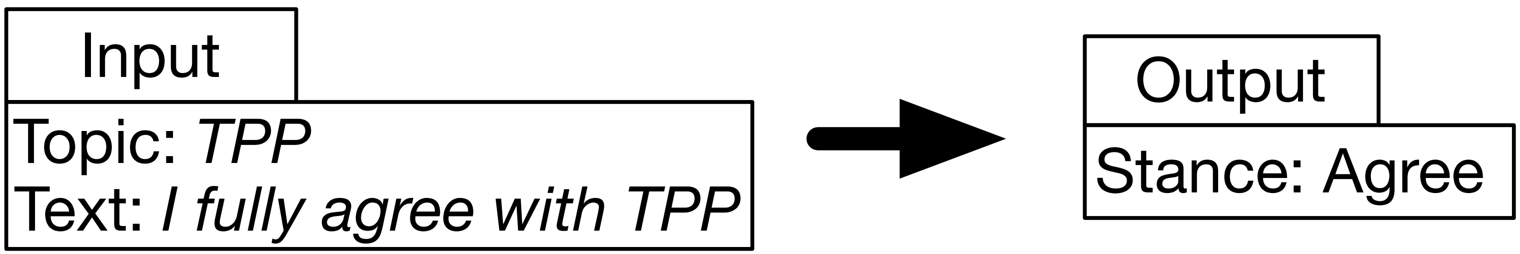
Other Topics You May Also Agree or Disagree: Modeling Inter-Topic Preferences using Tweets and Matrix Factorization

Akira Sasaki, Kazuaki Hanawa, Naoaki Okazaki, Kentaro Inui

Tohoku University

Stance classification

- **Goal**
  - Classify stances of texts in regard to a specific topic



- **Applications**
  - Public opinion survey from SNS data
  - Predicting voting actions

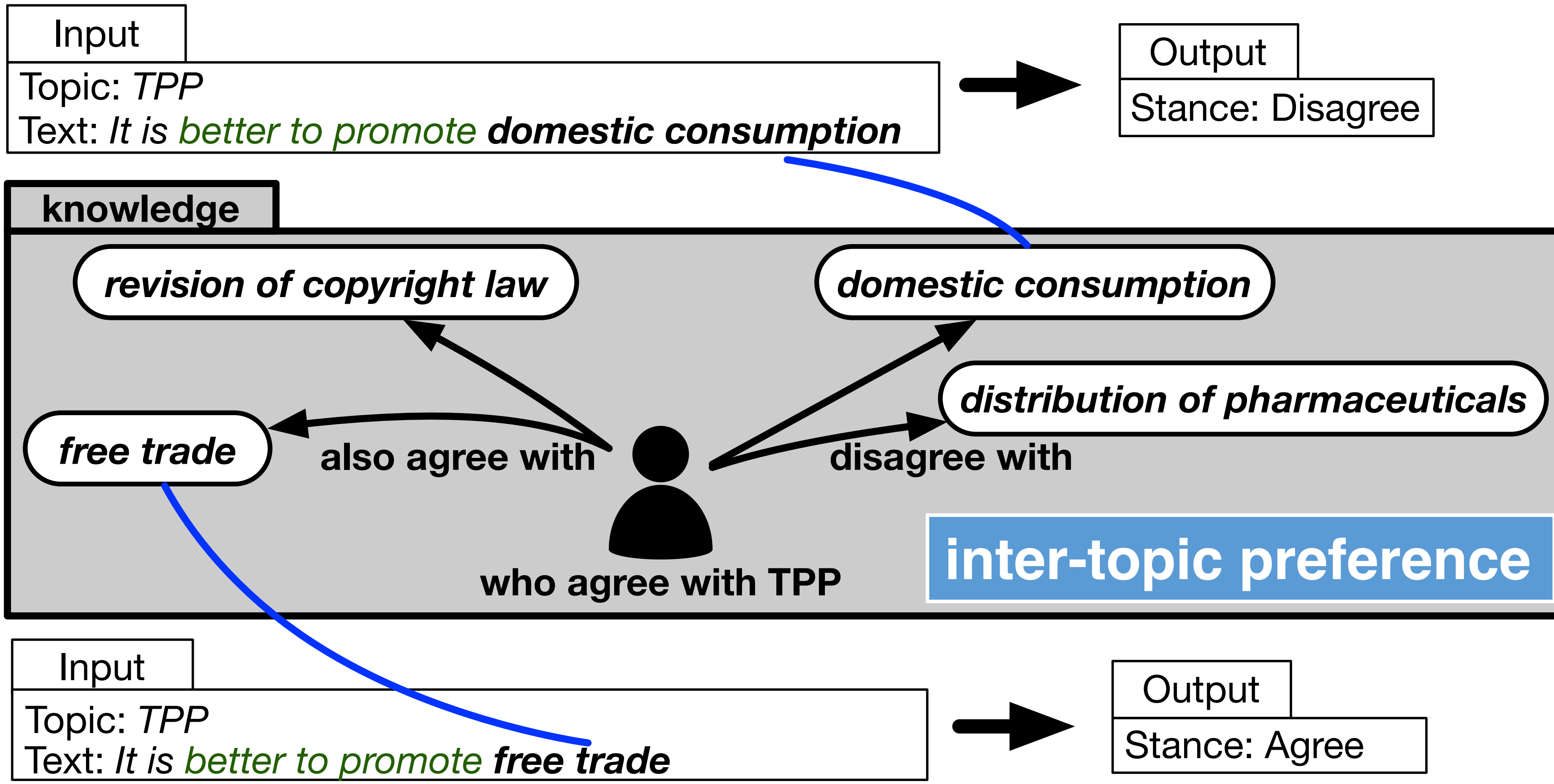
Difficulty of stance classification

Input		Output
Topic: *TPP* Text: *It is better to promote domestic consumption*	→	Stance: Disagree
Topic: *TPP* Text: *It is better to promote free trade*	→	Stance: Agree

People often talk about topics without explicitly mentioning the topic.

How can we classify stance from such a text?

Use of inter-topic preferences for stance classification



A relatively simple example

Input

Topic: *TPP*
Text: *I fully agree with* ***TPP***

→

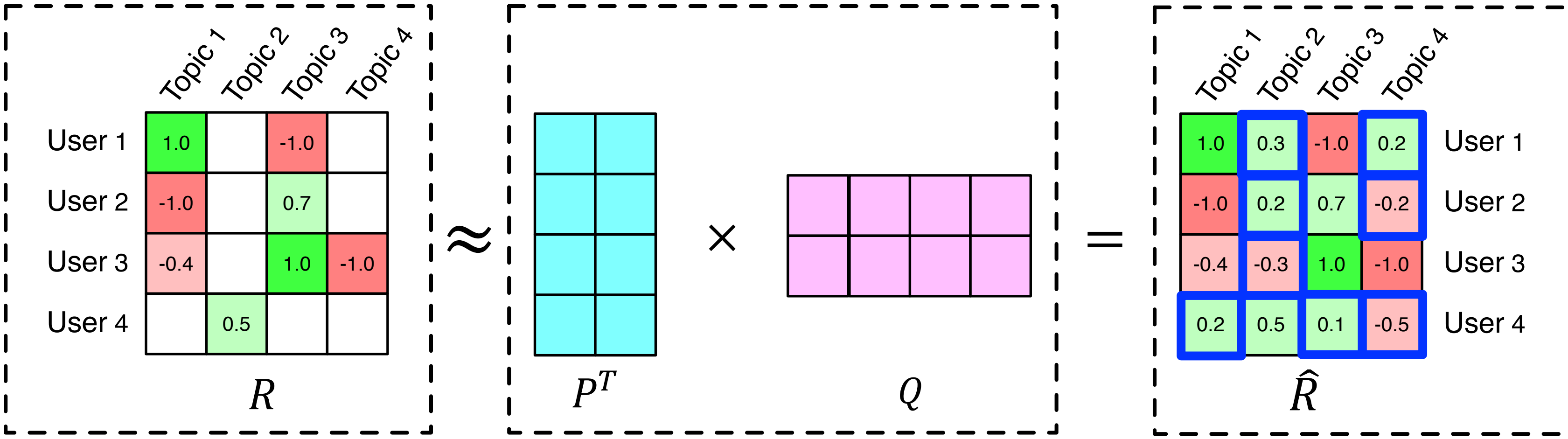
Output

Stance: Agree

Topic words and their surrounding words provide strong clues.
(Somasundaran&Wiebe, 2010), (Mohammad+, 2013)

※ Although datasets used in this work are in Japanese, we provide examples in English for readability.

