



Predicting Message Paths and Conversation Intent

HELLO!

Comm 100

Presentors:

Kristen Bystrom

Ye Liu

Alice Roberts

Ling Zhang

Qiang Zou

Wayne Ku

Nafis Abrar

Myung-chul Jeon

Tatchai Titichetrakun

Shyong Quin Yap

Rihan Yao

Executive Summary

1. Introduction
2. Our Process
3. Our Results
4. Future Intentions
5. Summary

Introduction

What is our goal?

- Given a data set from Comm 100 we predicted the intent of conversations as well as clustered each message.
- We also analyzed probability of conversation sentiment based on the clustering of the chat transcripts.

What does our data look like?

We were provided with several data sets to work with:

- ❑ Sample data includes 7,000 conversation transcripts. (1 transcript from each chat session.)
- ❑ Full data includes 59,082 conversation transcripts.
- ❑ The scripts consists of several languages including English and Mandarin.

Sample Chat Transcript

[Agent]Brian - Comm100: Hello oguz, this is Brian. How can I help you?

[Visitor]oguz: hi brian

[Agent]Brian - Comm100: Hi

[Visitor]oguz: we have a problem about chat session

[Visitor]oguz: 2 or 3 customers connected same session

[Agent]Brian - Comm100: What seems to be the issue?

[Visitor]oguz: and customers chat together

[Agent]Brian - Comm100: Hi your colleague has reported this issue yesterday

[Visitor]oguz: yes Our members complain very much in this situation

[Agent]Brian - Comm100: It was because those users need to be separated.

[Agent]Brian - Comm100: I have reported to the developers

[Visitor]oguz: yes, your personel give this link

<https://betlive100.com/chatserver/chatwindow.aspx?planId=146&siteId=223641>

[Agent]Brian - Comm100: And will let you know when fixed

[Agent]Brian - Comm100: Yes

[Visitor]oguz: thank you

[Agent]Brian - Comm100: you are welcome

[Visitor]: The visitor has left the chat.

2

Our Process

What did we do?

Prepared Data

1. Pulled Data
2. Cleaned Data
3. Structured Data

Vectorized Features

1. Found statistics on the words
2. Extracted keywords by TFIDF
3. Attained sentiment by TEXTBLOB

Clustering

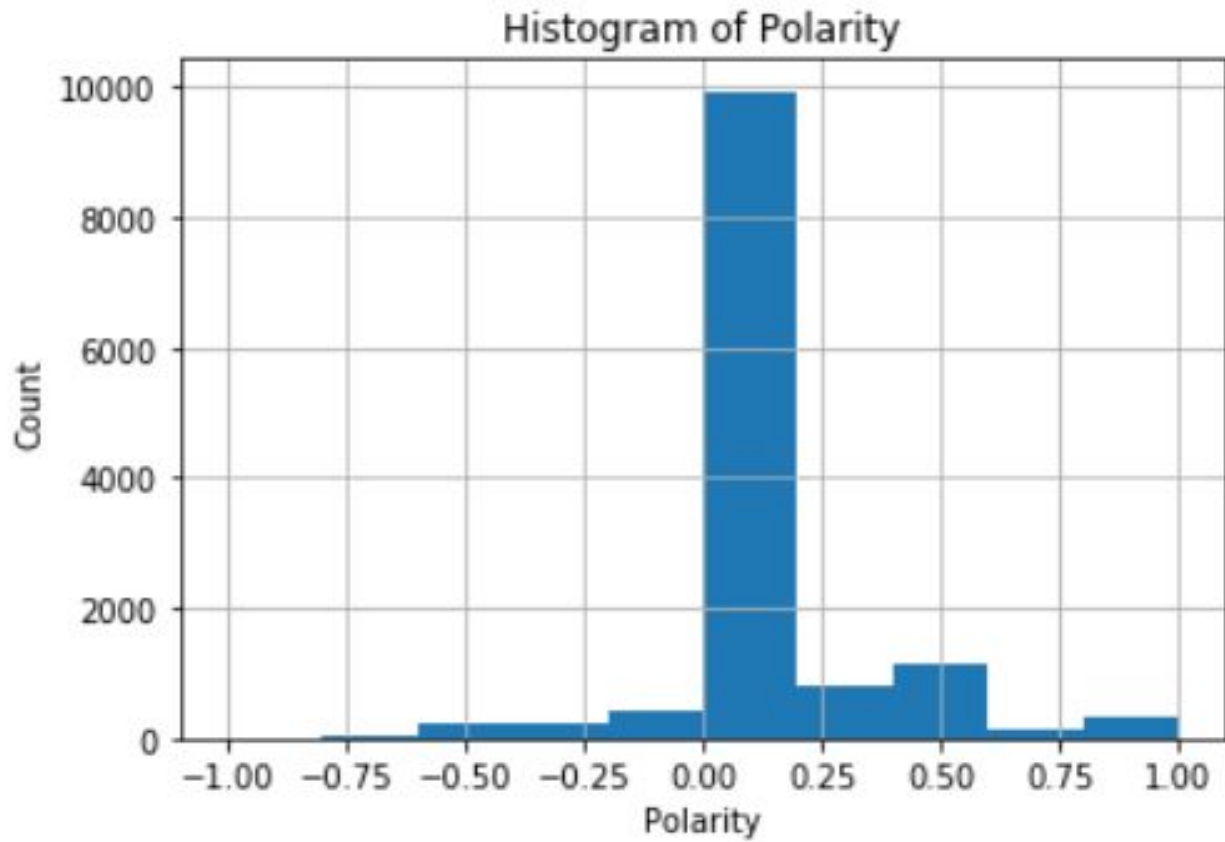
1. Refined keywords by LDA & K-Means
2. Performed NLP using spaCy

