Joint Modeling of Content and Discourse Relations in Dialogues

Kechen Qin[1], Lu Wang[1], and Joseph Kim[2]

[1]College of Computer and Information Science

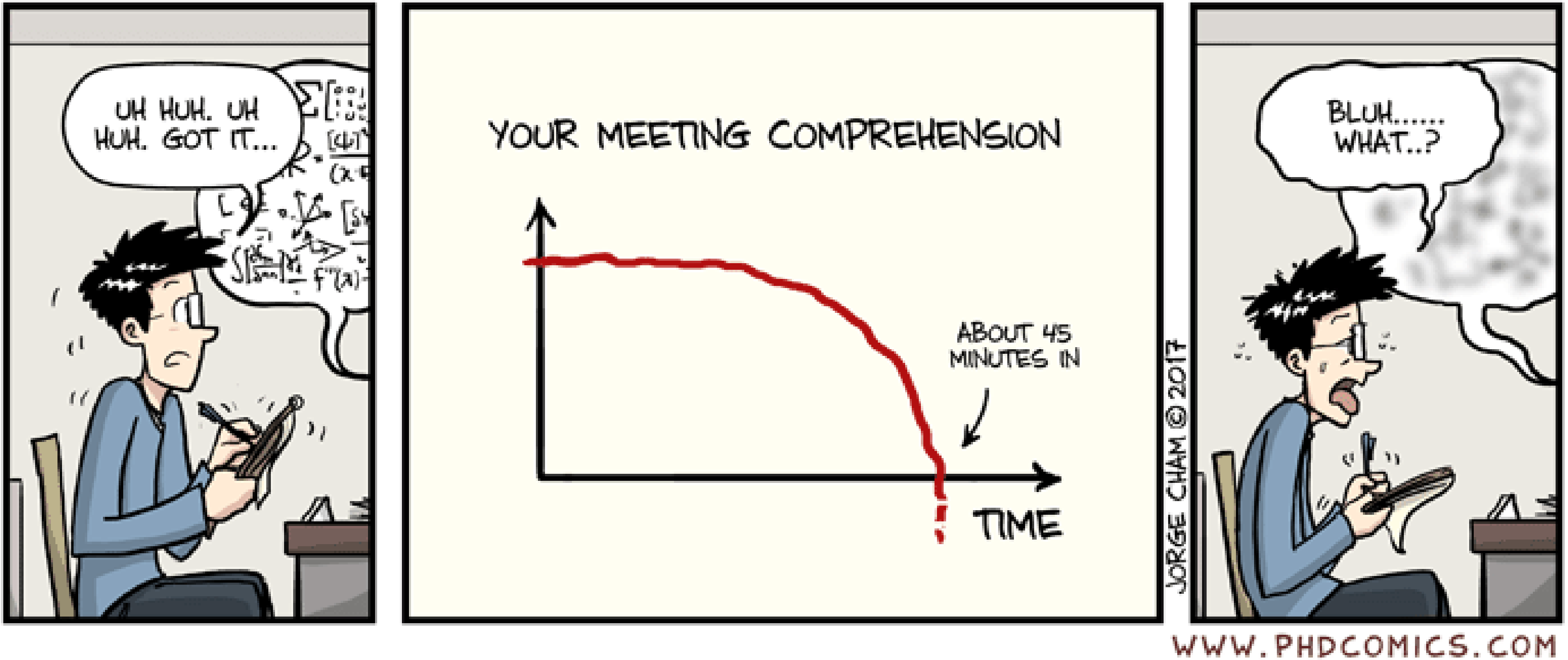
Northeastern University

[2]Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

qin.ke@husky.neu.edu

Motivation


UH HUH. UH HUH. GOT IT...
YOUR MEETING COMPREHENSION
ABOUT 45 MINUTES IN
TIME
JORGE CHAM © 2017
BLUH...... WHAT..?
WWW.PHDCOMICS.COM

