Natural Email Understanding

Automated FAQ Selection in IT Tickets Handling

Costanza Calzolari, Margherita Rosnati, Brian Regan {calzolac, mrosnati, bregan}@ethz.ch

ETH zürich

Executive Summary

Challenge:
This project aims to automate the handling of incoming tickets to the help desk leveraging an internal FAQ list

Solution:
Our solution, given an incoming ticket, either selects the most likely FAQ answers, or flags the lack of appropriate registered answer

Model:
The model consists of an unsupervised method and a supervised classifier which generate and learn a "question-to-FAQ" mapping

Results:
We obtain a 82.2% top3 accuracy and 50.4% top3 F1 score, thus unlocking the potential to improve the IT staff's efficiency in addressing tickets and identifying opportunities for new FAQs

Problem Statement and Related Work

- The project tackles two IT Services problems: similar tickets are answered repeatedly and it is tedious to keep the FAQs relevant
- The aim is to learn a mapping from tickets to FAQs, based on a set of unlabelled ticket conversations and an internal set of FAQs
- Research has tackled community based Q&A information retrieval [1]. However, most information retrieval systems rely on a large tagged dataset and/or answers rankings

Key Challenges

Model structure:
How to map tickets to FAQs in an unsupervised manner.
Solution: Embed tickets and FAQ answers and use a distance metric
Use manually labelled data for evaluation

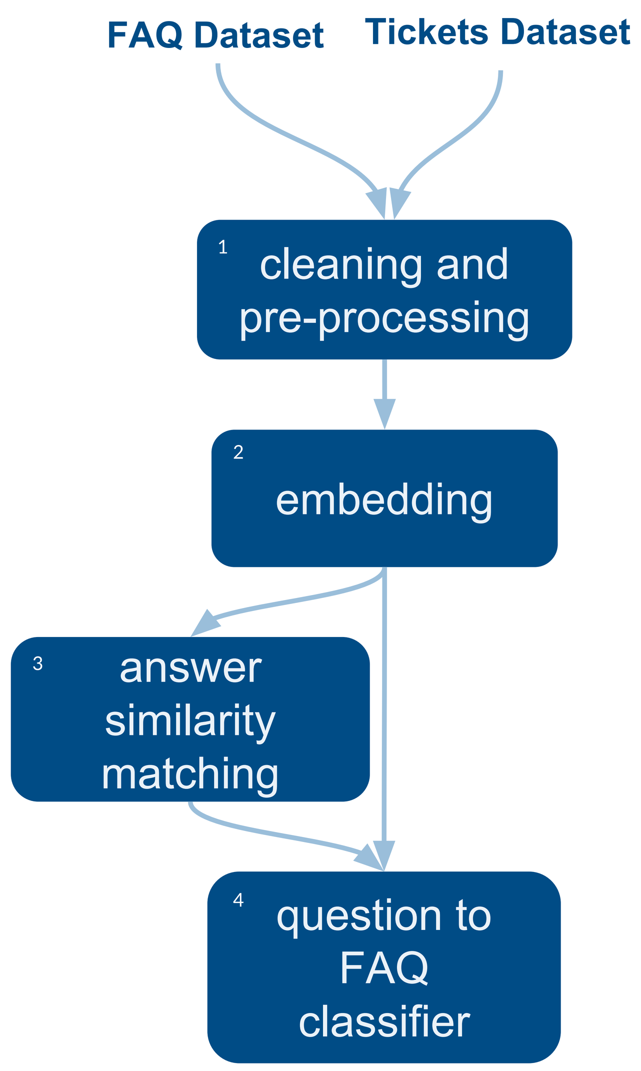
Ticket structure:
How to handle conversational nature of the tickets
Solution: Within a ticket, concatenate separately all questions and all answers

Ticket not answered by FAQ:
How to deal with large "non-FAQ" part of dataset?
Solution: Single unbalanced classifier performed better without the addition of a "non-FAQ" pre-classifier

One-shot classes:
How to train a classifier with FAQs corresponding to a single ticket?
Solution: Tested a classifier with a soft matched "question-to-FAQ" dataset, resulting in similar scores than current method

Pipeline

1. Cleaning and pre-processing
   Input: Raw FAQ and ticket datasets
   Process: translate, reduce and tokenize text
2. Embedding
   Input: Cleaned FAQs and tickets
   Process: Embed ticket questions, ticket answers and FAQ answers
3. Similarity clustering
   Input: embedded FAQ answer, embedded ticket answer
   Process: match tickets to FAQs using cosine similarity. Determine tickets to assign to "non-FAQ" class
4. Classifier
   Input: embedded ticket question, "ticket-to-FAQ" dataset
   Process: Classify ticket question based on "ticket-to-FAQ" dataset



1. Cleaning and Pre-processing

FAQs
Removals:
- Partially empty FAQs
- Duplicates
- "Junk" FAQs

→ **Final # FAQs: 199**

Tickets
Removals:
- Automated messages
- Emailing artefacts
- "Junk" tickets

Conversation handling:
- Concatenation

Model validation
- Removed 200 tickets for test/val.