From Chat Logs to Pandas

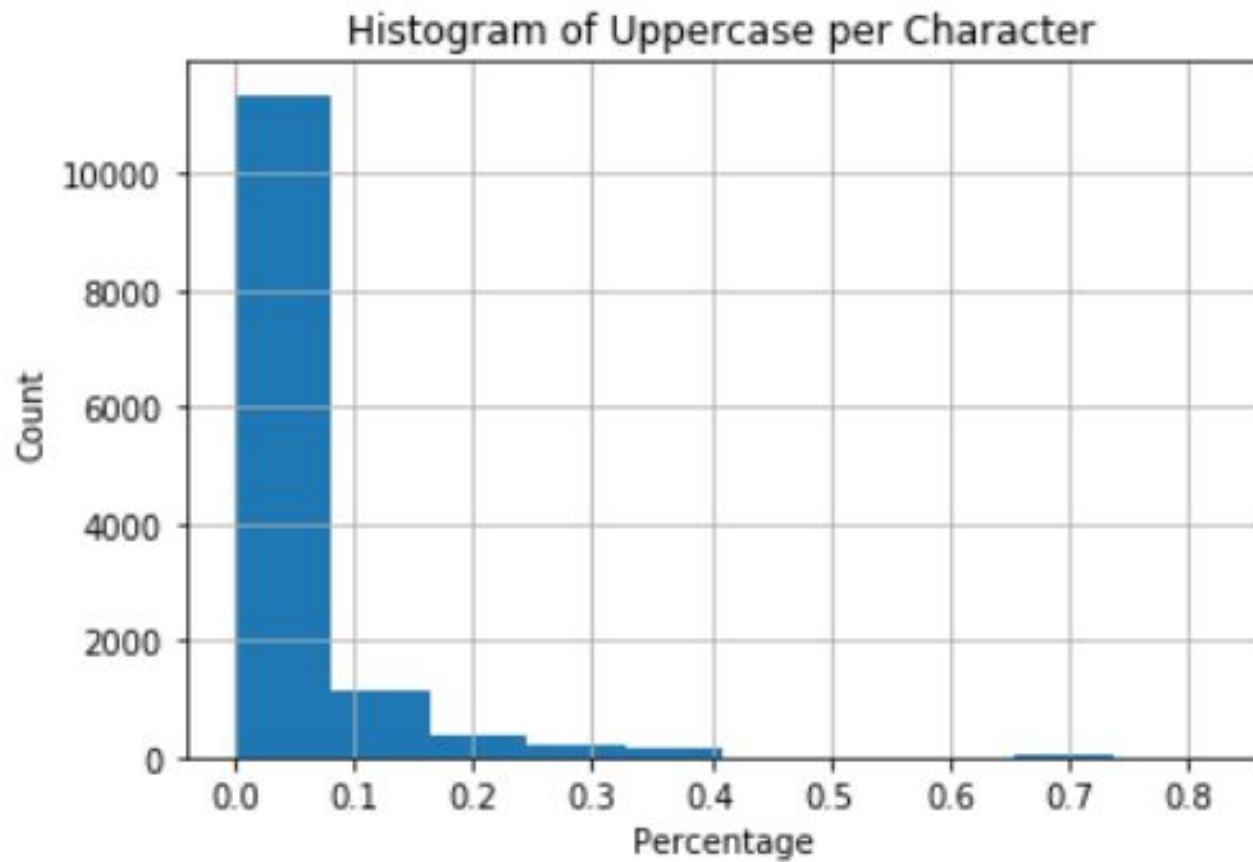
	ConversationID	MessageID	Message	Speaker	numWords	numChar	WordsPerChar	UpperPerChar	Question	Polarity
0	sample_data/801.txt	1	hi johnny	Visitor	3	10	0.300000	0.000000	0.0	0.0
1	sample_data/473.txt	1	norman my apologies i am leaving for a week a...	Visitor	17	77	0.220779	0.038961	0.0	0.0
2	sample_data/305.txt	1	the visitor has left the chat.	Visitor	7	31	0.225806	0.032258	0.0	0.0
3	sample_data/9.txt	1	girls	Visitor	2	6	0.333333	0.166667	0.0	0.0
4	sample_data/25.txt	1	hi	Visitor	2	3	0.666667	0.000000	0.0	0.0