Proposal: modeling inter-topic preferences via matrix factorization



Users' stances for each topics (**user-topic matrix**)

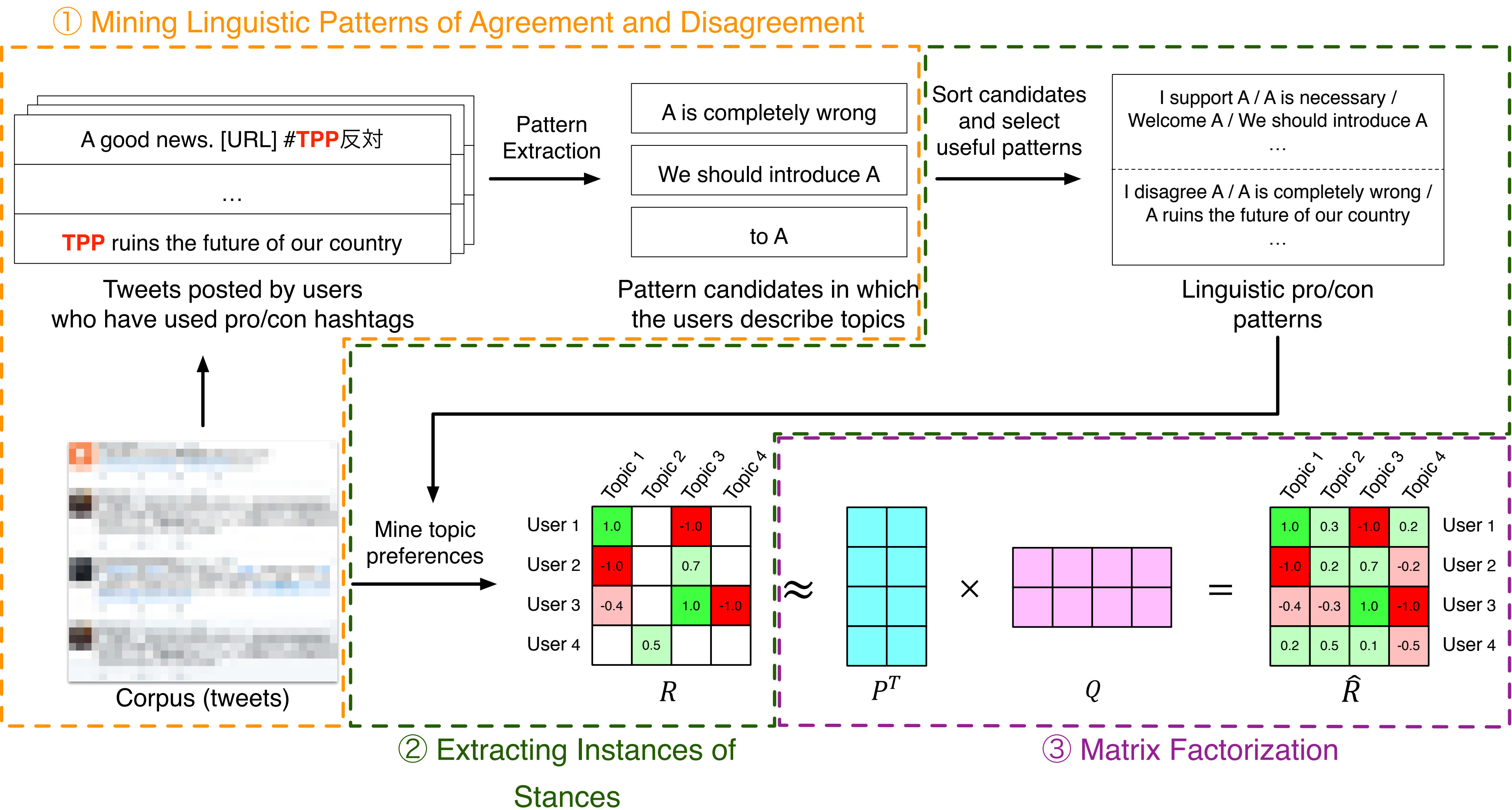
Compute users' dense feature vector and topics' dense feature vector via matrix factorization

Complete missing values by feature vectors

The aim of matrix factorization:

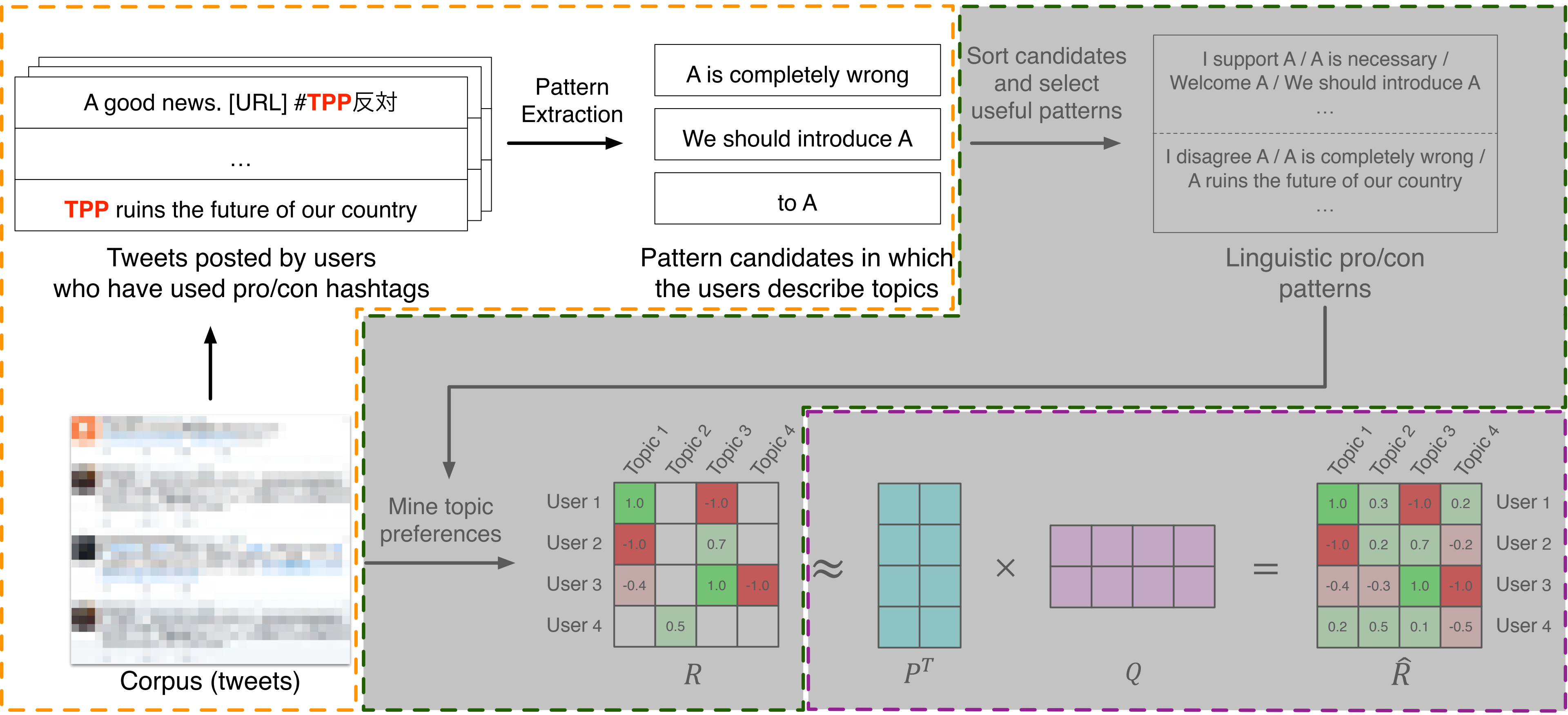
1. capture inter-topic preferences by dense feature vectors
2. reveal users' hidden stances by completion

