Novel approach for tagging of discourse segments in helpdesk emails

Suman Yelati, Rajeev Sangal

Language Technologies Research Center,

International Institute of Information Technology, Hyderabad, India.

Motivation

- High volume of email requests(queries) sent to the support desks

- Can some amount of automation be done?
  - Finding repetition of same/similar queries
  - Auto composing response
  - Natural language interface to database(s)

Discourse Parsing of Email

To identify how the discourse units in the email are related to each other, where discourse unit could be a sentence or a clause.

Example

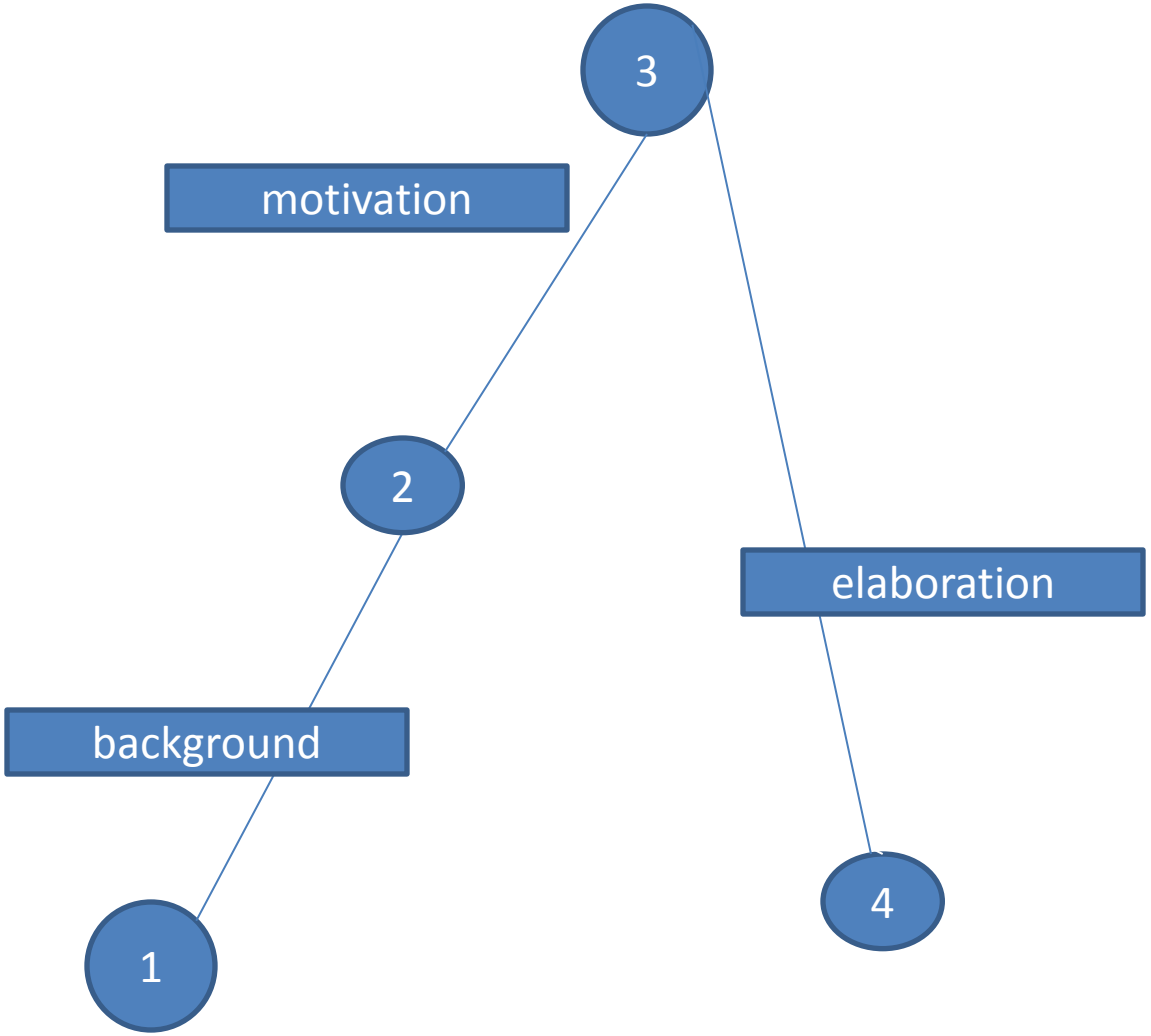
Hi,