3

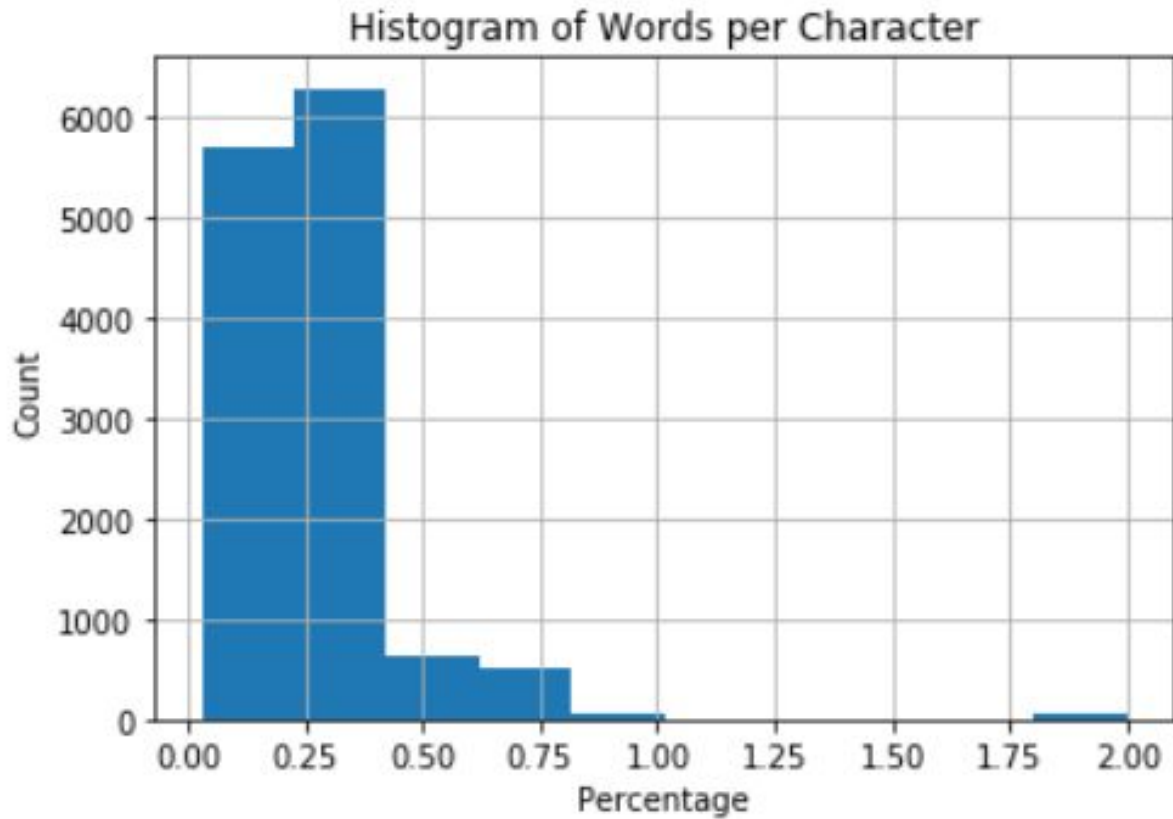
Our Results



Our Results



Our Results

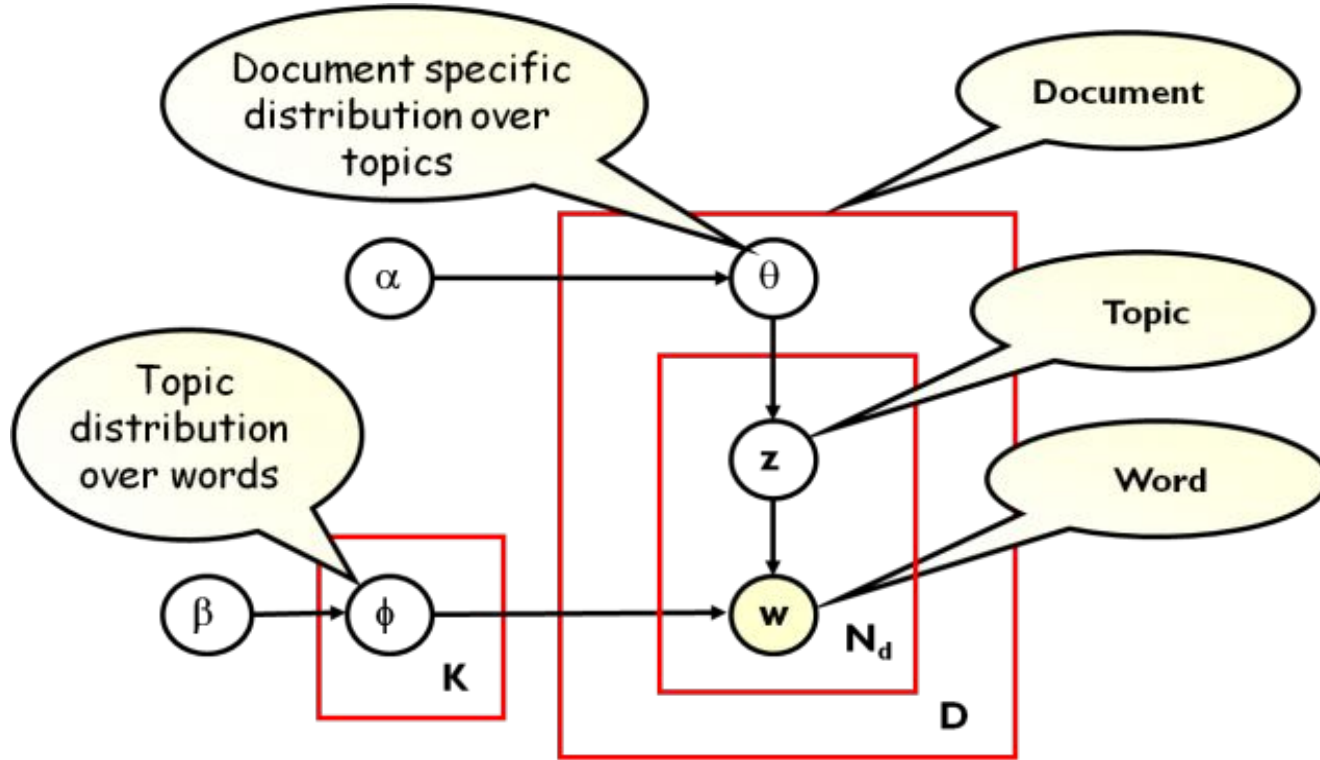


Our Results

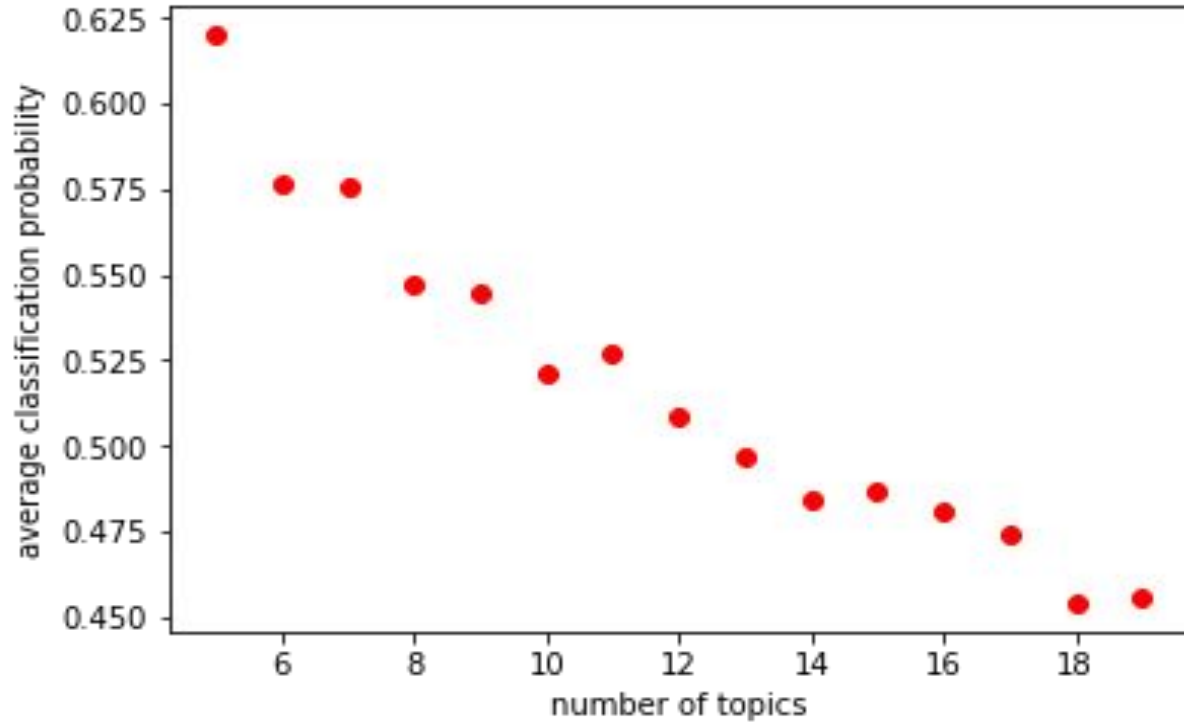
LDA dataframe