Questions

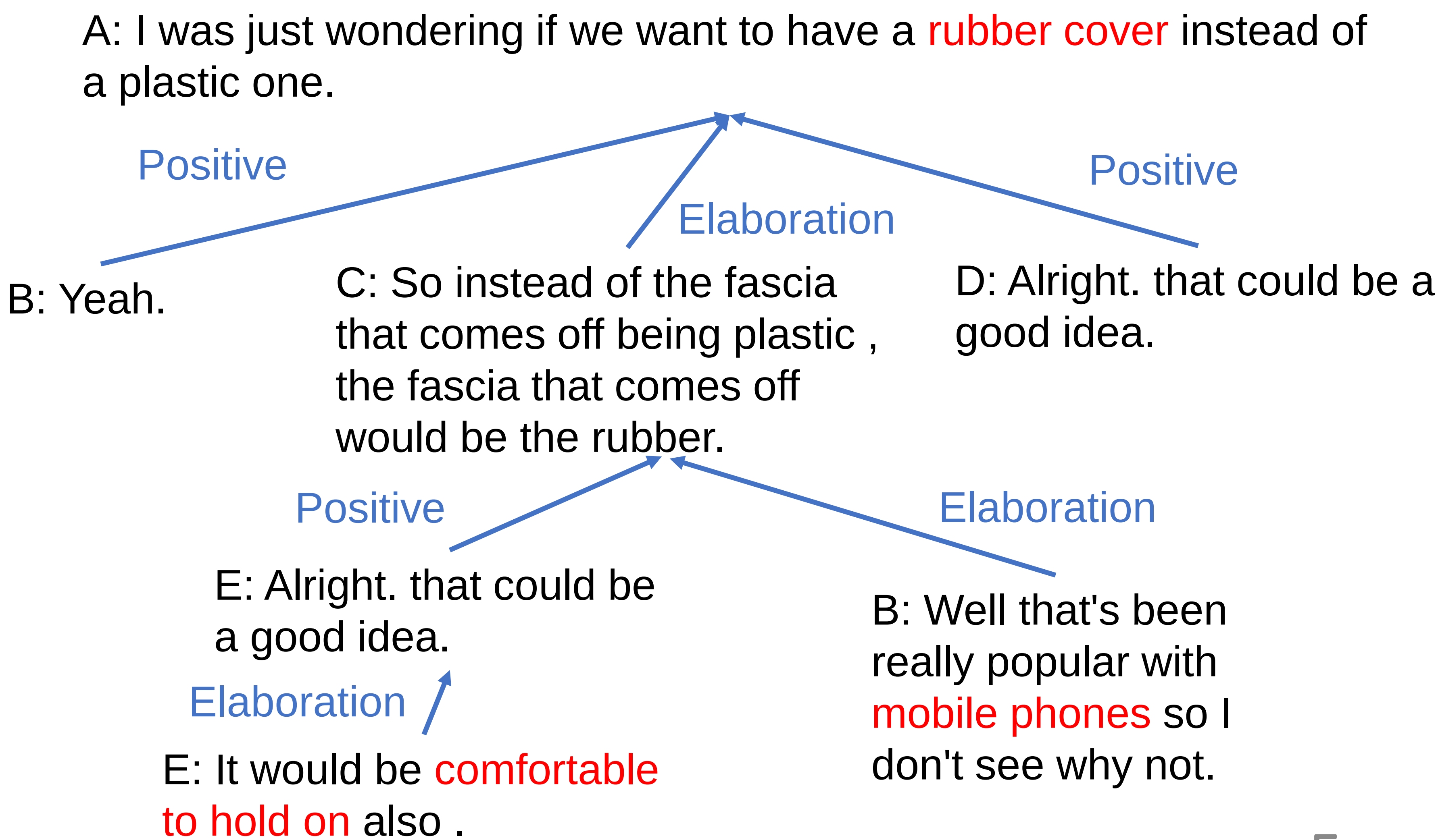
- How to pinpoint and extract salient content from a meeting?

- How to understand and model the effectiveness of a meeting?

Related Work

- Extract salient content from the meeting
  - Jones (1999)
  - Fernandez et al. (2008)
  - Riedhammer et al. (2010)
  - Wang and Cardie (2012)
- Leverage discourse information to extract important information from meetings
  - Murray et al. (2006)
  - Galley (2006)

Content and discourse are intertwined



Unfortunately

- Discourse parsing in dialogues is still a challenging problem

Contributions

- Propose a framework to model the interaction between discourse and content in the meeting
- Model the consistency of understanding to learn the effectiveness of the meeting

Outline

- Introduction
- Methodology
- Corpus and Annotation
- Evaluation
- Consistency of Understanding
- Conclusion

Notations

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

B: Is a kinetic one going to be able to supply enough power?

Notations

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

B: Is a kinetic one going to be able to supply enough power?

x: discourse Unit

on argument level

Notations

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

Uncertain

d: Discourse Relation

$d = [uncertain]$

B: Is a kinetic one going to be able to supply enough power?