The whole architecture



The whole architecture

① Mining Linguistic Patterns of Agreement and Disagreement

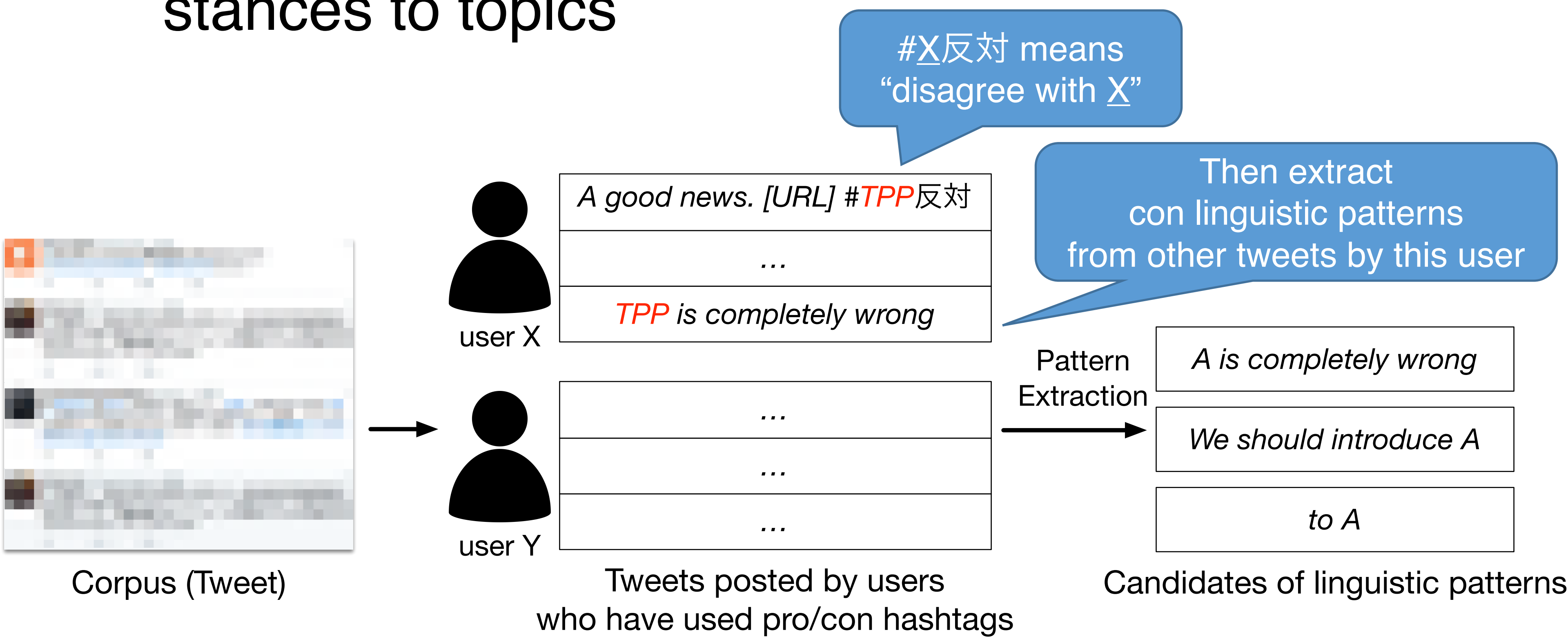


② Extracting Instances of Stances

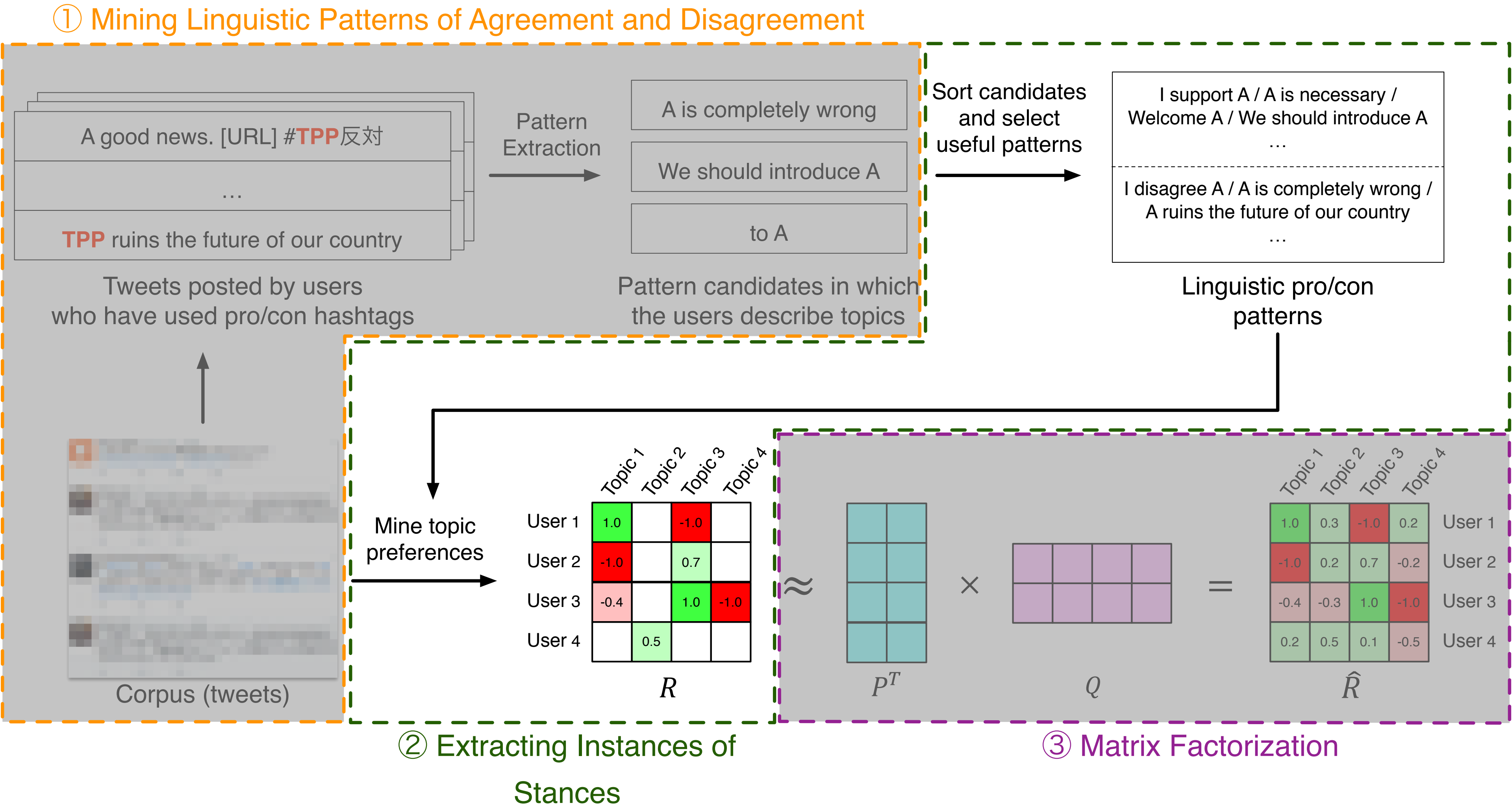
③ Matrix Factorization

Mining linguistic patterns of agreement/disagreement

- Focus on pro/con hashtags such as “#X賛成” or “#X反対” used by users who have strong stances to topics