data\keywords	account	bill	contact	desktop
Text 1	1	0	0	1
Text 2	0	1	0	0
Text 3	0	0	0	0
Text 4	0	1	1	1
Text 5	1	1	1	1
.....

Latent Dirichlet Allocation (LDA) Model

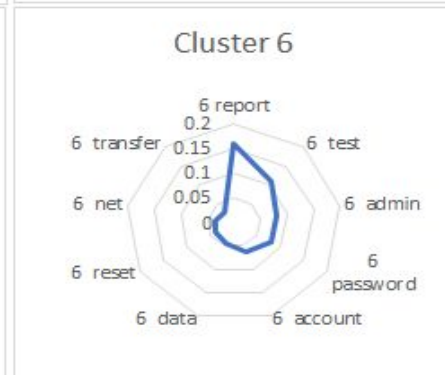
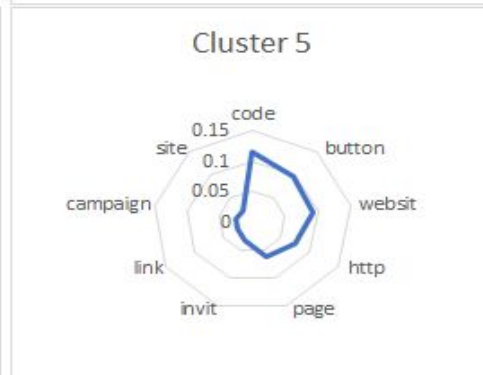
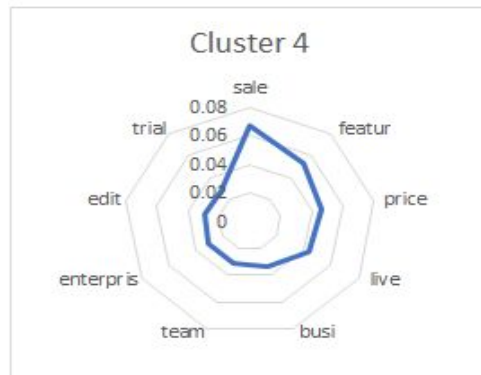
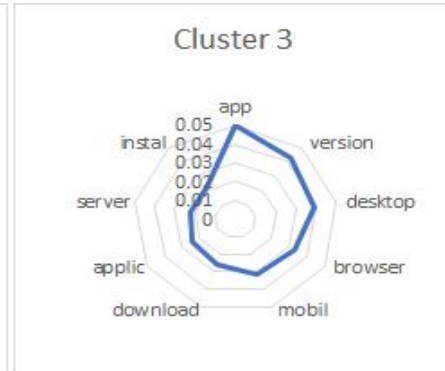
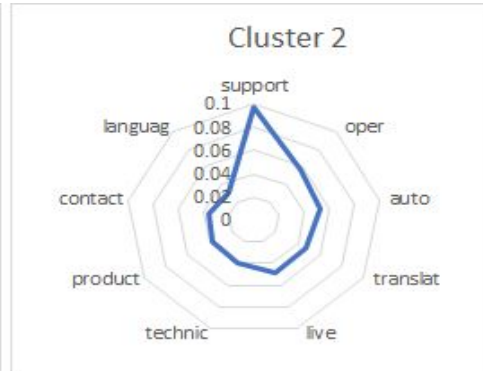
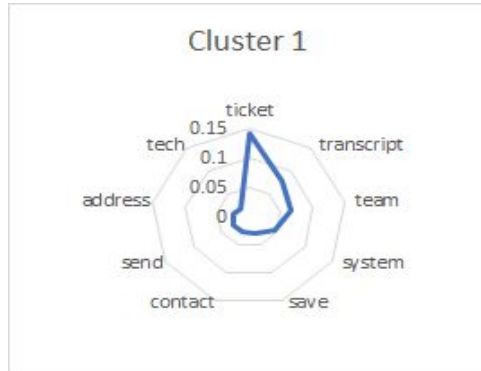


