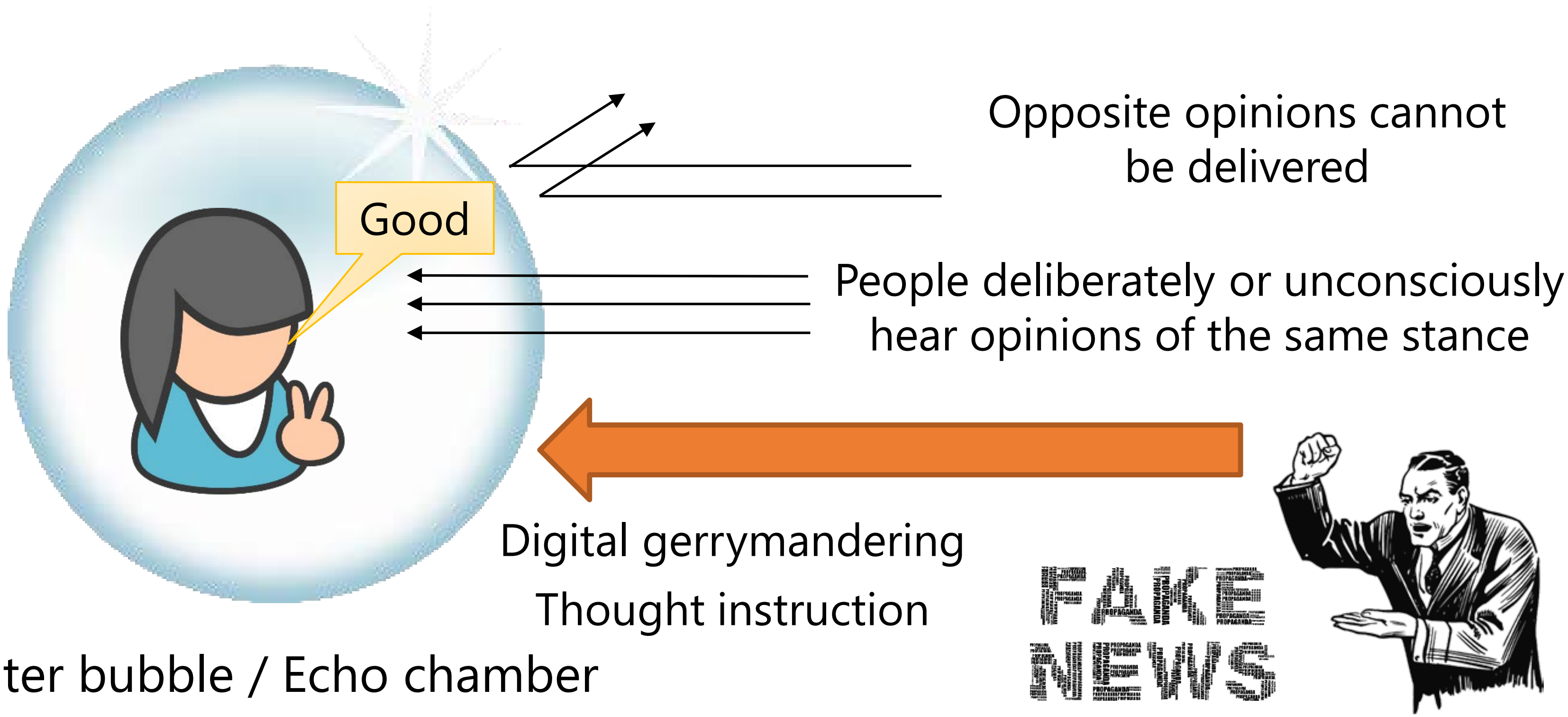
Knowledge Acquisition from Web and Opinion Analysis

Naoaki Okazaki (okazaki@c.titech.ac.jp)
School of Computing
Tokyo Institute of Technology
http://www.chokkan.org/
@chokkanorg

Information diffusion in the era of social media



We need the technology for consensus building and decision making under the variety of opinions

Stance detection (Mohammad+ 2016)

Topic: TPP

Impractical to prepare supervision data for every topic, e.g., Trump, nuclear plant…

I totally disagree with TPP. (No knowledge required) → Against

It's better to promote domestic consumption. → Against

It's better to promote free trade. → Favor

Suppressed by TPP

Promoted by TPP

agree

disagree

$P(y|\boldsymbol{x})$ $\hat{y}$

Topic-independent

Topic-specific

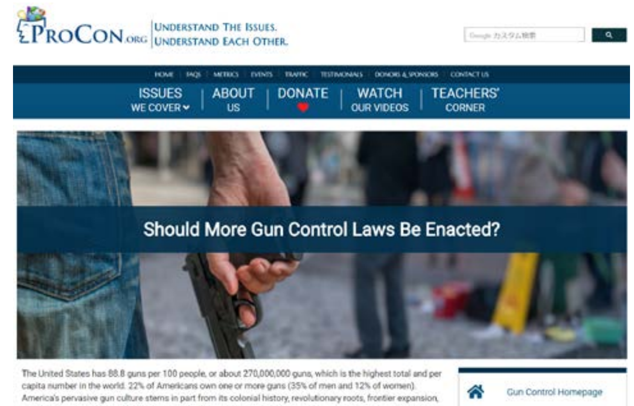
Acquire topic-specific knowledge not from supervision data but from other data (e.g., Wikipedia)

Stance detection needs topic specific/independent knowledge

Applications of stance detection

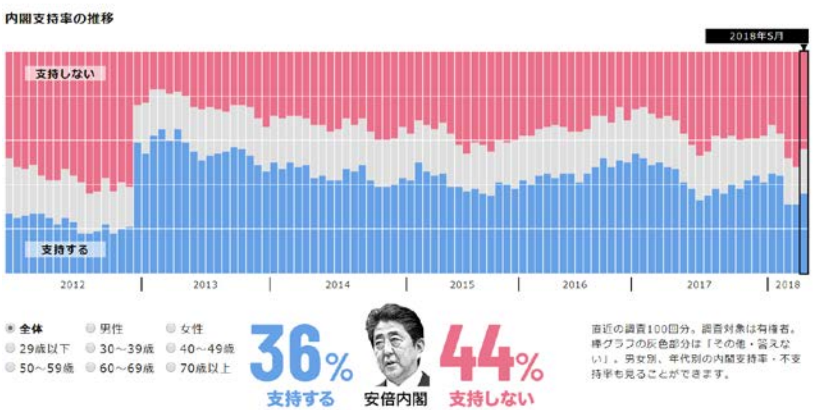
Automatic debating

https://www.youtube.com/watch?v=7g59PJxbGhY



Argumentation mining

https://www.procon.org/



Public opinion survey

http://www.asahi.com/politics/yoron/

Detecting fake news

http://www.fakenewschallenge.org/

Opinion analysis requires world knowledge
(Saint-Dizier 2016, Hanawa+ 2017, Moens 2018)

- An example that requires the world knowledge

Technology negatively influences how people communicate. Some people use their cellphone constantly and do not even notice their environment.