→ **Final # Tickets: 4007**

Pre-processing
- Translated with Google API [2]
- Punctuation removal
- IP/email removal
- Stop-word removal
- Numeric removal
- Stemming

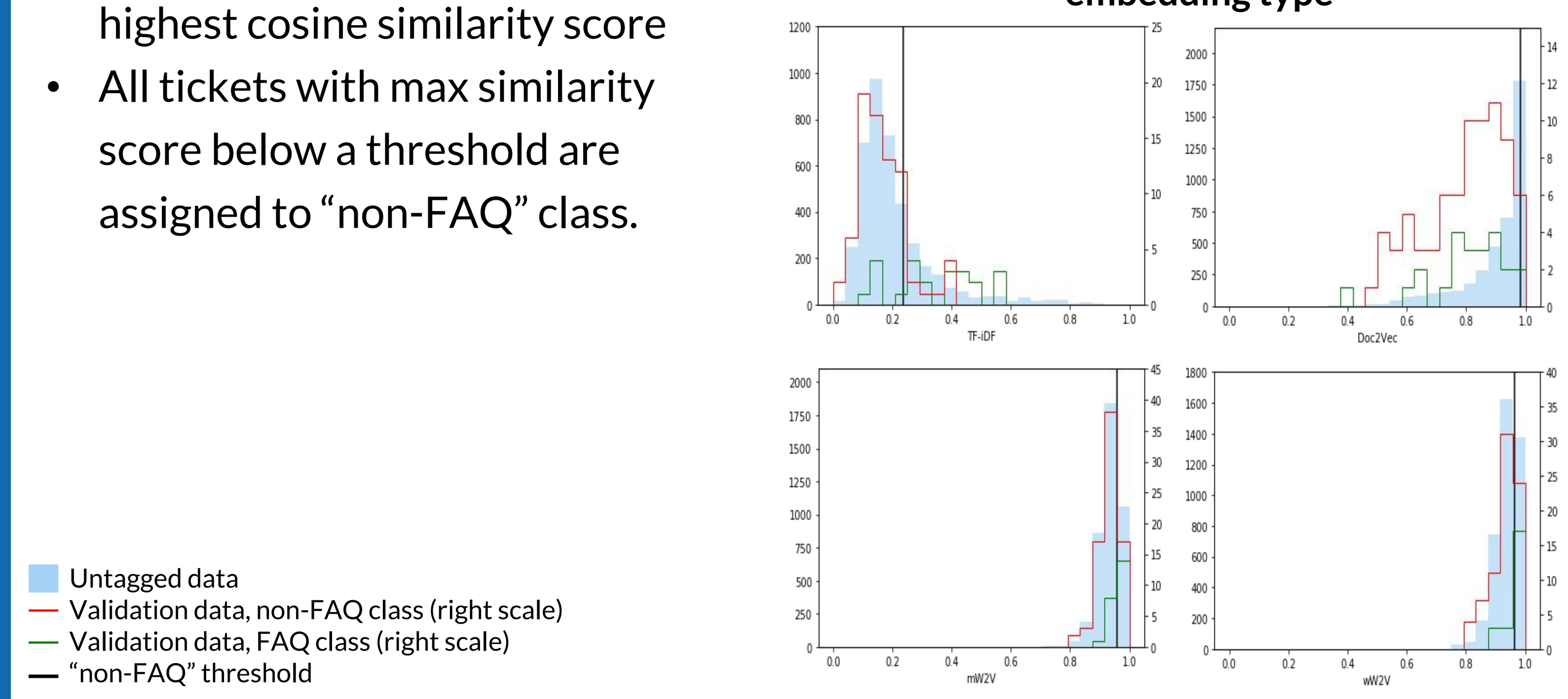
2. Embeddings

- **TF-iDF:** frequency based reflection of the importance of each word to a document [5]
- **Mean Word2Vec (mW2V):** average of the word2vec [3] representation over a document
- **Doc2Vec:** Distributed Memory paragraph vectors [4]
- **TF-iDF weighted word2vec (wW2V):** TF-iDF weighted average of the word2vec representations

3. Similarity Matching

- Allocate ticket to FAQ with highest cosine similarity score
- All tickets with max similarity score below a threshold are assigned to "non-FAQ" class.

Histograms of highest similarity score per embedding type



4. Classifier

We used a random forest classifier with 10 trees and unbalanced classes to map ticket questions to FAQ indexes.

Results

Results when looking at top 3 FAQ predictions per ticket over test set

Model	Avg. Precision	Avg. Recall	F1-Score
Predict all as Non-FAQ	72.0%	36.0%	48.0%
Doc2Vec	71.1%	36.0%	47.8%
Mean Word2Vec	83.4%	36.4%	50.7%
TF-iDF	**89.4%**	**38.1%**	**53.4%**
TF-iDF weighted Word2Vec	84.0%	36.9%	51.3%

Conclusion and Further Work

Findings:
- Similarity matching on answers performs better than over questions
- Simple 'word frequency'-based TF-iDF models outperform semantic representation models such as word2vec and doc2vec
- The size of the non-FAQ class complicates the task

Next Steps:
- Automatically identify recurring tickets to create new FAQs

References

[1] J. Jeon, W. B. Croft, and J. H. Lee. Finding similar questions in large question and answer archives. In Proceedings of the ACM Fourteenth Conference on Information and Knowledge Management, pages 76–83, 2005.
[2] https://console.cloud.google.com/apis/api/translate.googleapis.com/overview
[3] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.
[4] Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." International Conference on Machine Learning. 2014.
[5] G. Salton, C. Buckley. "Information Processing & Management." Information Processing & Management. 1988
[6] L. Maaten, G. Hinton. "Visualizing data using t-SNE." Journal of machine learning research. 2008

Acknowledgments We would like to extend a thank you to Professors Zhang, Krause and Feuerriegel and Bernhard Kratzwald for the academic guidance, Mark Buschor for the time given to us and food&lab for sustaining all of our meetings