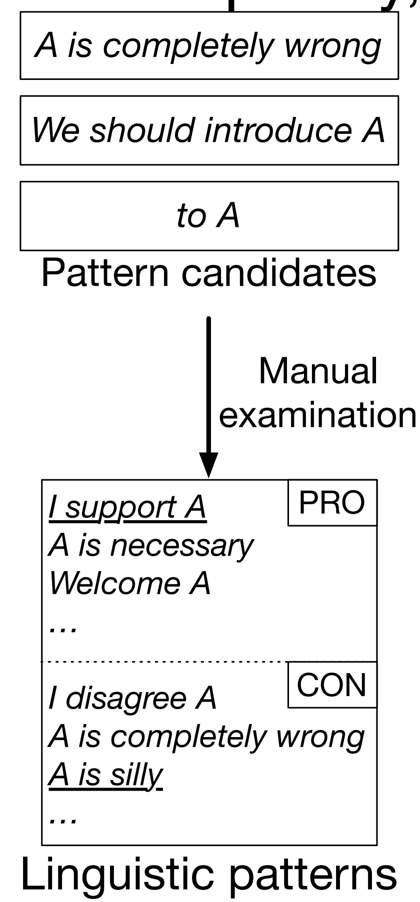


The whole architecture



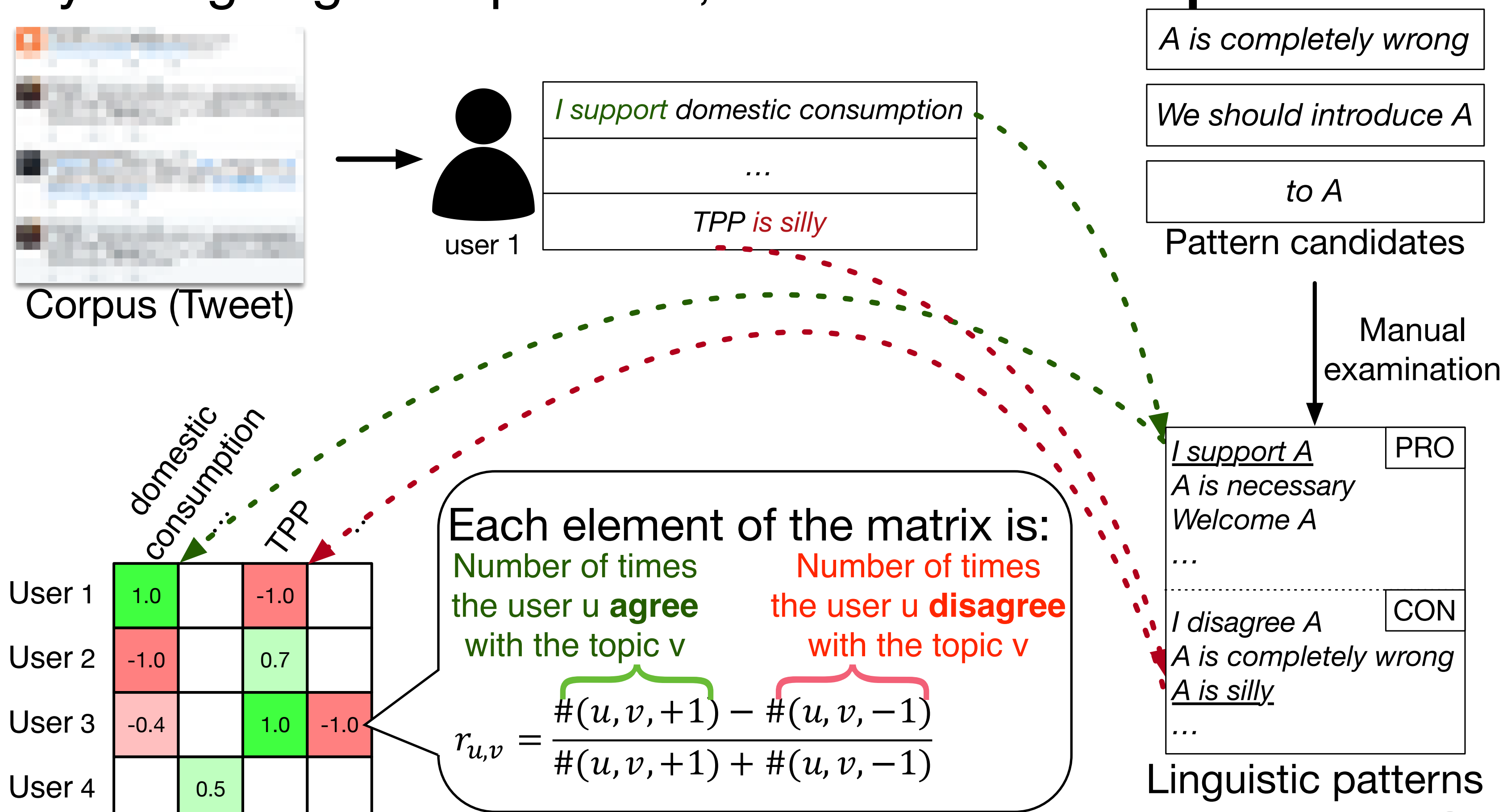
Extracting instances of stances

- Sort aforementioned pattern candidates by their frequency, and filter manually

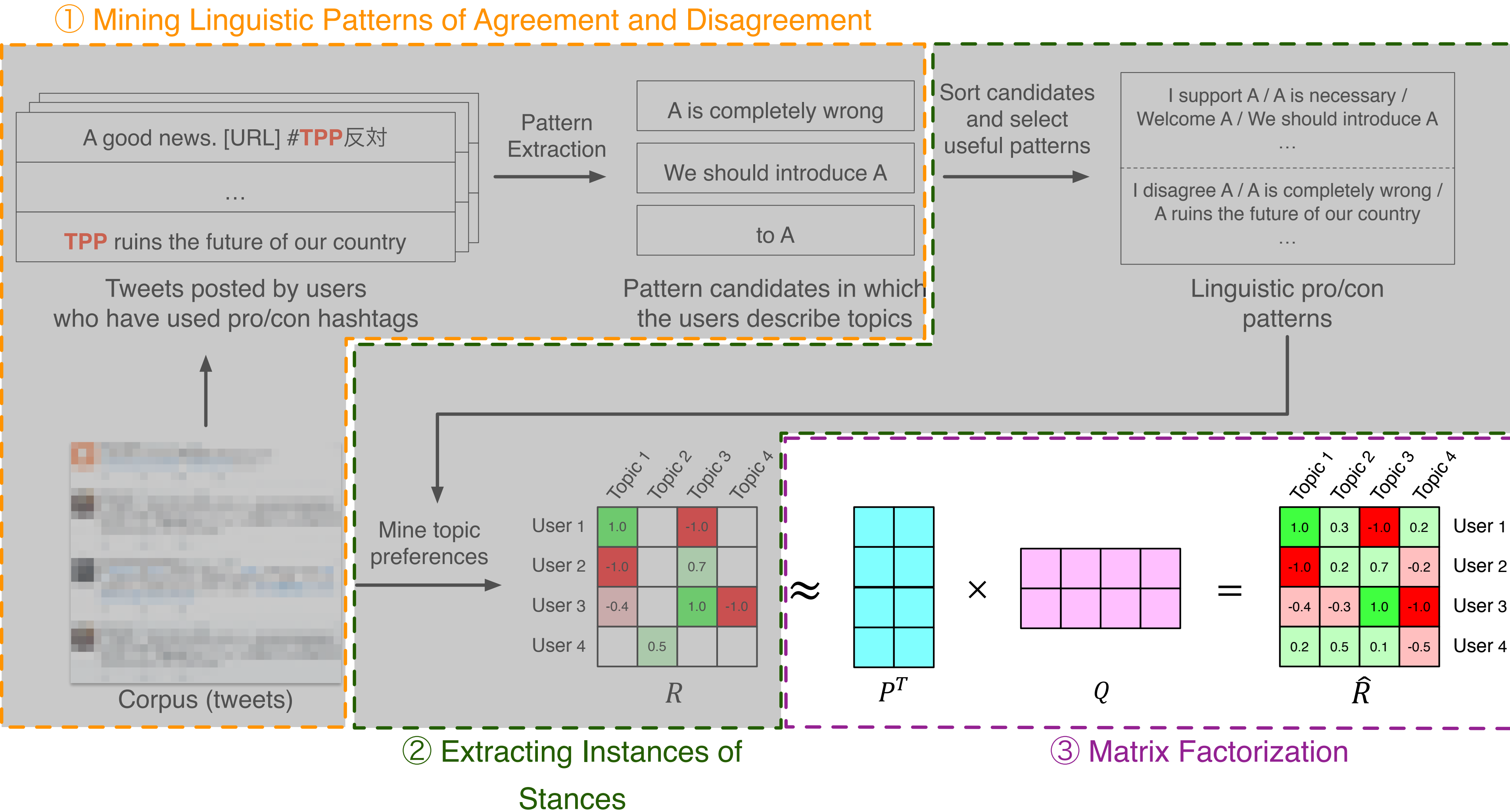


Extracting instances of stances

- By using linguistic patterns, we create **user-topic matrix**



The whole architecture

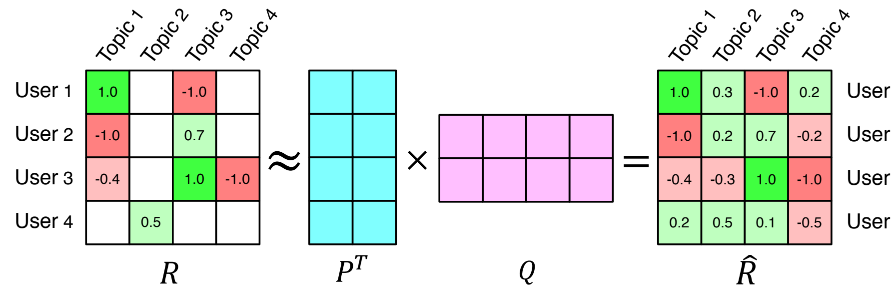


Matrix factorization

- By minimizing following objective function

$$\min_{P,Q} \sum_{(u,v)\in R} (r_{u,v} - \boldsymbol{p}_u{}^\intercal \boldsymbol{q}_v)^2 + \lambda_P ||\boldsymbol{p}_u||^2 + \lambda_Q ||\boldsymbol{q}_v||^2$$

$(u, v) \in R$: declared preference
$\boldsymbol{p}_u \in \mathbb{R}^k$: u column vectors of P (user vector)
$\boldsymbol{q}_v \in \mathbb{R}^k$: v column vectors of Q (topic vector)
$\lambda_P \geq 0, \lambda_Q \geq 0$: regularization coefficients