Clustering Accuracy vs Number of Topics



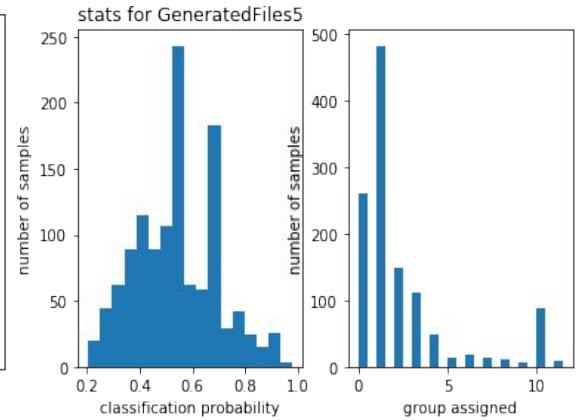
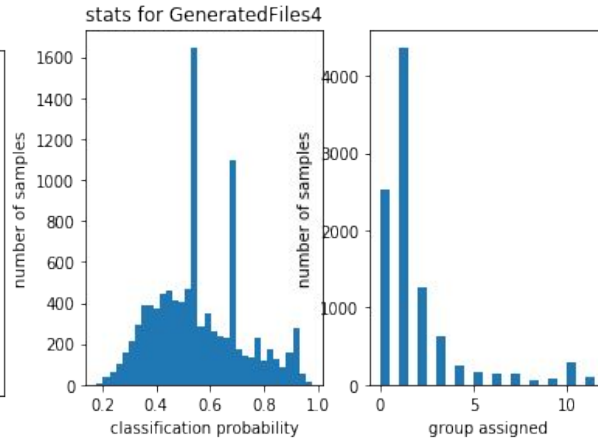
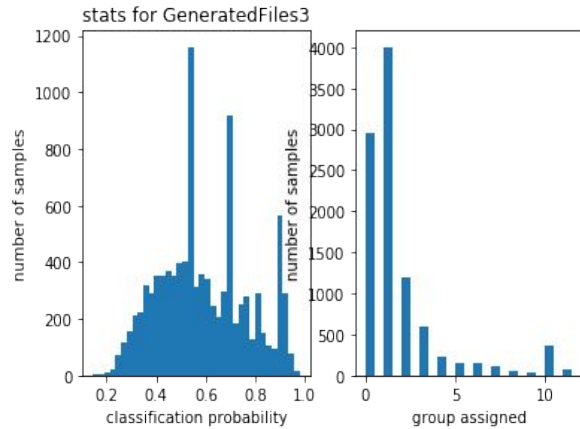
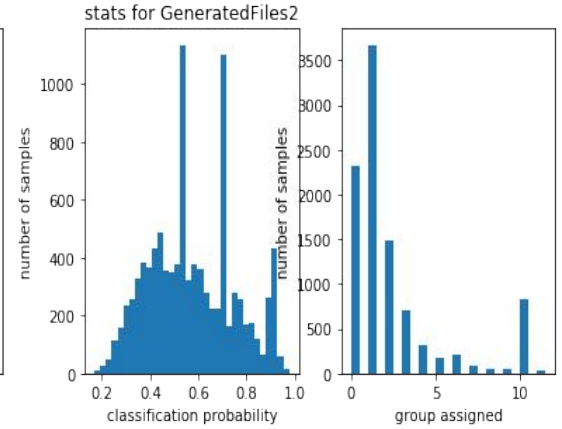
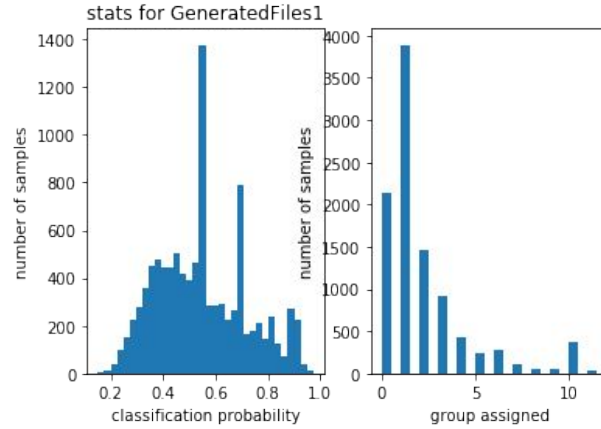
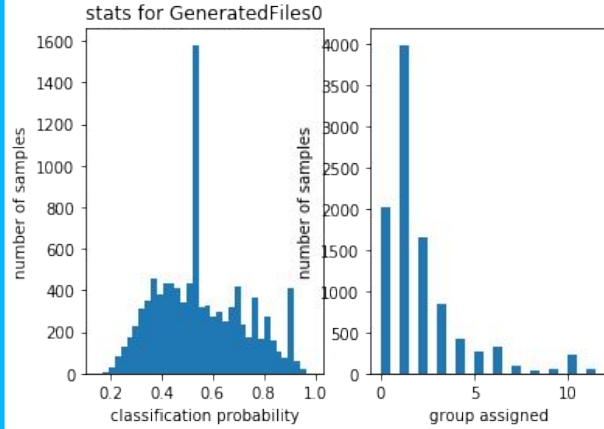
Our Results