Notations

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

B: Is a kinetic one going to be able to supply enough power?

c: Candidate Phrases

$\mathrm{c}_D = [unimportant, unimportant, important]$

$\mathrm{c}_B = [important]$

Generic Framework

Important candidate phrase

Discourse relation

$$P(c, d|x, w)$$

Input: discourse unit

Model parameters

Generic Framework

- A log-linear model

$$P(c, d|x, w) = \frac{\exp[w \cdot \varphi(c, d, x)]}{\sum_{c,d} \exp[w \cdot \varphi(c, d, x\ )]}$$

$$\propto \exp[w \cdot \varphi(c, d, x)]$$

Model Parameters

Features

Generic Framework

- A log-linear model

$$P(c, d \mid x, w) \propto \exp[w \cdot \varphi(c, d, x)]$$

$$\propto \exp\left[\underbrace{w_c \cdot \sum \varphi_c(c, x)}_{\text{Content features}} + \underbrace{w_\mathrm{d} \cdot \varphi_d(d, x)}_{\text{Discourse features}} + \underbrace{w_{cd} \cdot \sum \varphi_{cd}(c, d, x)}_{\text{Joint features}}\right]$$

Content features

Discourse features

Joint features

E.g., whether the head word of the phrase was mentioned in preceding turn

E.g., similarity between two discourse units

Generic Framework

- A log-linear model

$$P(c, d \mid x, w) \propto \exp[w \cdot \varphi(c, d, x)]$$

$$\propto \exp\left[w_c \cdot \sum \varphi_c(c, x) + w_{\mathrm{d}} \cdot \varphi_d(d, x) + w_{cd} \cdot \sum \varphi_{cd}(c, d, x)\right]$$

Joint features

E.g., whether phrases are salient when an elaboration relation is surrounded by two sentences with high similarity

Joint Learning

- SampleRank (Rohanimanesh et al., 2011)
  - Sampling-based search algorithm
  - Construct a sequence of configurations for sample labels as a Markov chain Monte Carlo (MCMC) chain
  - No limitations on the feature set

- Goyal and Eisenstein (2016)
  - On news articles summarization with Rhetorical Structure Theory (RST)
  - On sentence level
  - With simple summary features

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

Uncertain

B: Is a kinetic one going to be able to supply enough power?

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

Elaboration

B: Is a kinetic one going to be able to supply enough power?

Initialization:

Salient content phrases label: [unimportant, important, unimportant, important]

Discourse relation label: [Elaboration]

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

Uncertain

B: Is a kinetic one going to be able to supply enough power?

Old samples (initialization):
[unimportant, important, unimportant, important]
[Elaboration]

New samples:

Sampled salient content phrases:
[unimportant, unimportant, important, unimportant]

Sampled discourse:
[Uncertain]

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

Uncertain

B: Is a kinetic one going to be able to supply enough power?

Old samples (initialization):
[unimportant, important, unimportant, important]
[Elaboration]

New samples:
[unimportant, unimportant, important, unimportant]
[Uncertain]

Accept the new samples, if it improves the scoring function

$$if\ score(new) - score(old) > 0$$
$$old\_samples \leftarrow new\_samples$$

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

Uncertain

B: Is a kinetic one going to be able to supply enough power?

Old samples:
[unimportant, important, unimportant, important]
[Elaboration]

New samples:
[unimportant, unimportant, important, unimportant]
[Uncertain]

Accept the new samples, if it improves the scoring function

$if\ score(new) - score(old) > 0$

$\quad old\_samples \leftarrow new\_samples$

Update the parameters of the model based on old and new samples

Joint Inference

- Infer discourse and salient content iteratively
  - Dynamic Programming
    - Infer discourse relation
  - Integer Linear Programming
    - Infer salient phrase candidate

Joint Model with latent discourse

- Discourse relation as latent variable

Important candidate phrase

Discourse relation

$$P(c\,|x,w) \propto \sum_{d} P(c,d|x,w)$$

Input: Discourse unit

Model parameters

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

x

B: Is a kinetic one going to be able to supply enough power?

Old samples:
[unimportant, important, unimportant, important]
[discourse_type_1]

New samples:
[unimportant, unimportant, important, unimportant]
[discourse_type_2]

Accept the new samples, if it improves the scoring function

$$if\ score(new) - score(old) > 0$$
$$old\_samples \leftarrow new\_samples$$

Update the parameters of the model based on old and new samples

SampleRank

D: Two different types of batteries. Um can either use a hand dynamo, or the kinetic type ones.

x

B: Is a kinetic one going to be able to supply enough power?