  - is-a(cellphone, technology)
  - used-for(cellphone, communication)

⇒ The 2nd sentence is a warrant of the 1st sentence

- The world knowledge is essential to opinion analysis
  - 78% of relation recognitions between argumentation units require the world knowledge (Saint-Dizier 2016)
  - 40.2% of tweets require the world knowledge for identifying stances of tweets (Hanawa+ 2017)

Opinion analysis with the world knowledge

- Argumentation mining with the world knowledge
  - Gaps between claims (Boltuzic+ 2016)
  - Argument reasoning comprehension (Habernal+ 2018)
  - Stance detection with attention (Hanawa+, under review)

- Acquiring knowledge from the Web
  - From Wikipedia (Hanawa+ 2017)
  - From Twitter (Sasaki+ 2017; Sasaki+ 2018)

Filling the gap between claims (Boltuzic+ 2016)

Marijuana is not taxed, and those who sell it are usually criminals of some sort.

If something is not taxed, criminals sell it.

Criminals should be stopped from selling things.

Things that are taxed are controlled and regulated by the government.

Implicit premises to fill the gap

Legalized marijuana can be controlled and regulated by the government.

It is easy for humans to infer that these claims are in the same stance,
but this is extremely difficult for computers

Filling the gaps between claims (Boltuzic+ 2016)

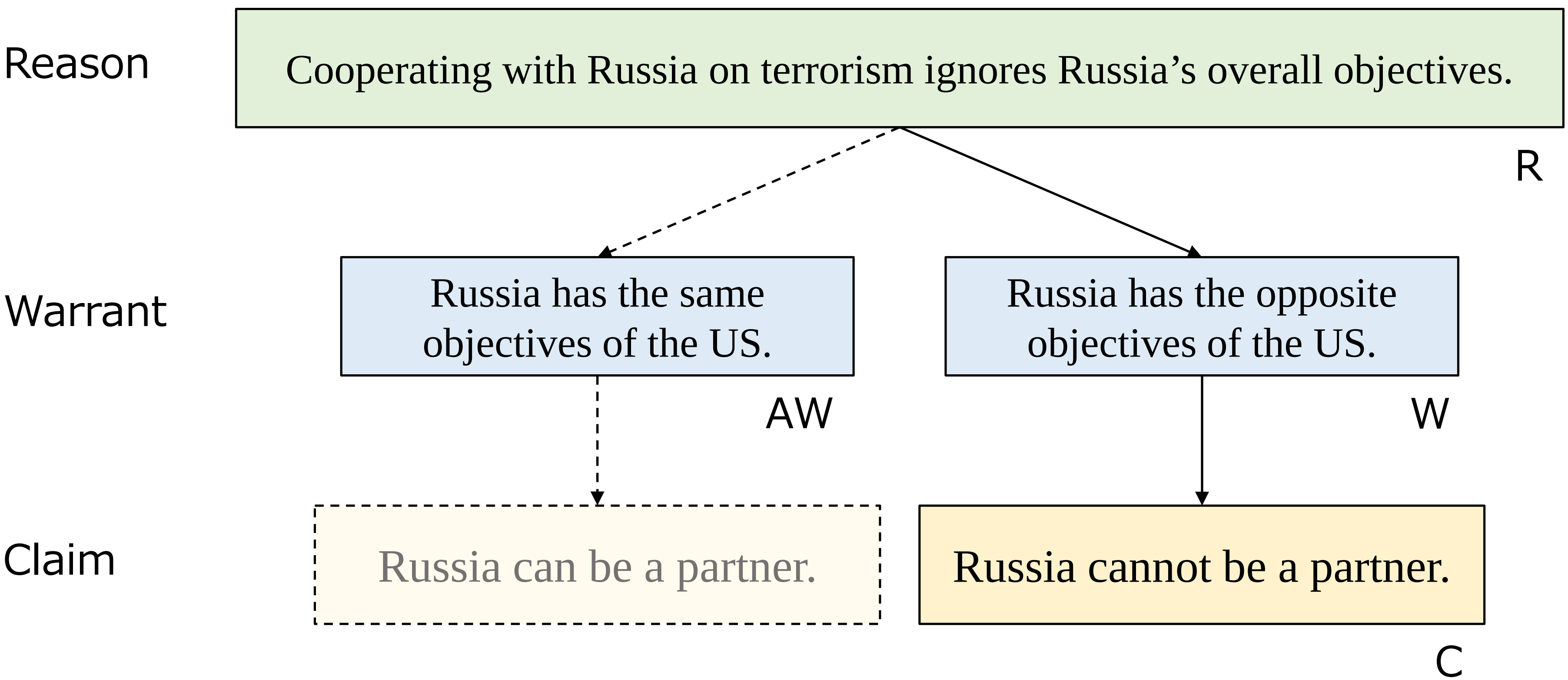
- Building a dataset for filling the gaps between claims
  - They use an existing dataset (Hasan+ 2014)
  - Three humans fill the gaps between 500 pairs of claims

- Findings
  - Negative correlation between the number of gaps and the similarity of claims
  - Filling the gaps improved the performance on automatic claim matching
  - No consistency in premises filled by humans (right figure)

User claim:	*It would be loads of empathy and joy for about 6 hours, then irrational, stimulant-induced paranoia. If we can expect the former to bring about peace on Earth, the latter would surely bring about WWIII.*
Main claim:	*Legalization of marijuana causes crime.*
A1 Premise 1:	*Marijuana is a stimulant.*
A1 Premise 2:	*The use of marijuana induces paranoia.*
A1 Premise 3:	*Paranoia causes war.*
A1 Premise 4:	*War causes aggression.*
A1 Premise 5:	*Aggression is a crime.*
A1 Premise 6:	*"WWIII" stands for the Third World War.*
A3 Premise 1:	*Marijuana leads to irrational paranoia which can lead to commiting a crime.*

Argument reasoning comprehension

(Habernal+ 2018) (cloze-style gap filling)