Radar Charts of each Cluster



Our Results

Our Results

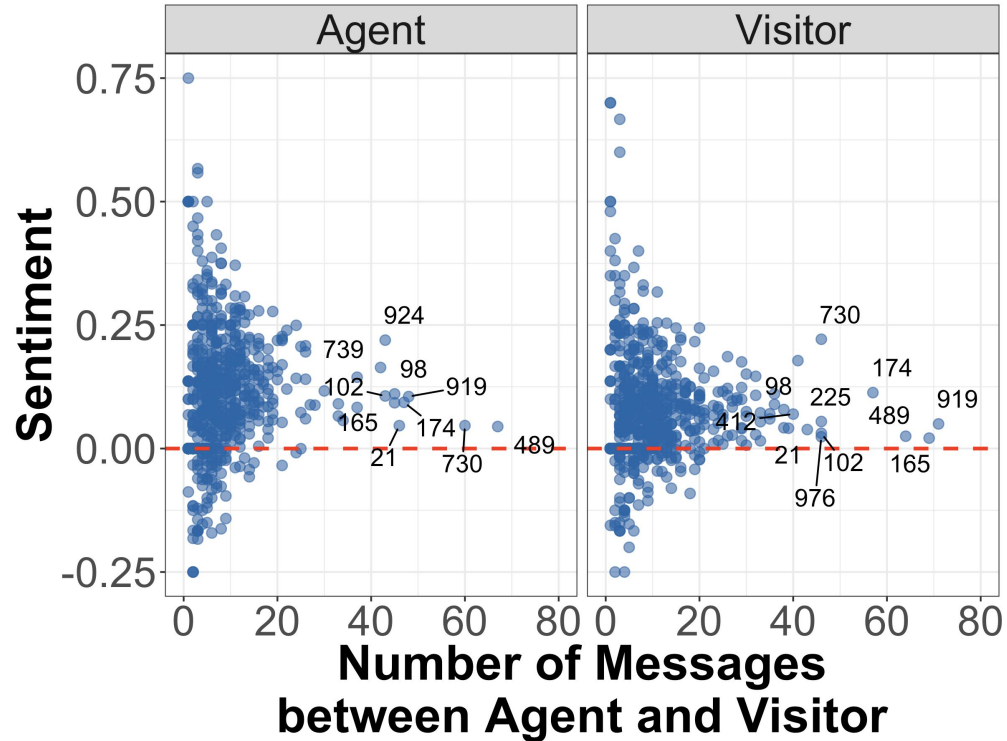


Correlation of clustering distributions on 6 datasets

1	0.99814471	0.98641715	0.97718553	0.98949556	0.9889491
0.99814471	1	0.99219326	0.98426065	0.99273614	0.99514591
0.98641715	0.99219326	1	0.9864876	0.98728002	0.99441838
0.97718553	0.98426065	0.9864876	1	0.99420914	0.98341255
0.98949556	0.99273614	0.98728002	0.99420914	1	0.99124636
0.9889491	0.99514591	0.99441838	0.98341255	0.99124636	1

Data Visualization based on sentiment analysis

So, how positive is our interaction with the customers?



4

Future Intentions

Markov Chain Message Modelling

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{pmatrix} 0 & p_{ij} & p_{ij} & p_i & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Let state 1 be the first message of the conversation.

Let the final state be the last message in the conversation.

Let states 2-(n-1) be the clusters from our analysis.

Let P be the transition matrix of probabilities of transitioning between message states i and j .

Markov Chain Message Modelling

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{pmatrix} 0 & p_{ij} & p_{ij} & p_i & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & p_{ij} & p_{ij} & p_{ij} & p_{ij} & p_{ij} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Let state 1 be the first message of the conversation.

Let the final state be the last message in the conversation.

Let states 2-(n-1) be the clusters from our analysis.

Let P be the transition matrix of probabilities of transitioning between message states.

