Assign topics to customer support conversations

Ott Kaarel Martens, Rasul Nabiyev, Joonas Praks
Project owner

- developer of the homonymous sales customer relationship management (CRM) tool
- ~90,000 customers around the world
Problem
Problem

- ~13k monthly conversations with the Customer Support
- Labeling conversations according to their content into topics would give better insight into how the work of Customer Support is distributed
Train a model to extract topics from the English conversations

Be able to predict new possible topics
Approach

- Doc2vec + clustering - Not good
Approach

- Doc2vec + clustering - Not good
- LDA - Old but gold, also good for trend analysis later on ✔
Approach

- Doc2vec + clustering - Not good
- LDA - Old but gold, also good for trend analysis later on
- Lda2vec - Failed, documentation not adequate
Approach

- Doc2vec + clustering - Not good
- LDA - Old but gold, also good for trend analysis later on ✔
- Lda2vec - Failed, documentation not adequate
- BERT embeddings + clustering ✔
Approach

- Doc2vec + clustering - Not good
- LDA - Old but gold, also good for trend analysis later on
- Lda2vec - Failed, documentation not adequate
- BERT embeddings + clustering
- Top2vec clustering - BERT was better, same in principle
Preprocessing

- Filter out bot messages (some labeled, others disguised)
- Omit languages other than English
- Removed the names of customer support agents
- Model specific preprocessing
Topics with BERT
Topic #1

- ('corona', 0.5908564881292043)
- ('virus', 0.530993725778748)
- ('planning', 0.48535634302408937)
- ('cancel', 0.2716410697082499)
- ('trish', 0.035470184121990364)
Topic #2

- ('amp', 0.08232454867679548)
- ('mg', 0.07504288054800366)
- ('customerid', 0.06587016788257814)
- ('latvian', 0.061217819703789006)
- ('campaign', 0.057916018801746294)
Topic #3

- ('australia', 0.05520498576540953)
- ('australian', 0.019919387533001893)
- ('gst', 0.01710706462716414)
- ('abn', 0.012073675704144174)
- ('nsw', 0.010174873507892207)
Workflow Automation
(product feature)
Account Lockdown
(new feature)
General customer browser issues
Showcasing a stable trend
Billing

0.053**"billing" 0.035**"seat" 0.035**"invoice" 0.025**"charge" 0.025**"payment"
Cancellation
Possible applications
Possible applications

- Real time
  - Feasible, but probably not fast/accurate enough
- Retroactive
  - On what subjects is customer support spending its time on?
  - Which product features should be improved for a better customer experience and less overhead on customer support?
What did we learn?

- Novelty is not always superior: a 20 years old model could still give better results. Promises made by new technology can be deceptive.
- Unsupervised learning has its limits
- Domain knowledge is powerful
- NLP is a broad and cool topic to explore!
Thank you for listening!
Link to our Colab

https://colab.research.google.com/drive/1ZraPpase_YTFwwbSuow1qyUBBmJlB0g?usp=sharing