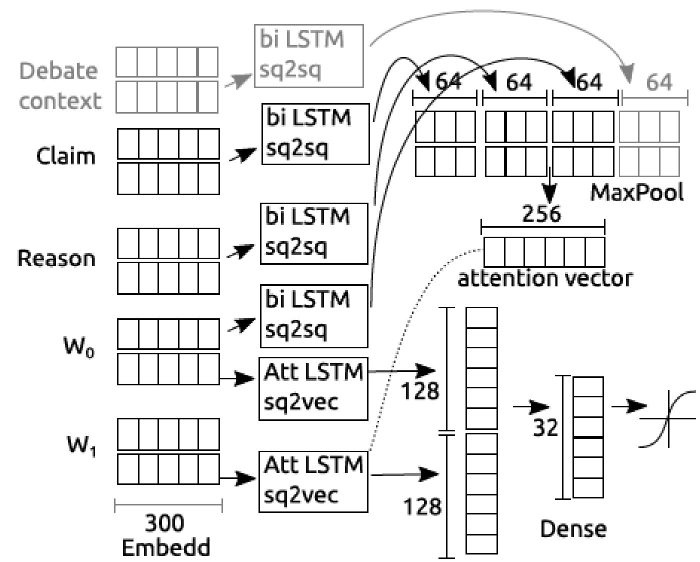
- Two warrants are given to a pair of a reason and claim
- Choose a claim that is suitable to connect the reason and claim
- The other warrant was prepared to conclude the opposite claim
  - This study call this warrant *alternative warrant (AW)*

Argument reasoning comprehension
(Habernal+ 2018)

- Creation of the dataset
  - Room for Debate in New York Times
  - Issued eight micro tasks to crowd sourcing
  - 1,970 tuples of (C, R, W, AW)
  - Human accuracy (choosing W instead of AW) was 79.8% (average workers) and 90.9% (trained workers)

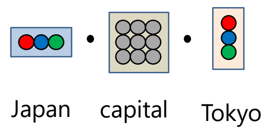
https://www.nytimes.com/roomfordebate/2017/01/19/media-in-the-age-of-trump

- Automatic classification of W and AW
  - Encoder-decoder model with attention mechanism
  - Accuracy was 56.0% (much lower than the human performance

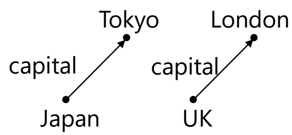


Integrating knowledge into DNNs

- RESCAL (Nickel+ 11)
  - $\mathrm{score}(s, r, t) = x_s^{\mathrm{T}} W_r x_t$
  - $x_s \in \mathbb{R}^d, x_t \in \mathbb{R}^d, W_r \in \mathbb{R}^{d \times d}$

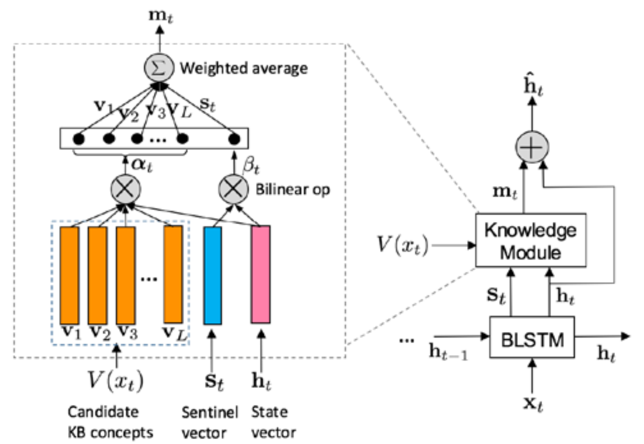


- TransE (Bordes+ 13)
  - $\mathrm{score}(s, r, t) = -\|x_s + w_r - x_t\|_2^2$
  - $x_s \in \mathbb{R}^d, w_r \in \mathbb{R}^d, x_t \in \mathbb{R}^d$

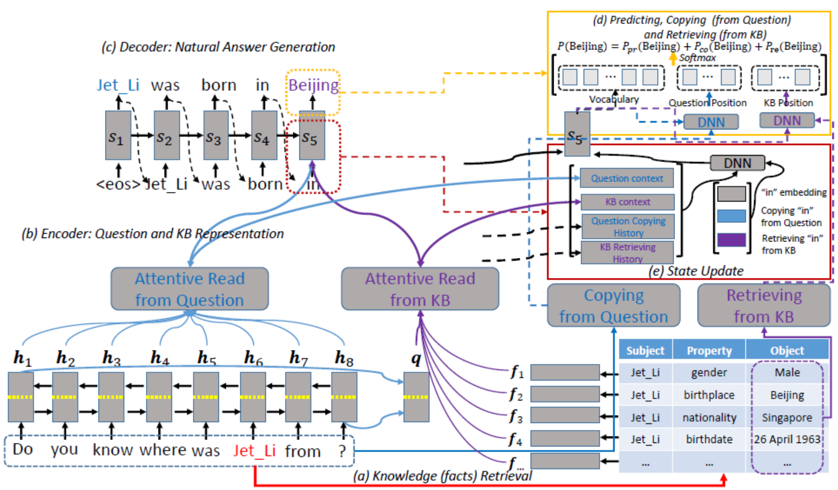


- Train W_r or w_r (and x_s, x_t) by max-margin loss

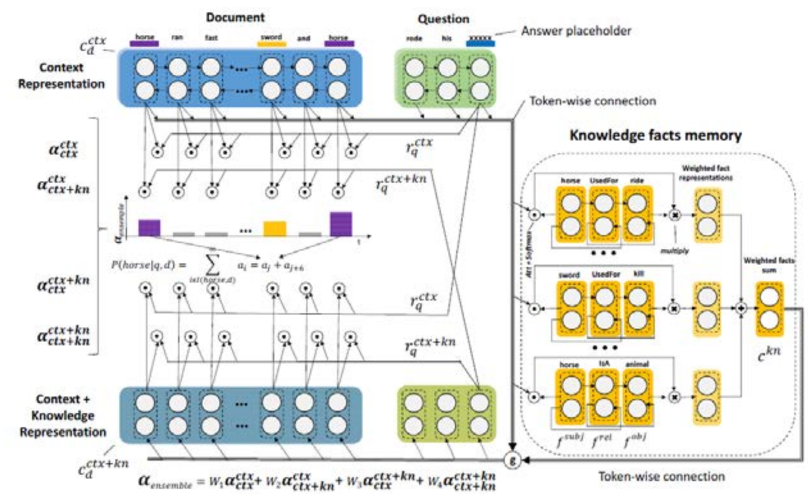
Representation learning for KBs



KBLSTM (Yang+ 17)



COREQA (Shizhu+ 17)



Knowledgeable Reader (Mihaylov+ 18)

Our attempts

Collaborators

Akira Sasaki
Recruit Technologies
(PhD at Tohoku University)