- We can complete missing values as follows:

$$\hat{r}_{u,v} \simeq \boldsymbol{p}_u{}^\intercal \boldsymbol{q}_v$$

- Based on preliminary experiments, we set parameters as $k = 100$, $\lambda_P = 0.1$, $\lambda_Q = 0.1$ (refer to the paper for more info)

- We use `libmf` to solve the optimization problem
https://github.com/cjlin1/libmf

Evaluation

- **Ex1: Determining the dimension parameter k**
  → RMSE decreased as the number of dimensions (k) increased

- **Ex2: Predicting missing stances**
  → 80-94% accuracy on predicting missing stances

- **Ex3: Correlation between human judgements**
  → Moderate correlation

Dataset

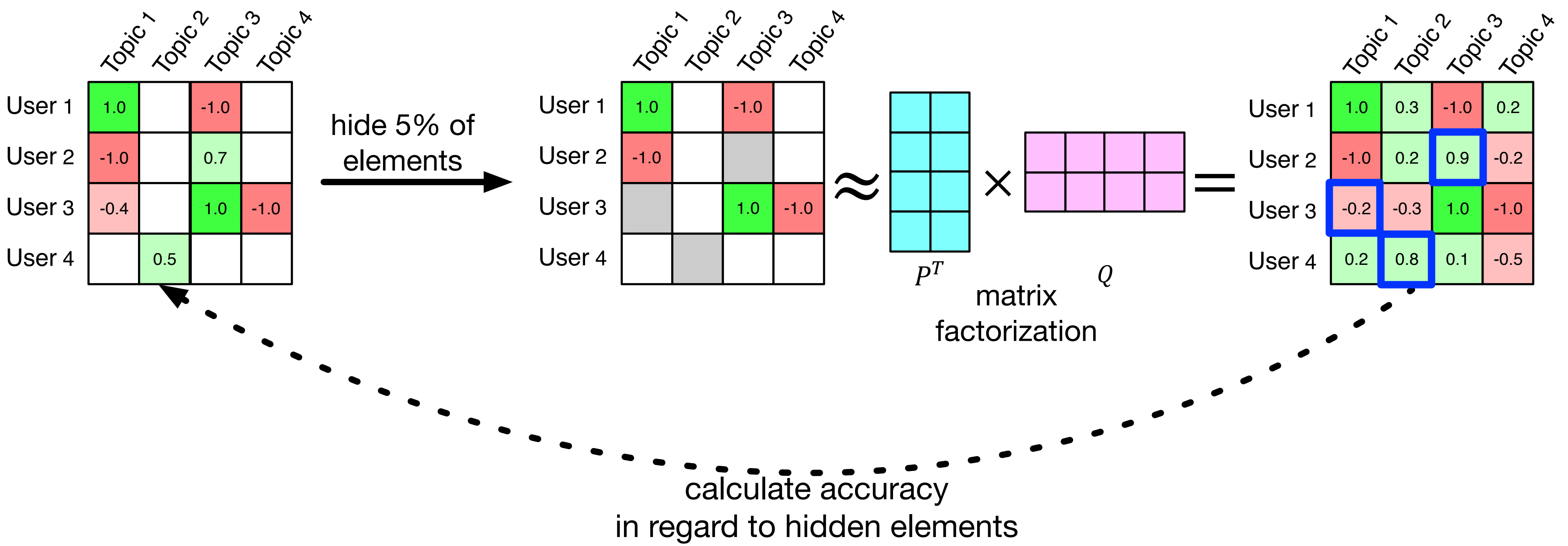
- Tweet corpus
  - about 35 Billion tweets crawled from Feb. 2013 to Sep. 2016
  - about 7 Million users
  - retweets are removed

- Collected data
  - 100 pro patterns and 100 con patterns (manually filtered)
  - about 25 Million tuples (agreement/disagreement declaration) corresponding to about 3 Million users and about 5,000 topics

- User-topic matrix
  - removed users and topics that appeared less than five times
  - about 10 Million tuples corresponding to about 270,000 users and about 2,300 topics
  - sparsity = 98.43%

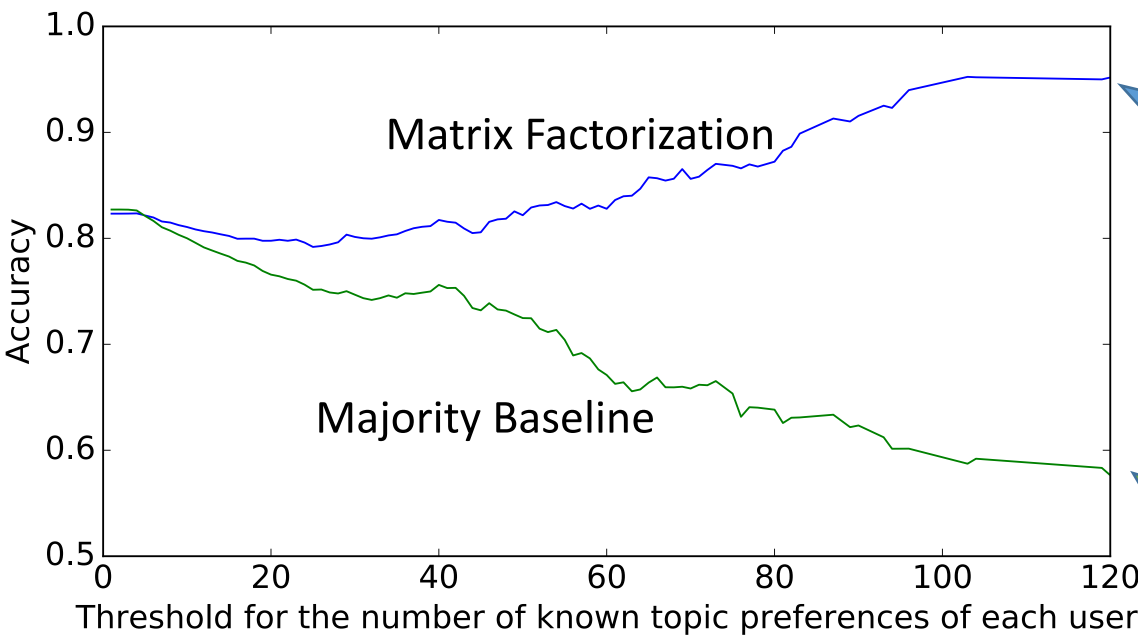
Ex2: Predicting missing stances

- **How accurately can user and topic vectors predict missing stances?**



Ex2: Predicting missing stances

- **How accurately can user and topic vectors predict missing stances?**
- majority baseline: predict missing values as majority one of agree/disagree in regard to the topic



Our approach predicts missing topic preferences by 80 – 94% accuracy

Since preferences of vocal users deviated from those of the average users, majority baseline decreased

Ex2: Predicting missing stances

- **Are predicted agreements/disagreements by matrix factorization are reasonable?**

user sample A

agree with		may also agree with
regime change capital relocation		vote of non-confidence to Cabinet same-sex partnership ordinance national people's government
disagree with	predict (matrix factorization) →	**may also disagree with**
Abe Cabinet Okinawa US military base nuclear weapons TPP		steamrollering war bill worsening dispatch law sendai nuclear power plant war bill

Our approach reasonably predicts missing values

Conclusion

- Modeled inter-topic preferences by matrix factorization

- Our approach accurately predicts missing stances by 80-94% accuracy

- Future work
  - Use methods of targeted sentiment analysis instead of using linguistic patterns

