which introduce decorative formatting. Requests also come from various sources and not all are customer requests to be treated such as answers to marketing campaigns. Labels are noisy too. Indeed, the classes overlap, some classes are uninformative ("Unknown"), there are more than 100 classes so agents tend to use a restricted amount of favorite classes and also the agents get an initial label suggestion which is sometimes left unchanged.

3 Classification

The goal is to classify the intent of request into one of more than a hundred classes. For this, we use word embeddings pre-trained with word2vec (Mikolov, 2013) and a Convolutional Neural Network (Kim, 2014) per language. We regularize using layer normalization, dropout and L2 weight norm penalization. Some select metadata is prepended to the e-mail text to improve accuracy.

4 Improving Data Quality

As discussed, the data can be heavily noisy, both input (incoming request) and the output (intent label). We have developed an internal crowd-sourcing tool for text labeling. We typically create to kind of jobs: 1) Given a written request, we ask five agents to label the same request. This helps improving the label quality and creating gold test sets. Manual inspection shows a 20% reduction in label error in these sets. 2) Given the intent, we ask the agents to come up with hypothetical requests landing in that category. This helps gather less noisy input data and enable new use-cases.

5 Relative improvements

Previously there was a legacy system which classified each request mostly based on metadata using a set of regex rules. Compared to that, our new system has reduced the intent recognition error by 50.7% over the last year.

References

Yoon Kim, 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, *EMNLP 2014*, pages 1746-1751, Doha, Qatar.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado and Jeff Dean, 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26, NIPS 2013*, pages 3111-3119, Lake Tahoe, USA.