Kazuaki Hanawa
Tohoku University

Kentaro Inui
Tohoku University
Riken AIP

Stance detection with external knowledge

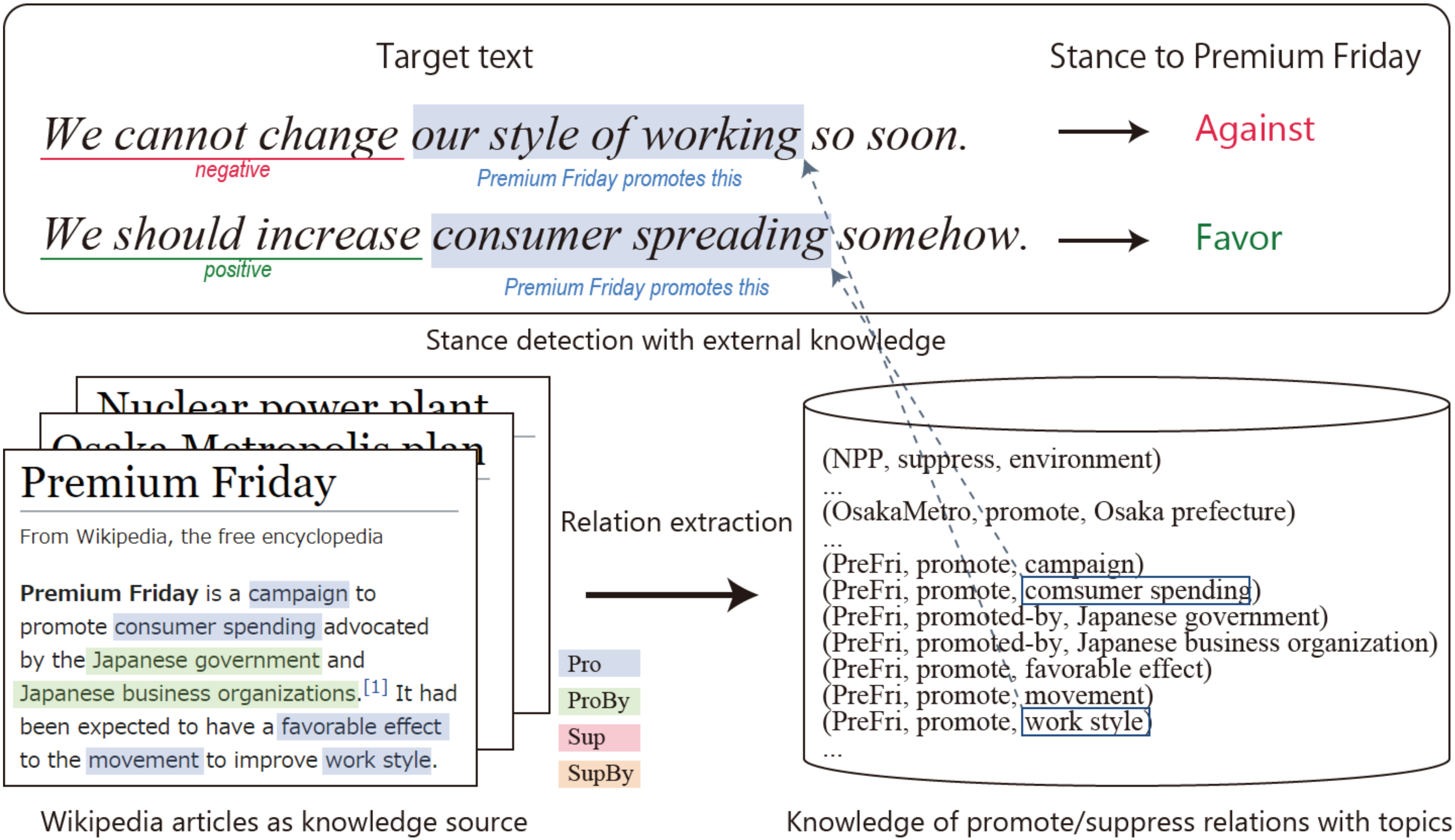


Table of Contents

- Building a corpus for stance detection
  - Does external knowledge really matter?
- Acquiring topic knowledge
  - Reading Wikipedia articles for extracting causal (promote/suppress) relations
  - Analyzing SNS posts for extracting inter-topic preferences
- Detecting stance by attending knowledge

A dataset for stance detection
(Hanawa+, under review)

- Includes 7 topics, 2000 tweets for each topic
- Labeled by crowd workers to 👍, 👎, or none
- Uses tweets where 4 out of 5 workers agreed
  - Other tweets were discarded (shown as 🆖 below)

Topic	👍	👎	none	🆖
大阪都構想 (Osaka Metropolis plan)	239	259	380	1122
安保法案 (2015 Japanese military legislation)	168	352	262	1218
プレミアムフライデー (Premium Friday)	153	744	218	885
Trans-Pacific Partnership (TPP)	53	802	230	915
原発 (nuclear power plant)	47	783	202	968
集団的自衛権 (right of collective self-defense)	160	468	196	1177
共謀罪 (Anti-Conspiracy Bill)	86	592	308	1014
Total	906	4000	1795	7299

Necessity of domain knowledge
(Hanawa+, under review)

- We manually examined the necessity of external knowledge by using 10% samples of the dataset
- We found that **40.1% of the examined instances require the topic-specific knowledge** for detecting stances

Necessary knowledge	%	Statement example
No topic knowledge	56.3	Nuclear power plant is absolutely necessary (NPP: 👍)
Promote/suppress (in Wikipedia)	26.3	Custom should function (TPP: 👎)
Promote/suppress (not in Wikipedia)	13.9	I'm worrying about gene-altered foods (TPP: 👎)
Other types of knowledge	2.5	Do you want another Public Order and Police Law? (2015 Japanese military legislation: 👎)

Reading Wikipedia articles for relation extraction (Hanawa+ 2017)

Premium Friday

From Wikipedia, the free encyclopedia

Premium Friday is a campaign to promote consumer spending advocated by the Japanese government and Japanese business organizations.[1] It had been expected to have a favorable effect to the movement to improve work style.