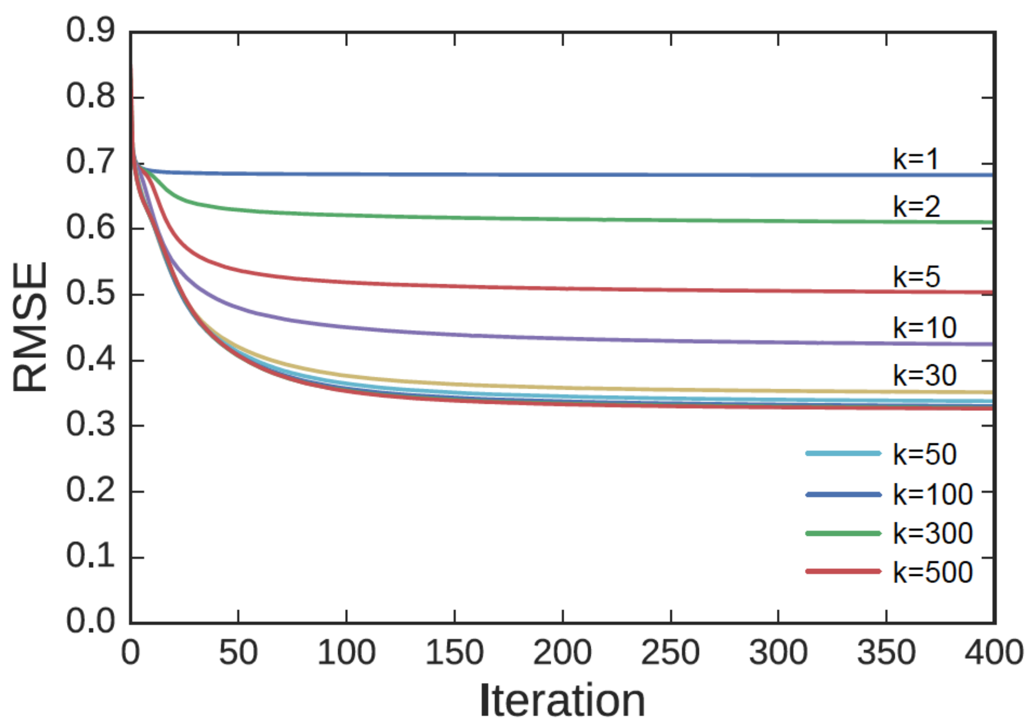
  - Extend our approach to other domains
    - product, company, music, etc

Appendix

Ex1: Determining the dimension parameter k

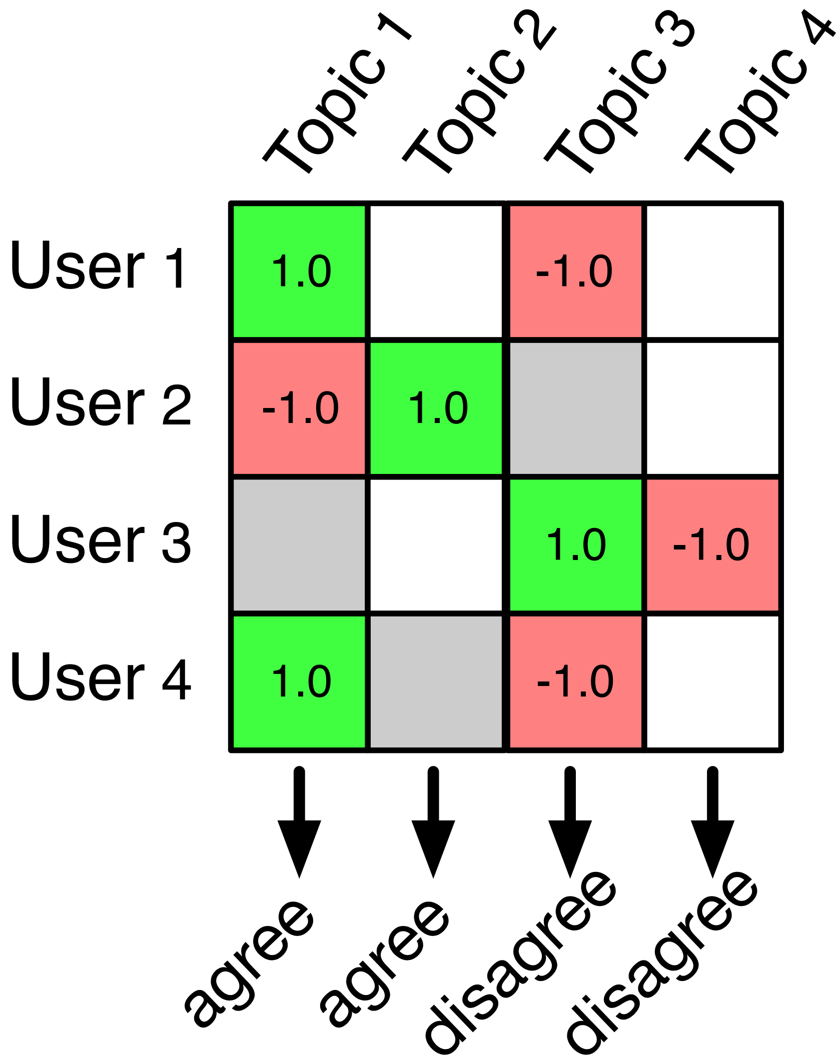
- We observed that the reconstruction error decreased as the iterative method of `libmf` progressed

- Based on this result, we concluded that $k = 100$ is sufficient for reconstructing the original matrix R



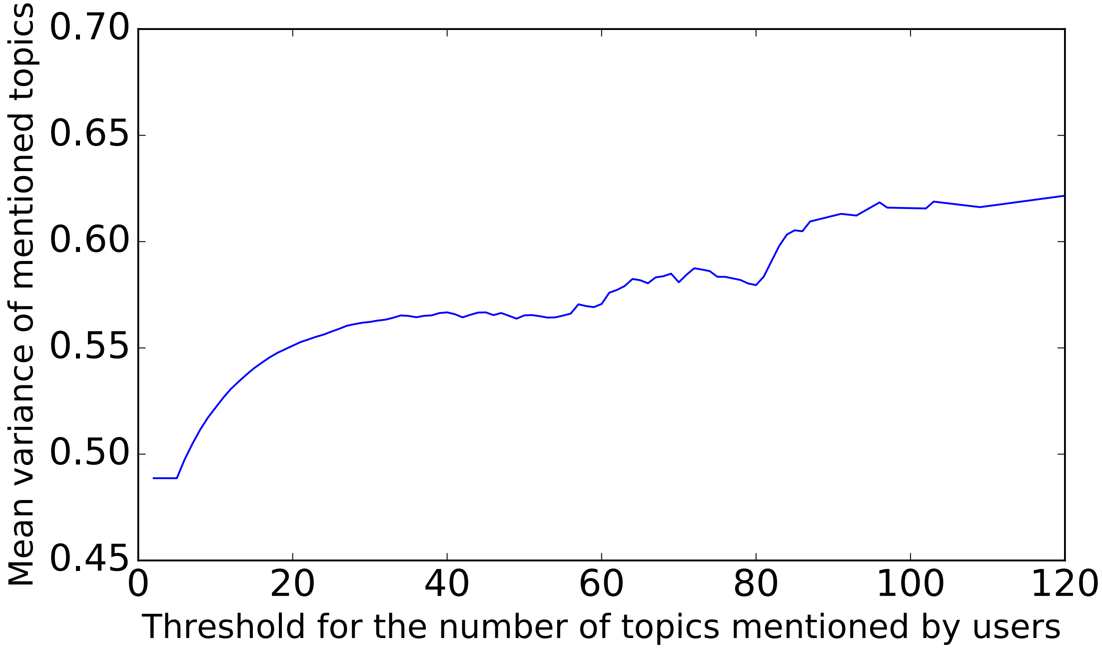
Ex2: Predicting missing stances

- majority baseline: predict missing values as majority one of agree/disagree in regard to the topic



Ex2: Predicting missing stances

- Since preferences of vocal users deviated from those of the average users, majority baseline decreased



Ex3: Correlation between human judgements

- Created a dataset of pairwise inter-topic preferences by using a crowdsourcing service

Q. People who agree with topic A also agree with topic B?

A1. those who agree/disagree with topic A may also agree/disagree with topic B → +1