Markov Chain Message Modelling

$P =$

	1	2	3	4	5	6
1	0	p_{ij}	p_{ij}	p_i	p_{ij}	p_{ij}
2	0	p_{ij}	p_{ij}	p_{ij}	p_{ij}	p_{ij}
3	0	p_{ij}	p_{ij}	p_{ij}	p_{ij}	p_{ij}
4	0	p_{ij}	p_{ij}	p_{ij}	p_{ij}	p_{ij}
5	0	p_{ij}	p_{ij}	p_{ij}	p_{ij}	p_{ij}
6	0	0	0	0	0	1

Let state 1 be the first message of the conversation.

Let the final state be the last message in the conversation.

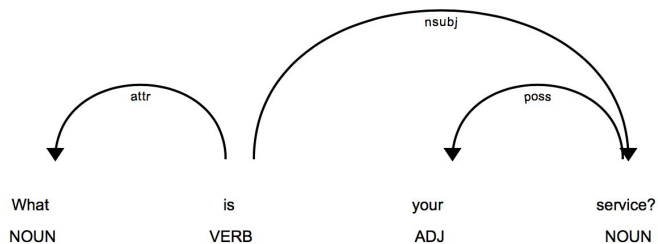
Let states 2-(n-1) be the clusters from our analysis.

Let P be the transition matrix of probabilities of transitioning between message states.

Sentiment Analysis

Text interpretation using SpaCy

- ▣ Identifying sentences.
- ▣ Identifying the interdependencies between the words.



Rasa NLU trainer based on SpaCy and Scikit-Learn

- ▣ Determine intent of a particular sentence.
 - ▣ Associating intent with entities and values

Input: "Can I take a price for comm100 for 8 user?"

Intent: Inquiry (confidence)

Entity: Purchase

Value: Price

Determining the intent of conversations based on sentiment analysis

- ❑ What is the main cause of negative interactions in the transcript?
- ❑ What similarities or keywords do the positive and negative messages have in common?
- ❑ Improve services based on the features that most people are complaining about.

A Better Chatbot?

Context Specific

Give the right answer

Starting with identifying the topic

**Level 1
Categories
By
Topics of
Conversations**

- 1. Setup & Getting Started**
- 2. Settings & Customizations**
- 3. Monitor & Chat**
- 4. Transcripts & Reports**
- 5. Billing & Management**
- 6. Desktop & Mobile Apps**
- 7. API & Webhooks**
- 8. Troubleshooting**

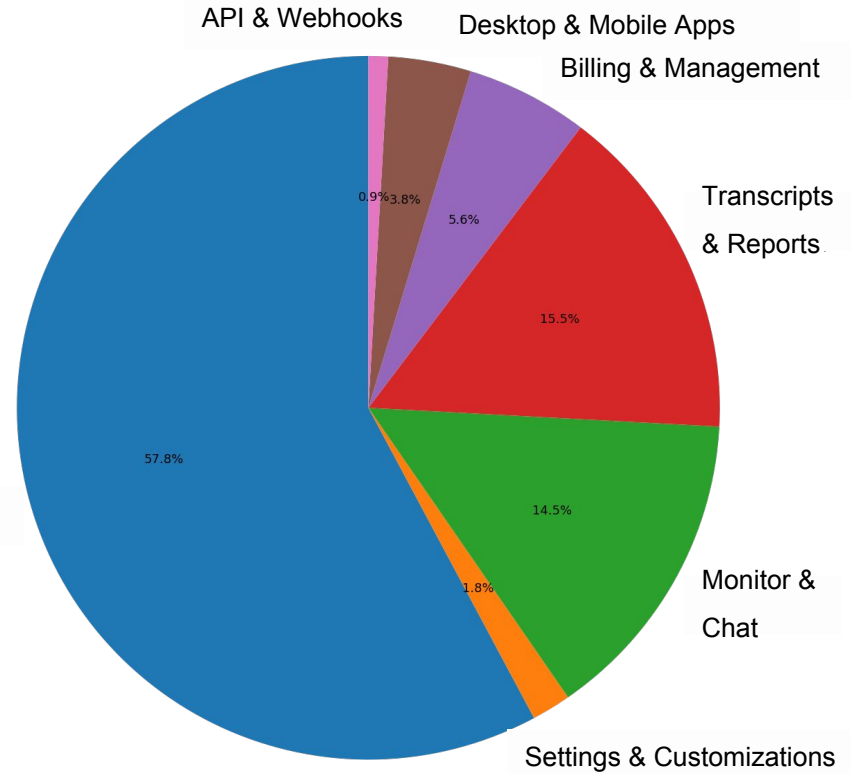
Settings & Customization Keywords



Results

Setup & Getting Started

Count



Next Steps?

Additional Sub-categories

Integration With Knowledge Base

Chatbot Development

5

Summary