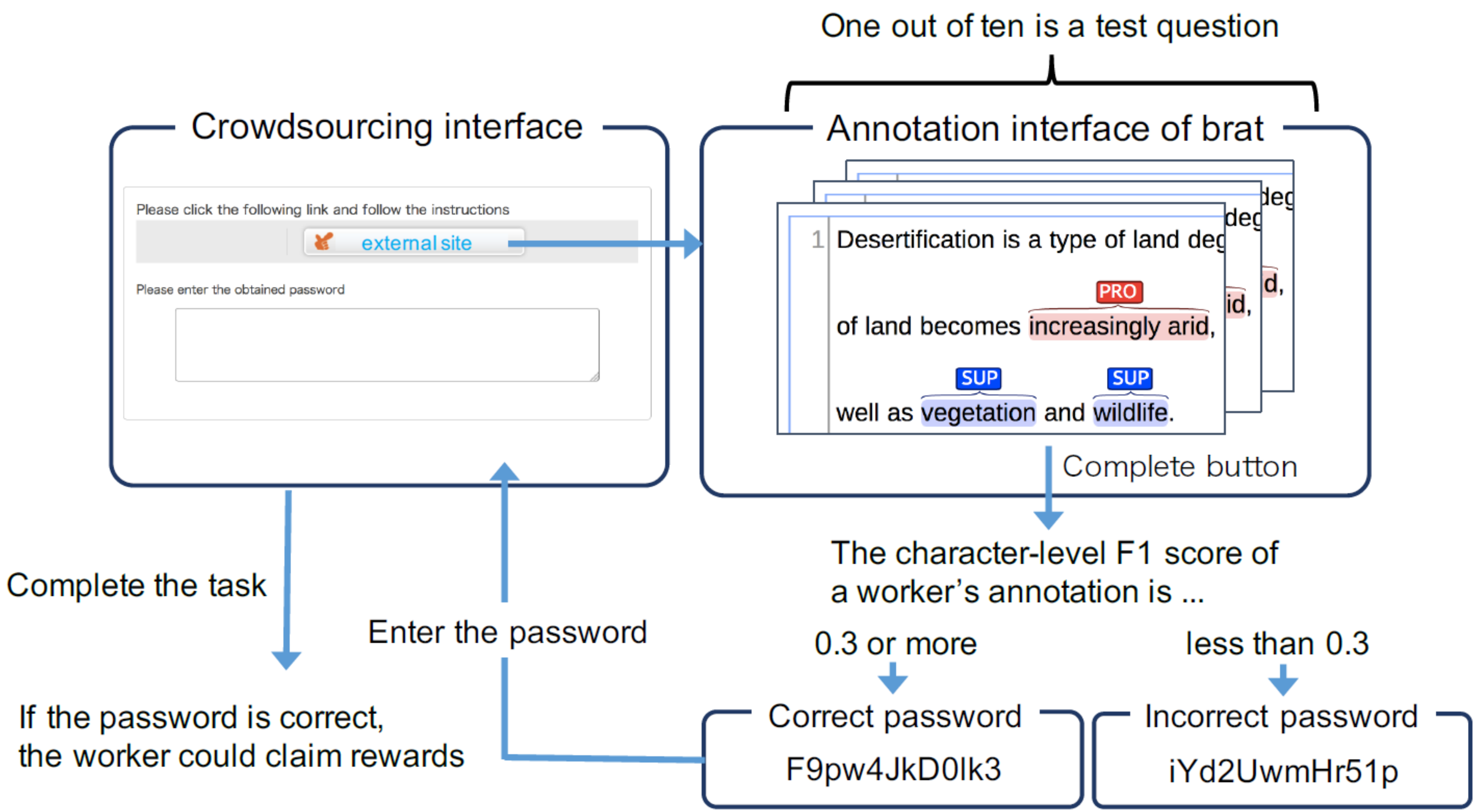
Background [edit]

It is a campaign promoted by the Japanese business sector led by the Government of Japan and Japan Business Federation recommending to spend more enriched life every last Friday each month. Tied up with a movement to improve work style, it is recommending to end work at 15:00. Since last Friday of a month end tends to be when salary has just been paid, it is advocating to spend the afternoon shopping or travelling. This has been implemented on February 24, 2017.

Promote Promoted_by Suppress Suppressed_by

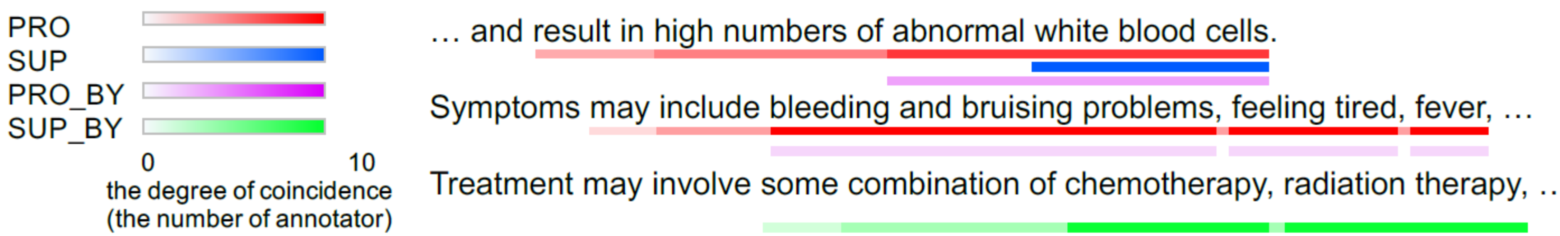
- Treat a title of a Wikipedia article as a subject of a relation
  - We can avoid various problems (e.g., coreferences, paraphrases) in RE

Collecting annotations via crowdsourcing
(Hanawa+ 2017)



Annotation results
(Hanawa+ 2017)

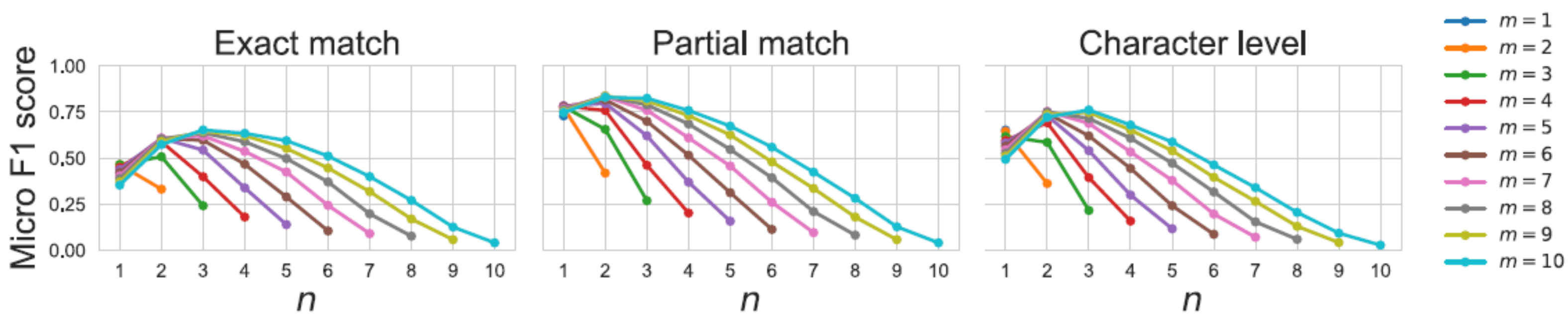
- 10 annotations for 1494 articles of 9 categories:
  - Social issues, Disasters, Diseases and disorders, Innovation, Policy, Finance, Energy technology, Biomolecules, Nutrients
- Annotation excerpt for “Leukemia” article



  - We did not specify annotation boundaries to the workers (e.g., noun or verb phrases)
  - Nested spans observed between PRO and SUP
  - Cloud workers often confuse direction of causality
- Annotation results are available at:
  - http://www.cl.ecei.tohoku.ac.jp/wikipedia_pro_sup/

Quality assessment of annotations
(Hanawa+ 2017)