I am a ABC Co-ordinator for XYZ Mobile otherwise Known as PQR @Hyderabad. We wanted to make use of ABC assets but the problem is we do not have access to it from our domain – INHYDPQR . Is there any way we can access the site from our domain. if it is technically feasible please let me know what needs to be done from our end and also if there is any fiscal impact please let me know the same .

Thanks,

Auth-Name.

Discourse Parse of the Email



1

- I am a ABC Co-ordinator for XYZ Mobile otherwise Known as PQR @Hyderabad

2

- We wanted to make use of ABC assets but the problem is we do not have access to it from our domain - INHYDPQR

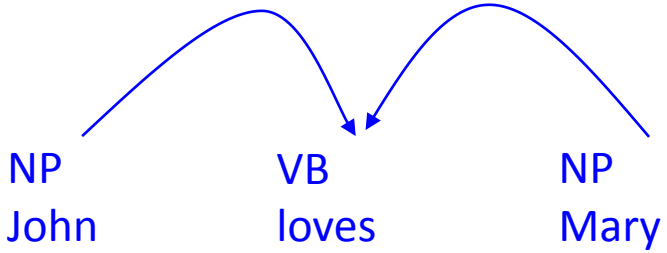
3

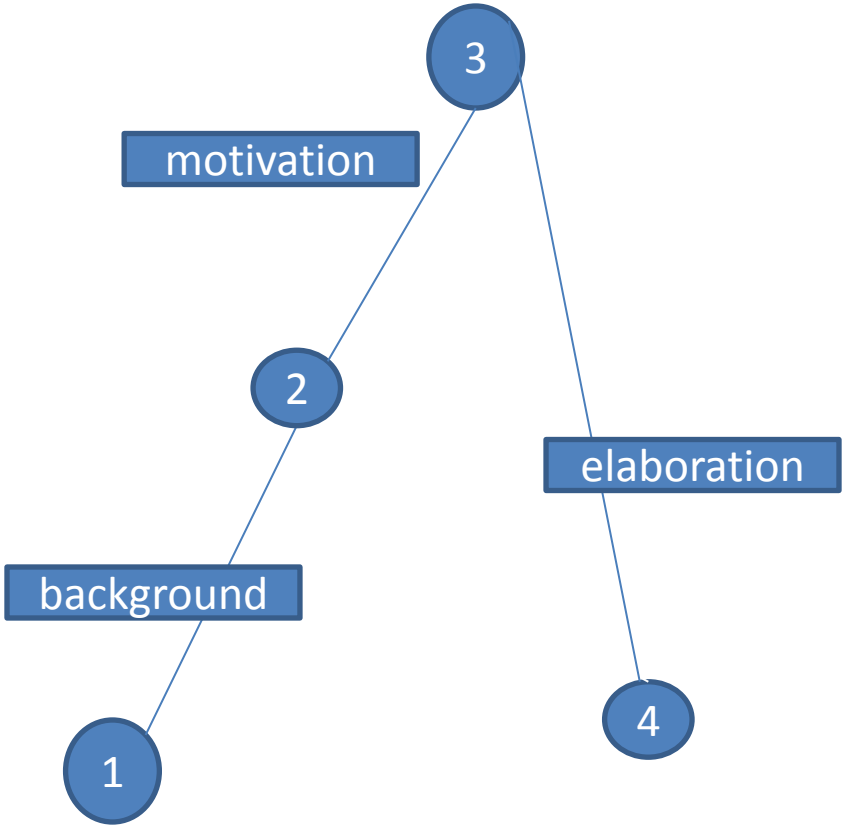
- Is there any way we can access the site from our domain .