Old samples:
[unimportant, important, unimportant, important]
[discourse_type_1]

New samples:
[unimportant, unimportant, important, unimportant]
[discourse_type_2]

Accept the new samples, if it improves the scoring function

$$\text{if } score(new) - score(old) > 0$$
$$old\_samples \leftarrow new\_samples$$

Update the parameters of the model based on old and new samples

Outline

- Introduction
- Methodology
- Corpus and Annotation
- Evaluation
- Consistency of Understanding
- Conclusion

Meeting Corpora

- AMI meetings (Carletta et al., 2006)
  - Annotated with abstractive summaries, argumentative discourse units, and argumentative discourse relations (Twente Argumentation schema by Rienks et al. 2005)

- ICSI meetings (Janin et al., 2003)
  - Annotated with salient content label

Outline

- Introduction
- Methodology
- Corpus and Annotation
- Evaluation
- Consistency of Understanding
- Conclusion

Evaluation

- Content selection - Extractive summarizer
- Discourse relation prediction - Discourse parser

Baselines and Comparisons: Summarization

- Longest Dialogue Act
- Centroid Dialogue Act
- Support Vector Machine (SVM)
- Keyword Extractive Approach (Liu et al., 2016)
  - Heuristic method using linguistic features
  - For fair comparison, we change it to keyphrase
  - State-of-the-art

Extractive Summary

	Length of Summary	ROUGE_1_F1	ROUGE_SU4_F1
Longest Dialogue Act	30.9	23.1	15.3
Centroid Dialogue Act	17.5	20.8	11.3
SVM Baseline	49.8	27.5	11.8
Keyword Extraction (Liu et al., 2016)	62.4	36.2	13.5
Joint Model	66.6	41.1	20.9
Joint Model with Latent Discourse	85.9	42.4	21.3

Rouge_1: Unigrams
Rouge_SU4: Skip-bigrams with at most 4 words in between

Discourse Relation Prediction

- 9 discourse relations in predefined discourse relations set from Twente Argumentation schema by Rienks et al. (2005)

Baselines and Comparisons: Discourse

	Accuracy	F1
Majority Label	51.2	7.5
SVM Baseline	51.2	22.8

Support Vector Machine (SVM)

- 5-fold Cross Validation
- With the same feature set as our joint model

Baselines and Comparisons: Discourse

	Accuracy	F1
Majority Label	51.2	7.5
SVM Baseline	51.2	22.8
Neural Language Model (Ji et al., 2016)	54.2	21.4

Neural Language Model (Ji et al., 2016)

- State-of-the-art
- Propose a novel latent variable recurrent neural network architecture for jointly modeling sequences of words and discourse relations

Baselines and Comparisons: Discourse

	Accuracy	F1
Majority Label	51.2	7.5
SVM Baseline	51.2	22.8
Neural Language Model (Ji et al., 2016)	54.2	21.4
Joint Model	59.2	23.4

Outline

- Introduction
- Methodology
- Corpus and Annotation
- Evaluation
- Consistency of Understanding
- Conclusion

Consistency of Understanding

- Compare participant summaries to determine whether participants report the same decisions (Kim et al., 2016)

- Binary Classification Task
  - consistent vs. inconsistent

Our Model - Features

- Consistency Probability
  – Probability of consistent understanding:
  $$max_{c,d}P(c,d|x,w_{consistent})$$

  - Probability of inconsistent understanding:
  $$max_{c,d}P(c,d|x,w_{inconsistent})$$

Our Model - Features

- Consistency Probability

- Discourse Relation
  - Based on our study, there is high correlation between discourse information and consistency of the meeting
  - Positive, Negative (+)
  - Request, Specialization (-)
  - Unigram and bigram discourse relations

Our Model - Features

- Consistency Probability
- Discourse Relation

- Word Entrainment (Nenkova et al., 2008)
  - People tend to use similar words as the meeting proceeds
  - This phenomenon is very likely to be detected in effective meetings, when participants are on the same page in the meeting

Baselines and Comparisons

- Support Vector Machine (SVM)
  - Leave-one-out
  - Unigram and bigrams

- Hidden Markov Model (HMM) (Kim et al., 2016)
  - State-of-the-art
  - Discourse and head gesture

Results

	Accuracy	F1
Majority Label	66.7	40.0
SVM Baseline	51.2	50.6
Hidden Markov Model (Kim et al., 2016)	60.5	50.5
Oracle Discourse Relation	69.8	62.7
Oracle Word Entrainment	61.2	57.8