- Agreement between the gold-standard data and n-match aggregation from m annotations



  - Recommended setting for obtaining good agreement:
    - Extracting spans at least two annotators agreed
    - Receive at least five annotations for each article

- Applying 2-match aggregation to the data:
  - 7624 PRO, 2923 SUP, 5387 PRO_BY, and 1127 SUP_BY annotations

Automatic extraction of causal relations (Hanawa+ 2017)

- Using 2-match aggregation as training data
  - IOB2 notation (e.g., B-PRO, I-PRO)
  - One layer bi-directional LSTMs for labeling words
  - Occurrences of title phrases are replaced with `__TITLE__`

Label	precision	recall	F1
PRO	0.507	0.364	0.424
SUP	0.354	0.275	0.310
PRO_BY	0.470	0.344	0.397
SUP_BY	0.259	0.178	0.211

  - F1 scores are relatively low (but understandable)
    - Annotator agreement was approximately 0.5 F1 score

Acquiring topic knowledge from SNS
(Sasaki+ 2017)

- Similar idea to item recommendation
  - “Other items you may also like” (based on purchase history)
  - “Other topics you may also like/dislike” (based on tweets)
- Collect a number of tuples (user, topic, 👍 or 👎) from Twitter
- Store the tuples as a 2D matrix
- Apply matrix factorization to complete missing values

	Topic 1	Topic 2	Topic 3	Topic 4
User 1	1.0		-1.0	
User 2	-1.0		0.7	
User 3	-0.4		1.0	-1.0
User 4		0.5		

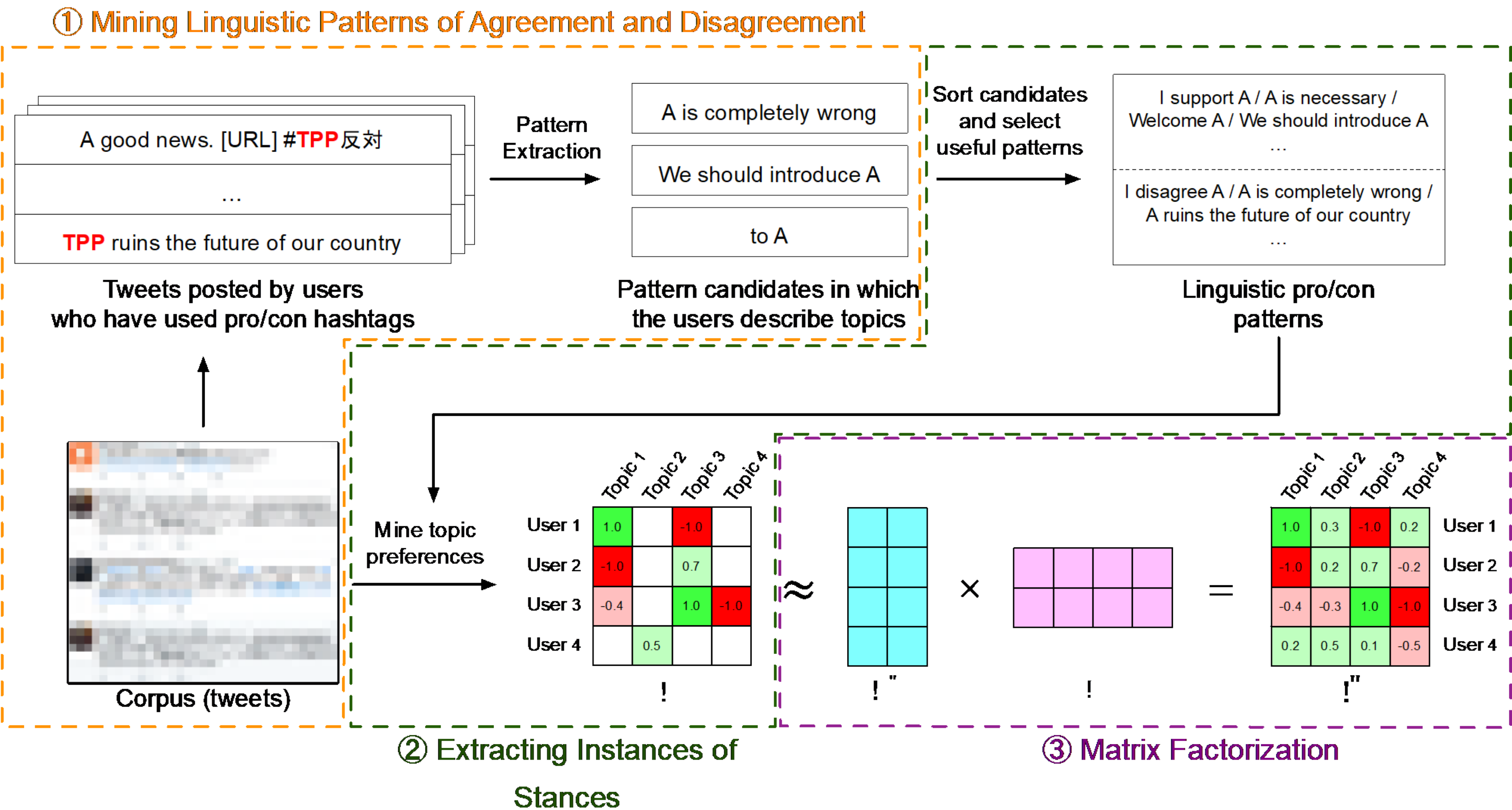
R

Topic 1	Topic 2	Topic 3	Topic 4	
1.0	0.3	-1.0	0.2	User 1
-1.0	0.2	0.7	-0.2	User 2
-0.4	-0.3	1.0	-1.0	User 3
0.2	0.5	0.1	-0.5	User 4

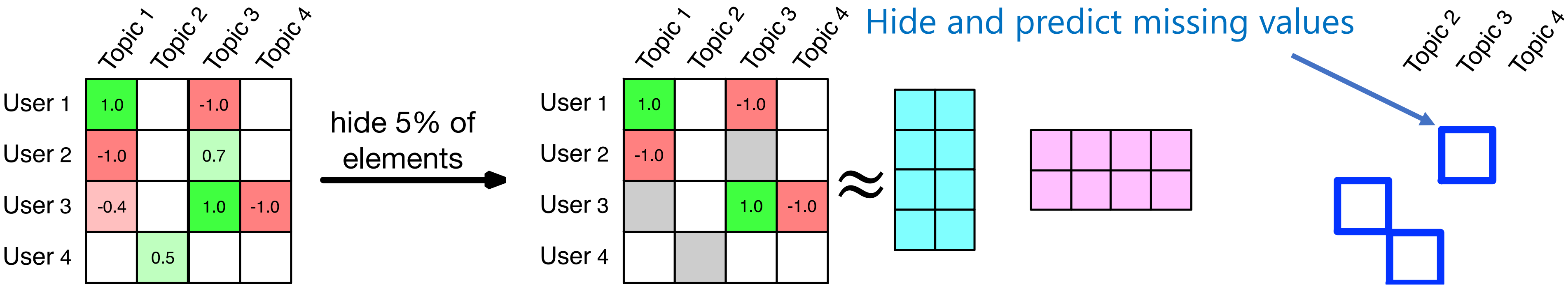
r and topics’ dense
feature vector via matrix factorization