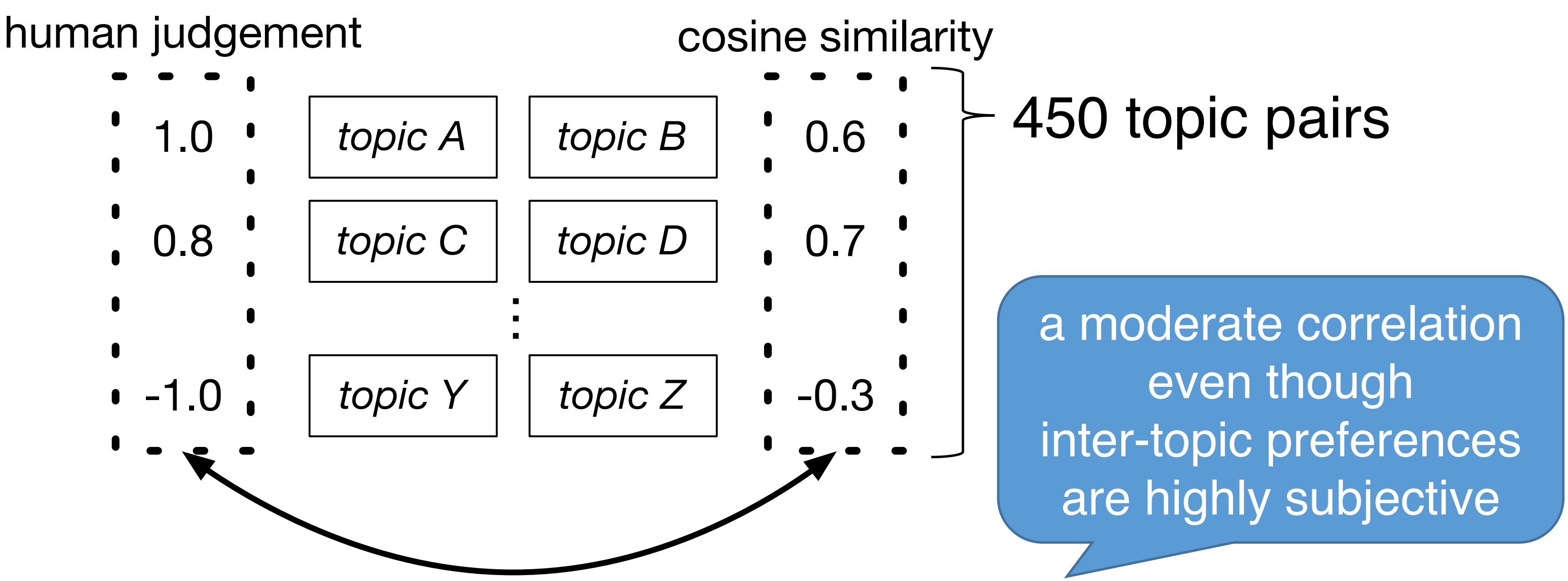
A2. those who agree/disagree with topic A may conversely disagree/agree with topic B → -1

A3. otherwise (no associaction between topic A and topic B → ±0

- Obtained 6-10 human judgements for every topic pair, then computed the mean of the points

Ex3: Correlation between human judgements

- Compared human judgements and similarity between vectors of pairs



Spearman’s rank correlation coefficient = 0.2210

Sub1: Example of predicted missing topic preference (qualitative)

User	Type	Topic
A	Agreement (declared)	regime change, capital relocation
	Disagreement (declared)	Okinawa US military base, nuclear weapons, TPP, Abe Cabinet, Abe government, nuclear cycle, right to collective defense, nuclear power plant, Abenomics
	Agreement (predicted)	same-sex partnership ordinance (0.9697), vote of non-confidence to Cabinet (0.9248), national people's government (0.9157), abolition of tax (0.8978)
	Disagreement (predicted)	steamrollering war bill (-1.0522), worsening dispatch law (-1.0301), Sendai nuclear power plant (-1.0269), war bill (-1.0190), constructing new base (-1.0186), Abe administration (-1.0173), landfill Henoko (-1.0158), unreasonable arrest (-1.0113)
B	Agreement (declared)	visit shrine, marriage
	Disagreement(declared)	tax increase, conscription, amend Article 9
	Agreement (predicted)	national people's government (0.8467), abolition of tax (0.8300), same-sex partnership ordinance (0.7700), security bills (0.6736)
	Disagreement (predicted)	corporate tax cuts (-1.0439), Liberal Democratic Party's draft constitution (-1.0396), radioactivity (-1.0276), rubble (-1.0159), nuclear cycle (-1.0143)

Table 1: Examples of agreement/disagreement topics predicted for two sample users A and B, with predicted score $\hat{r}_{u,v}$ shown in parenthesis.

Sub2: Similarity between topic vectors

- **Do the topic vectors obtained by matrix factorization capture inter-topic preferences?**

Topic: Liberal Democratic Party (LDP)

Top 7 of similar topics	cosisne similarity
Abs's LDP	0.3937
resuming nuclear power plant operations	0.3765
bus rapid transit (BRT)	0.3410
hate speech countermeasure law	0.3373
Henoko relocation	0.3353
C-130	0.3338
Abe administration	0.3248

Synonymous topics successfully have similar vectors

Sub2: Similarity between topic vectors

- **Do the topic vectors obtained by matrix factorization capture inter-topic preferences?**

Topic: Liberal Democratic Party (LDP)

Top 7 of similar topics	cosisne similarity
Abe's LDP	0.3937
resuming nuclear power plant operations	0.3765
bus rapid transit (BRT)	0.3410
hate speech countermeasure law	0.3373
Henoko relocation	0.3353
C-130	0.3338
Abe administration	0.3248

Topics promoted by LDP also have similar vectors

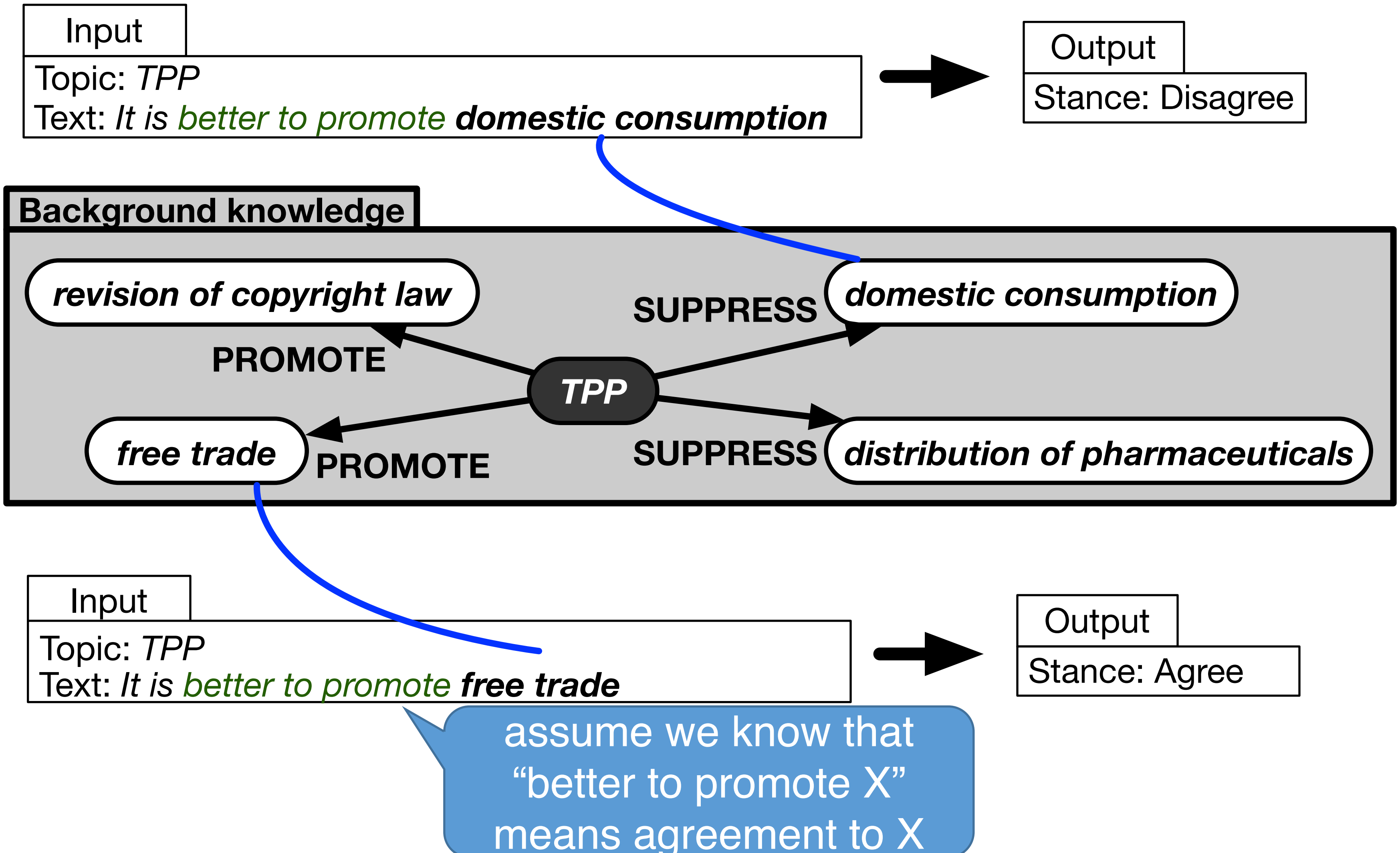
Sub2: Similarity between topic vectors

Topic	Topics with a high degree of cosine similarity
Liberal Democratic Party (LDP)	Abe's LDP (0.3937), resuming nuclear power plant operations (0.3765), bus rapid transit (BRT) (0.3410), hate speech countermeasure law (0.3373), Henoko relocation (0.3353), C-130 (0.3338), Abe administration (0.3248), LDP & Komeito (0.2898), Prime Minister Abe (0.2835)
constitutional amendment	amendment of Article 9 (0.4520), enforcement of specific secret protection law (0.4399), security related law (0.4242), specific confidentiality protection law (0.4022), security bill amendment (0.3977), defense forces (0.3962), my number law (0.3874), collective self-defense rights (0.3687), militarist revival (0.3567)
right of foreigners to vote	human rights law (0.5405), anti-discrimination law (0.5376), hate speech countermeasure law (0.5080), foreigner's life protection (0.4553), immigration refugee (0.4520), co-organized Olympics (0.4379)

