Assigning topics to customer support messages

Birgit Sõrmus, Hendrik Šuvalov, Marti Ingmar Liibert
Team

- Ingmar - 1st year Computer Science MSc
- Hendrik - 1st year Data Science MSc
- Birgit - 3rd year Computer Science BSc

- Roland Kriibi - Machine Learning Team Lead at Pipedrive
Problems

- Assigning topics to customer support messages
- Giving overview of customers’ problems
- Directing customers’ problems to right people
Preprocessing

- Data preprocessing
  - Language detection and removing non-English conversations
  - Removal of escaped personal data in the following format: {{name}} etc
  - Lemmatizing
  - Removing stopwords
  - TF-IDF vectorizing
Model

- LDA - Latent Dirichlet Allocation
- Optimising number of topics and learning decay using perplexity of the model
Optimising number of topics and learning decay
Resulting topics

- **Email sync**
  - “With our workflow automations you can indeed send automatic emails based on triggers, for example if a deal move from stage a to b send email x”

- **User activities**
  - “Mark an existent scheduled activity as done OR create a new one when a call has finished”

- **Account billing**
  - “Reaching out, I checked your account and I confirm the subscription has been cancelled successfully, it will expire on Jan 4, the last day of your billing cycle”

- **Mailigen**
  - “Errors appear on his side, in his system, because Mailigen e-mails do not know why t received differently from others and are not delivered to the inbox (e-mail is accepted, no bounce is formed). Thanks!”
Validating

Trained new models with month-specific data to see if a new topic, Coronavirus, emerges in March.

Proves that the model can:

- Assign correct topics to conversations
- Discover new topics
Covid frequency by topics throughout the year
Covid frequency in topics by months
Lessons

- Language detection is not perfect
- Data exploration beforehand is important
- Optimising training speed
Thank you!

https://github.com/ingmarliibert/intercom-topic-assignment