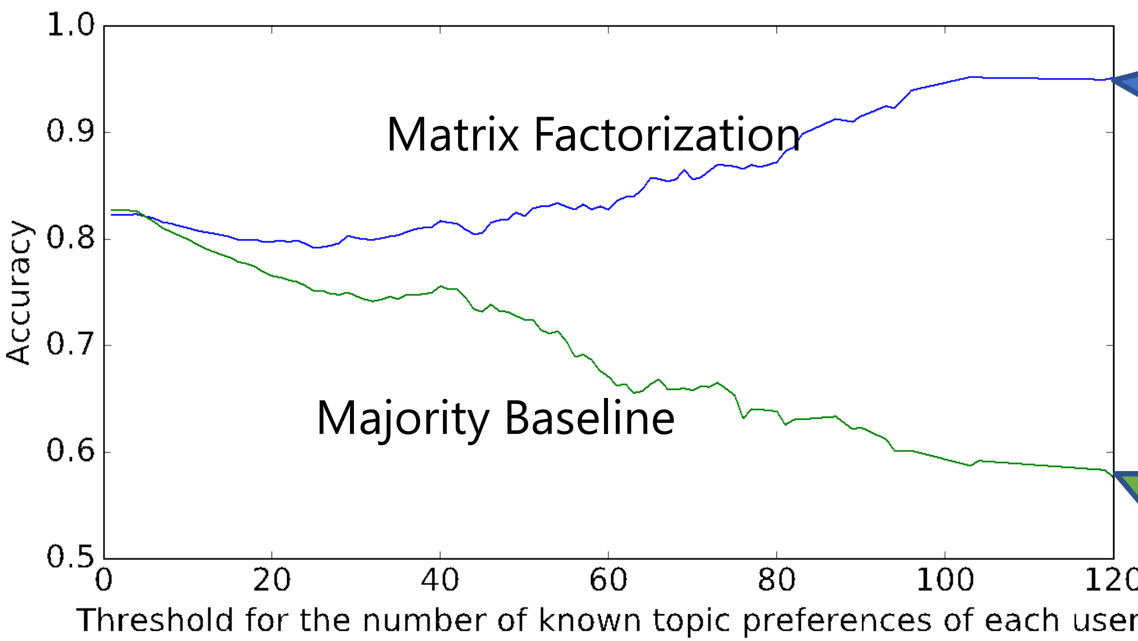
Complete missing
values by
feature vectors

Acquiring topic knowledge from SNS
(Sasaki+ 2017)



Experiment: predicting missing stances
(Sasaki+ 2017)





Our approach predicts missing topic preferences of 82-95% accuracy

Majority baseline cannot predict preferences of vocal users, whose preferences are deviated from those of the average

Example of predicted stances
(Sasaki+ 2017)

Agreed with:
- Regime change
- Capital relocation

Disagreed with:
- Abe's cabinet
- Okinawa military base
- Nuclear weapons
- TPP

Prediction by matrix factorization

May also agree with:
- Same-sex partnership (0.9697)
- Vote for NO to the cabinet (0.9248)

May also disagree with:
- Nuclear power plant (-1.0269)
- War bill (-1.0190)
- Construction of a new base (-1.0186)

Stance classification with users' posts
(Sasaki+ 2018)

- Sasaki+ (2017) model the inter-topic preferences, but could not utilize the posts from the users
- This study considers users' posts as well as inter-topic preferences by using Factorization machines instead of Matrix Factorization

- Factorization machines:
  - Predict a target variable $\hat{y}$ for a given feature vector $(x_1 \ \dots \ x_n)$

$$\hat{y} = w_0 + \underbrace{\sum_{i=1}^{n} w_i x_i}_{\text{First order}} + \underbrace{\sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j}_{\text{Second order}}$$

  - Parameters $(w_0 \ w_1 \ \dots w_n)$ and $(\boldsymbol{v}_1 \ \dots \ \boldsymbol{v}_n)$ trained by ttfm

Applying factorization machines
(Sasaki+ 2018)

- Target variable
  - The stance of a user towards a topic

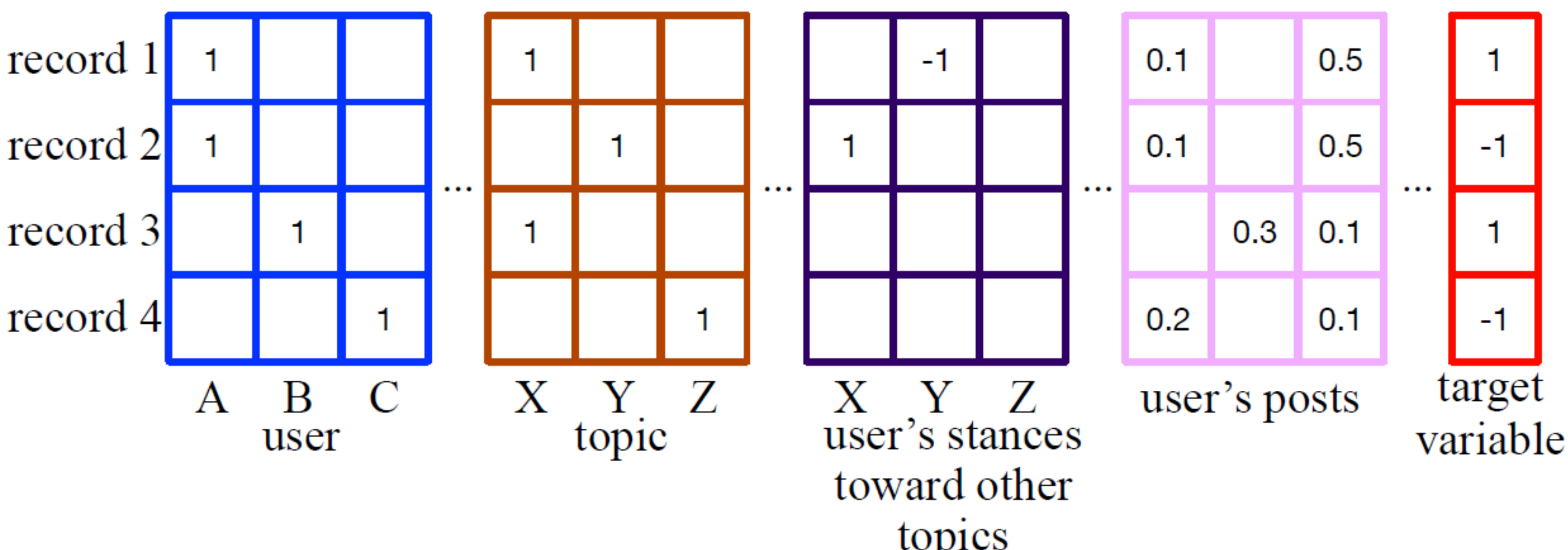
$$\frac{\#\text{positive} - \#\text{negative}}{\#\text{positive} + \#\text{negative}}$$