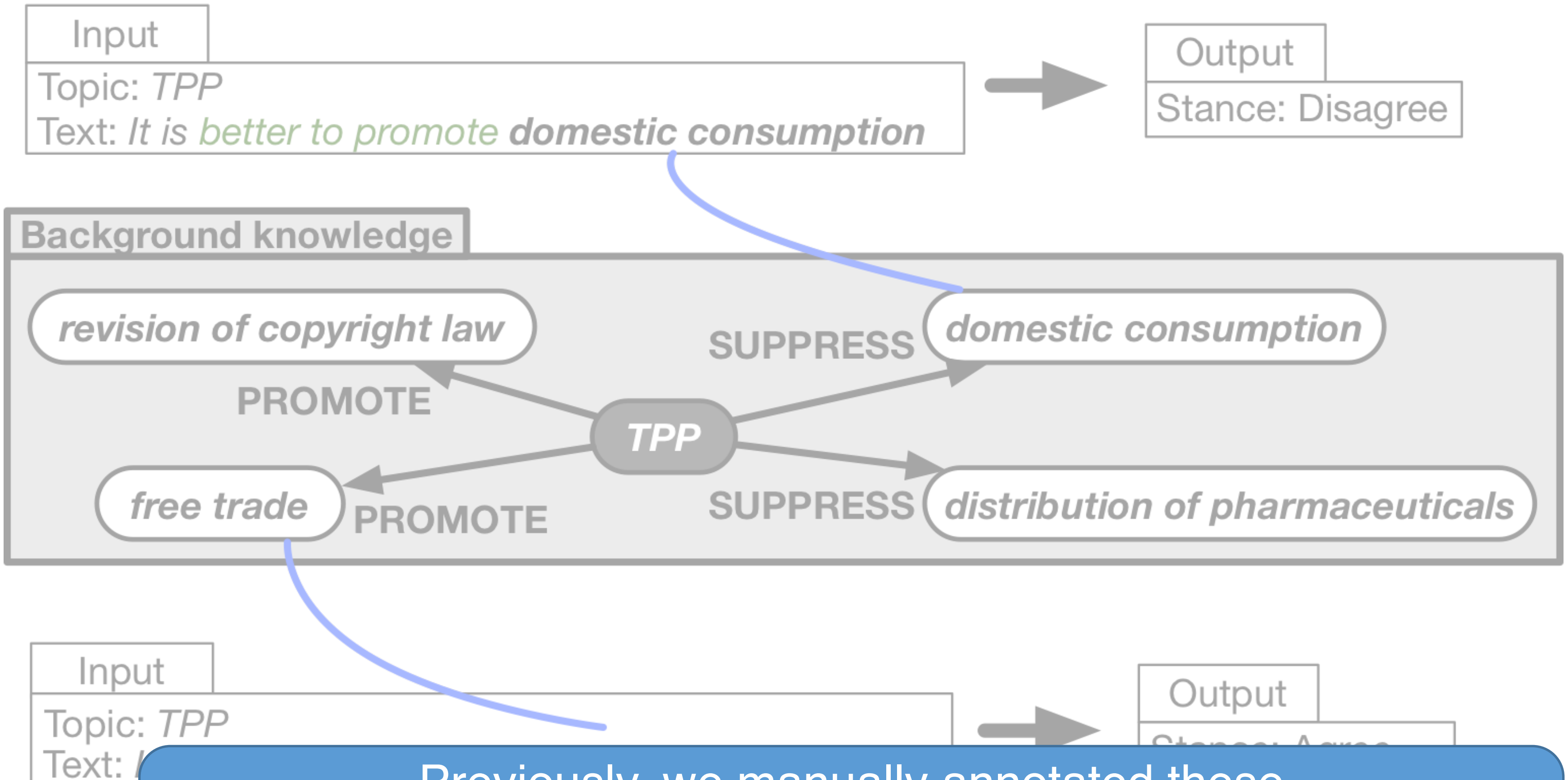
Table 2: Topics identified as being similar to the three controversial topics shown in the left column.

Unused slides

How can we use intrinsic knowledge in stance classification?



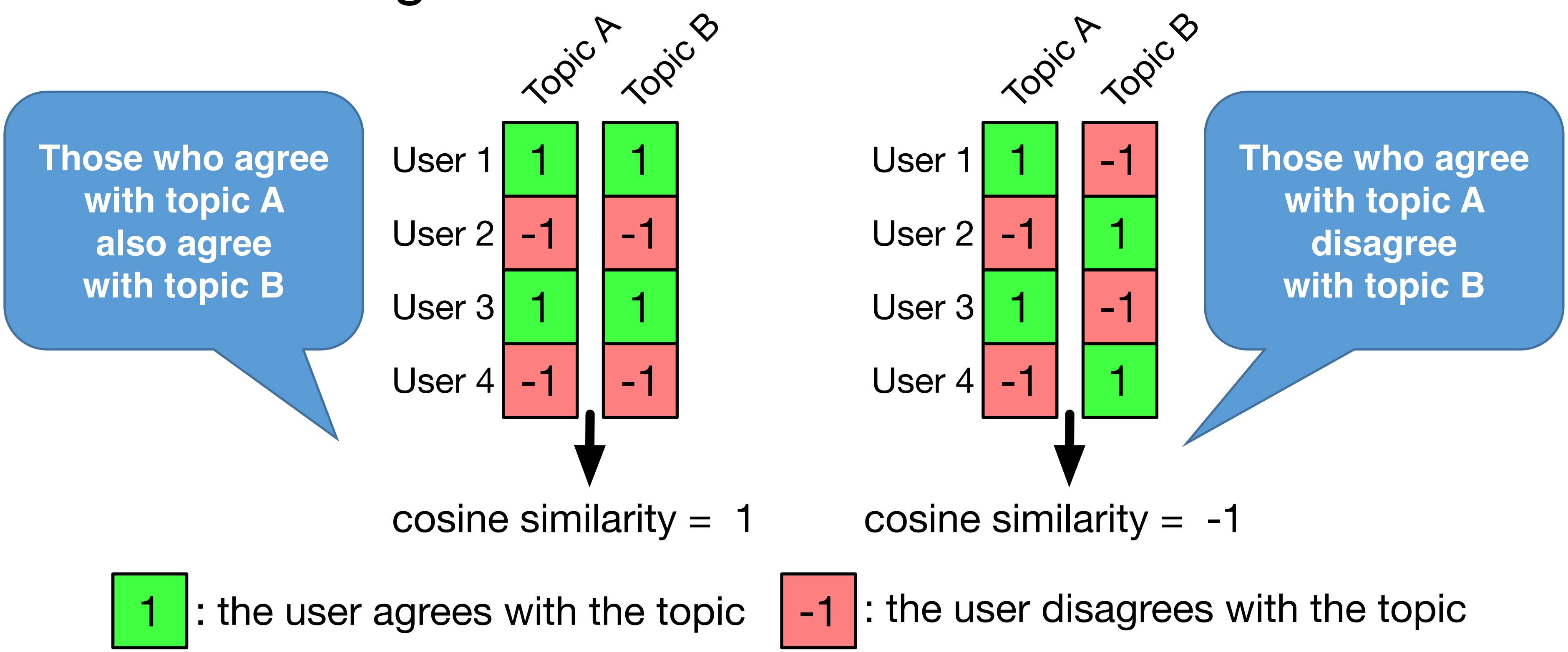
How can we use intrinsic knowledge in stance classification?



Previously, we manually annotated these PROMOTE/SUPPRESS knowledge and utilized in stance classification (Sasaki+, WI2016)

Challenge for modeling inter-topic preference

- Intuitively, we can see a topic as a vector consisting of users' declared stances



Challenge for modeling inter-topic preference

- However, a lot of people declare agreement/disagreement to only a few topics

	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		

Empty cell means undeclared stance

Other usage of inter-topic preference

- Public opinion survey
  - analyze people’s political ideology at low cost (cf. public opinion poll, census)
  - finer-grained than liberal/conservative
- Electoral campaigns
  - we can assume *“those who agree with topic A also vote for party B”*