4

- if it is technically feasible please let me know what needs to be done from our end and also if there is any fiscal impact please let me know the same .

Hypothesis





POS(Parts of Speech) Tag ←→ Discourse Segment Tag Analogy

2-Stage discourse parsing

- First stage: Label the discourse segments into classes that would help find relations between them.

- Second stage: Generate the parse trees with the help of label info that was generated in stage 1.

Tagging of discourse segments in Email

- Classification of units in email into one of the six classes.
  - Greet
  - Query
  - Context of the Query
    - background
    - goal
    - concern
  - Address

Example for Classification of Discourse Segments

Hi,	=> greet
I am a ABC Co-ordinator for XYZ Mobile otherwise Known as PQR @Hyderabad.	=> background
We wanted to make use of ABC assets but the problem is we do not have access to it from our domain – INHYDPQR .	=> concern
Is there any way we can access the site from our domain.	=> query
if it is technically feasible please let me know what needs to be done from our end and also if there is any fiscal impact please let me know the same .	=> query
Thanks,	=> greet
Auth-Name.	=> address

Definitions

Class Name and definition	Example(s)
Greet: Greetings and Pleasantries	Hi Sir, Respected Sir/Ma'am, best regards, eagerly awaiting your response
Background: Context setting by the user for the query. Mainly informative	I have written the PGEE entrance exam on feb-19th 2009. My name is ABCD and I work at the Hyderabad center.
Goal: What is the user trying to achieve.	I want to know when I will get my result.
Concern: Any issue faced by the author	I am not able to login to the server with the given details.
Query: The request being made	Can you please tell me when I can get the result of the exam.
Address: Description/Signature of author	Abhishek. Application Number: 328839 Mobile: 9490055922

Data creation

- Post Graduate admission help-desk emails
- Technical help desk emails

- Number of discourse units in the emails : 1705

- Issues in annotation:
  - When there are two types of speech acts in the same sentence.
  - E.g: I am trying to connect to internet but the password does not work. (Goal / Concern?)
  - Query > Concern > Goal > Background > Address > Greet

Distribution of Data

Greet	Background	Goal	Concern	Query	Address
23.87%	23.81%	2.64%	11.90%	17.24%	23.17%

Observations

- Baseline: Word n-grams(up to trigrams) – **33%**
- Stemming :
  - Reduces accuracies
- Generalizing sentences : -- **60.3%**
  - He went home.
  - <person> went home.
  - NN went NN
- Some rules to predict the class and add that class as a feature for learning -- **68.3%**

Observations (cont)

- POS n-grams

Can	you
Will	she
Shall	he
Could	

- Handpicked features
  - Greet: Hi/Dear/Hello/regards/thanks
  - Query: please/?/Start with a question word
  - Address: Capitalization
  - Concern: but, not
- **74%**
- SVM (Support Vector Machines) used for learning

Confusion matrix

	Greet	Background	Address	Concern	Query
Greet	92.52	2.24	4.49	0	1
Background	0.74	84.9	8.91	2.23	0.99
Address	0.76	0.51	98.73	0	0
Concern	0.49	49.02	0.49	48.53	1.96
Query	0.34	14.24	2.03	0.68	82.71