Results

	Accuracy	F1
Majority Label	66.7	40.0
SVM Baseline	51.2	50.6
Hidden Markov Model (Kim et al., 2016)	60.5	50.5
Oracle Discourse Relation	69.8	62.7
Oracle Word Entrainment	61.2	57.8
Our Model	68.2	63.1

Conclusion

- Propose a flexible framework to jointly model content and discourse. We achieve good performance on discourse recognition and salient content extraction tasks

- By using the outputs of our model, our system is able to learn the consistency prediction task

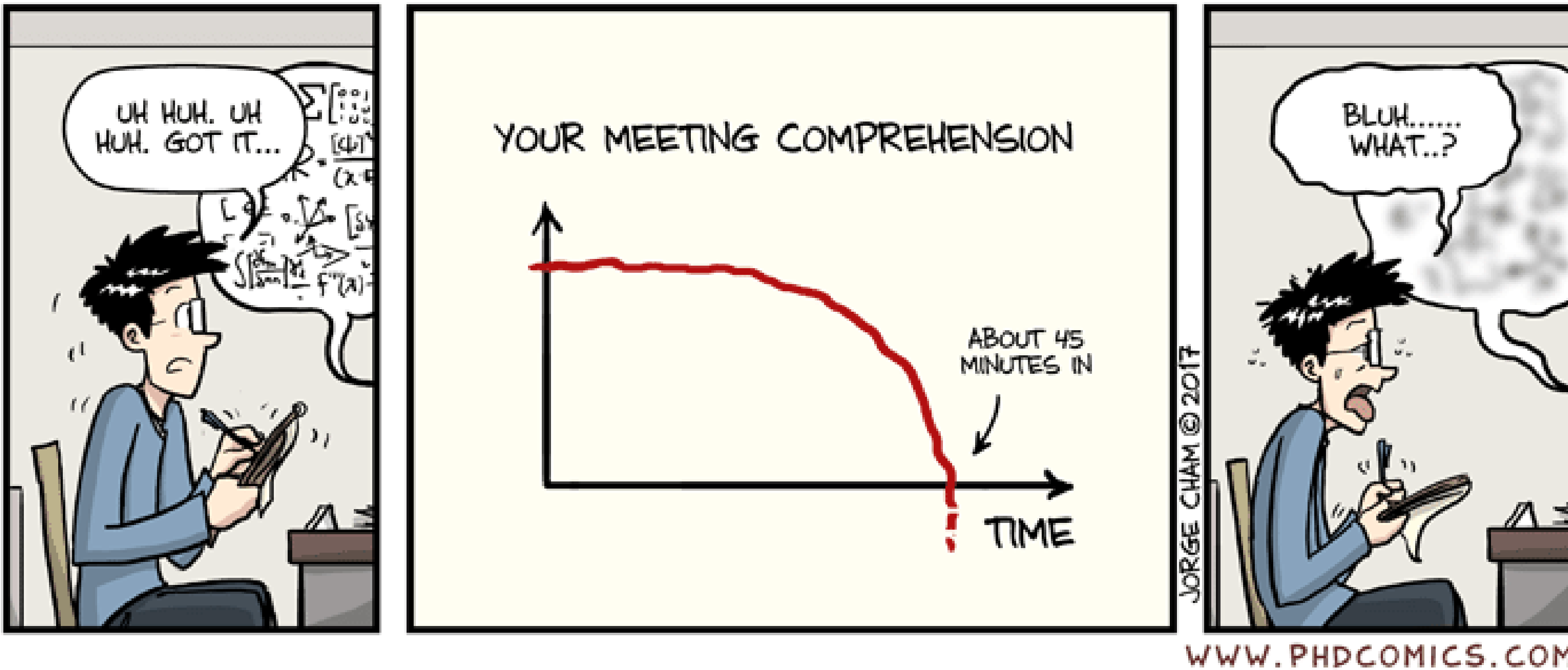
Future Work

- How to model idea flows among participants?
  - Which fraction of ideas are discussed and what is the outcome?
  - Which fraction of ideas are not discussed thoroughly and why?
- How can we leverage the discourse to capture the idea generation process?

Resources

- Project website (code & data):
  - http://www.ccs.neu.edu/home/kechenqin/paper/acl2017.html
- Consistency data download:
  - http://people.csail.mit.edu/joseph_kim/data/cou_ami.zip
- Contact:
  - qin.ke@husky.neu.edu

All for survival!



Thank you!

- Any Questions?