    - From -1 (negative stance) to +1 (positive stance)

- Features
  - User identifier
  - Topic identifier
  - User's stance towards other topics
  - User's post

Example (without users' posts)
(Sasaki+ 2018)

- The user A is favor to the topic X, but against to the topic Y.
- Record 1 presents the stance toward topic X as the target variable and the stance toward Y as other topics.
- Record 2 presents the stance toward topic Y as the target variable and the stance toward X as other topics.



Example (without users' posts)

(Sasaki+ 2018)

- Record 1: $1 = w_0 + w_{\text{user:A}} + w_{\text{topic:X}} - w_{\text{other:Y}} + \langle \boldsymbol{v}_{\text{user:A}}, \boldsymbol{v}_{\text{topic:X}} \rangle - \langle \boldsymbol{v}_{\text{topic:X}}, \boldsymbol{v}_{\text{other:Y}} \rangle - \langle \boldsymbol{v}_{\text{other:Y}}, \boldsymbol{v}_{\text{user:A}} \rangle$
- Record 2: $-1 = w_0 + w_{\text{user:A}} + w_{\text{topic:Y}} + w_{\text{other:X}} + \langle \boldsymbol{v}_{\text{user:A}}, \boldsymbol{v}_{\text{topic:Y}} \rangle + \langle \boldsymbol{v}_{\text{topic:Y}}, \boldsymbol{v}_{\text{other:X}} \rangle + \langle \boldsymbol{v}_{\text{other:X}}, \boldsymbol{v}_{\text{user:A}} \rangle$

	user				topic				user's stances toward other topics				user's posts				target variable
	A	B	C		X	Y	Z		X	Y	Z						
record 1	1				1					-1			0.1		0.5		1
record 2	1			…		1		…	1			…	0.1		0.5	…	-1
record 3		1			1									0.3	0.1		1
record 4			1				1						0.2		0.1		-1

Features for users' posts

(Sasaki+ 2018)

- Features from uni-grams, bi-grams, dependencies in users' posts

Feature type	Examples
1-gram	war
2-gram	(war, bill)
adnominal	(terrible, bill)
adjective → noun phrase	(long, working hours)
noun phrase → adjective	(train, plentiful)
noun phrase → verb	(salary level, return)

Stance detection by using users' posts
(Sasaki+ 2018)

Used information				Numbers of stances stated					Numbers of stances stated				
Topic	User	Other	Posts	≥ 0	≥ 5	≥ 10	≥ 30	≥ 50	≤ 0	≤ 5	≤ 10	≤ 30	≤ 50
✓	✓	✓	✓	62.80	62.30	63.35	72.55	85.46	65.35	62.99	62.67	62.66	62.71
✓	✓	✓		62.62	62.69	63.45	69.78	87.22	64.97	62.53	62.44	62.50	62.52
✓	✓		✓	63.34	63.22	63.76	73.70	88.11	65.24	63.40	63.21	63.18	63.24
✓	✓			62.97	62.39	63.64	70.59	88.11	65.11	63.14	62.80	62.86	62.87
✓		✓	✓	65.99	66.40	66.83	74.39	**89.43**	**66.99**	65.78	65.81	65.86	65.90
✓		✓		63.95	63.82	63.39	66.44	74.45	65.10	64.10	64.04	63.90	63.91
✓			✓	**66.45**	**66.57**	**67.23**	**75.09**	88.55	66.91	**66.37**	**66.25**	**66.31**	**66.36**
Majority baseline				63.67	62.25	60.99	55.82	55.51	65.23	64.47	64.18	63.78	63.70
Matrix factorization (topic&user)				61.12	64.17	64.56	72.55	80.18	54.31	59.63	60.48	60.95	61.05

- Can we predict the stance of every user towards a topic?
  - Users' posts increased the accuracy of stance detection
  - The more topics a user refers their stances to, the higher performance the stance detection achieves
- Accuracy for stance detection for the users who declared no stance (about 70% of the users) is estimated around 65%

Conclusions and future work

- An attempt to incorporate knowledge to DNNs
  - Building a corpus for stance detection
  - Acquiring topic knowledge from Wikipedia
  - Detecting stance by attending knowledge

- Future work
  - Expand the source for acquiring external knowledge
  - Explore an end-to-end architecture of knowledge acquisition and stance detection
    - Currently they are split into two separate models

References

- Filip Boltuzic, Jan Šnajder. 2016. Fill the gap! Analyzing implicit premises between claims from online debates. *Proc. of ArgMining*, pp. 124-133.
- Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, Benno Stein. 2018. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. *Proc. of NAACL-HLT*, pp. 1930-1940.
- Kazuaki Hanawa, Akira Sasaki, Naoaki Okazaki, Kentaro Inui. 2017. A crowdsourcing approach for annotating causal relation instances inWikipedia. *Proc. of PACLIC*.
- Shizhu He, Cao Liu, Kang Liu, Jun Zhao. 2017. Generating Natural Answers by Incorporating Copying and Retrieving Mechanisms in Sequence-to-Sequence Learning. *Proc. of ACL*, pp. 199-208.
- Todor Mihaylov, Anette Frank. 2018. Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge. *Proc. of ACL*, pp. 821-832.
- Marie-Francine Moens. 2018. Argumentation mining: How can a machine acquire common sense and world knowledge? *Argument & Computation*, 9:1–14.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. *Proc. of SemEval*, pp. 31-41.
- Patrick Saint-Dizier. 2016. Challenges of argument mining: Generating an argument synthesis based on the qualia structure. *Proc. of INLG*, pp. 79-83.
- Akira Sasaki, Kazuaki Hanawa, Naoaki Okazaki, Kentaro Inui. 2017. Other topics you may also agree or disagree: Modeling inter-topic preferences using tweets and matrix factorization. *Proc. of ACL*, pp. 398-408.
- Akira Sasaki, Kazuaki Hanawa, Naoaki Okazaki, Kentaro Inui. 2018. Predicting stances from social media posts using factorization machines. *Proc. of Coling*, (to appear).
- Christian Stab, Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3):619-659.
- Bishan Yang, Tom Mitchell. 2017. Leveraging Knowledge Bases in LSTMs for Improving Machine Reading. *Proc. of ACL*, pp. 1436-1446.