Error analysis

Dear sir,

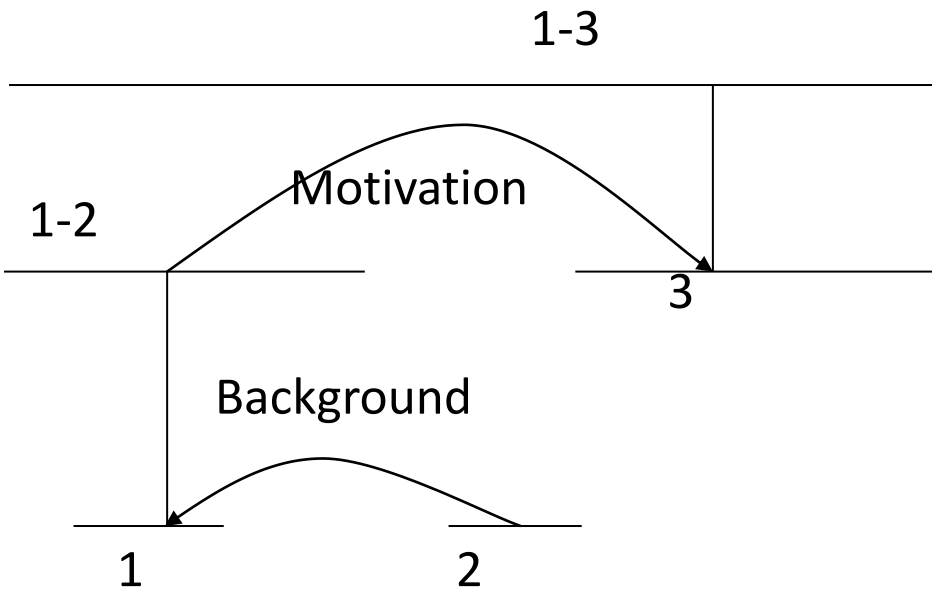
My name is ABC, and I am in XYZ department. **Internet has not been working on my system for last 3 days**. I have raised a support-ticket couple of days back but no action has been taken yet. This delay in handling support tickets is now very frequent. Request you to make a serious note of this and enquire into the situation to make sure that support-tickets are solved promptly.

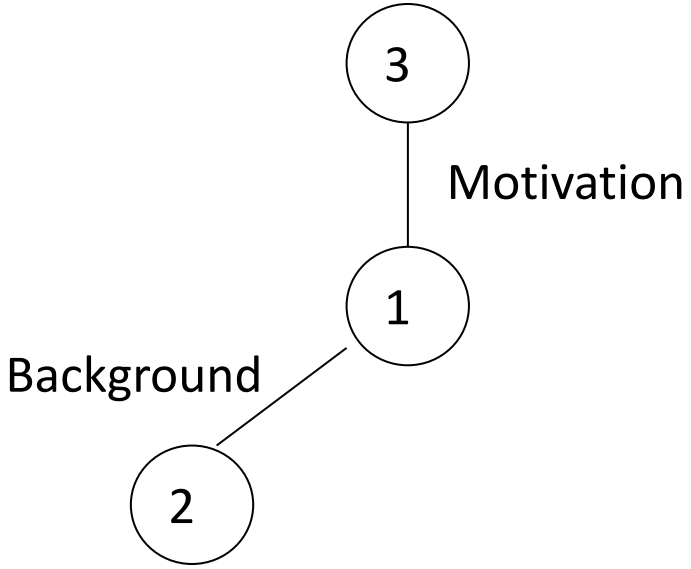
Thanks,

ABC.

Discourse parse of email

- RST corpus of the emails





Parsing emails

- 850 relations
- Applying MALT parser
- CONLL format to represent

Results

Best case scenario.

GOLD Tags for Training parser and GOLD Tags for Testing parser

==========

Labeled attachment score = **84.44 %**

Unlabeled attachment score = **90.07 %**

More results

GOLD Tags for Training parser and AUTO Tags for Testing parser

==========

Labeled attachment score: 174 / 302 * 100 = 57.62 %

Unlabeled attachment score: 222 / 302 * 100 = 73.51 %

AUTO Tags for Training parser and GOLD Tags for Testing parser

==========

Labeled attachment score: 159 / 302 * 100 = 52.65 %

Unlabeled attachment score: 189 / 302 * 100 = 62.58 %

Auto Tags for Training parser and Auto Tags for Testing parser

========== =========

Labeled attachment score: 142 / 302 * 100 = 47.02 %

Unlabeled attachment score: 180 / 302 * 100 = 59.60 %

Observation

Labeling of discourse segments is a useful sub stage for finding the discourse parse.

Future work

- Finer discourse segments
- Finer tags
- Richer set of relationships

Merci

thank you

Questions/Comments