

自然言語処理による 議論マイニング

東京工業大学

情報理工学院 情報工学系 知能情報コース

岡崎 直観 (okazaki@c.titech.ac.jp)

<http://www.chokkan.org/>

@chokkanorg

議論マイニング (argument mining)

(Gurevych+ 2016)

- **議論学**に基づきながら, **語用論**のレベルで与えられたテキストの**談話構造**を解析
 - **議論学**: 議論のモデル・理論 (例: トゥールミン)
 - **語用論**: 発話の意味を超え, 発話の機能も考慮
 - **談話構造**: 文の境界を越えた関係
- 関連技術
 - 感情分析: 対象物に対する話者の評価を推定
 - 談話構造解析: 文章の構造を推定
 - 含意関係認識: 文間の意味的な関係を推定

実例でみる議論マイニング

① セグメンテーション

- 文章から議論の「単位」を認定
- 議論単位の「役割」を推定
 - 役割の定義も色々ある

同時に解く手法／
別々に解く手法の
両方がある

大学の講義はすべて英語化すべきだ。優秀な学生が世界中から集まるし、日本人学生が世界で活躍する可能性が広がる。

文章1

大学の国際競争力が下がるだけだ。授業の質は低下するし、日本人学生の英語力はそこまで高くない。中高の教育も含めてトータルに見直すべきだ。

文章2

主張 (claim)

事実・前提 (premise)

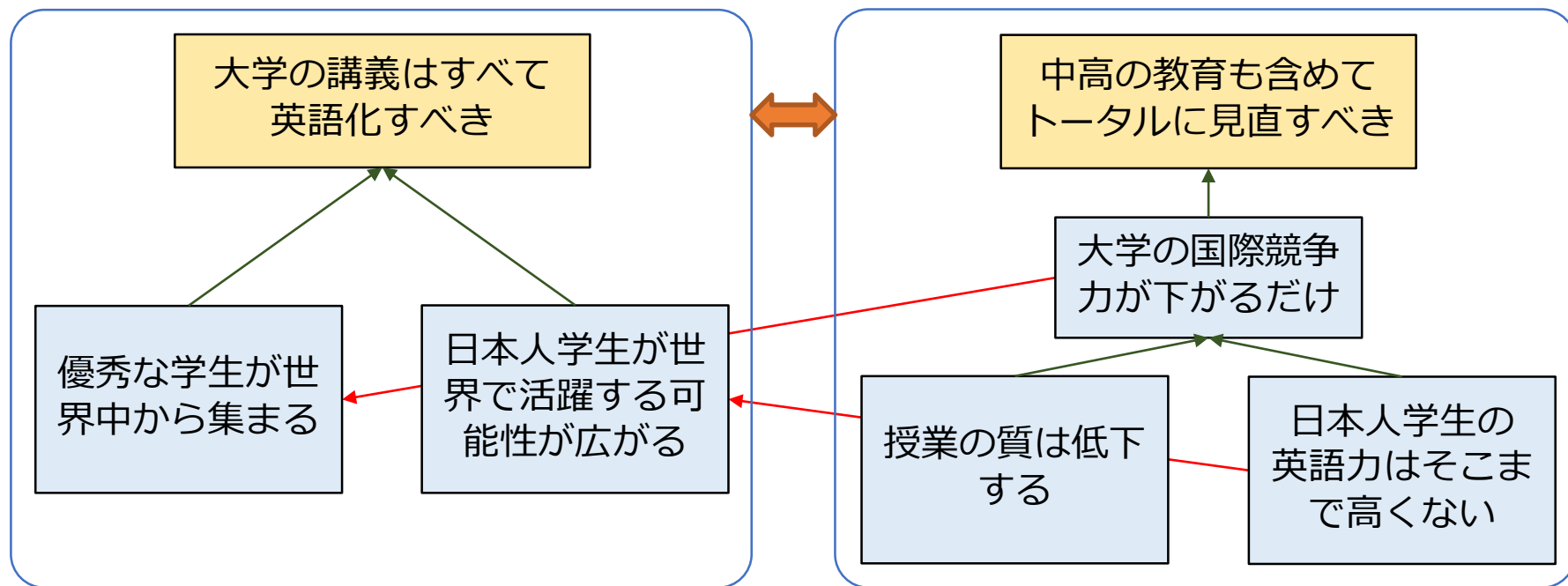
実例でみる議論マイニング

② 関係認識・構造化

• 議論単位の間関係を推定

- **Support**: 根拠や支持
- **Attack**: 反論や批判

この関係の定義も色々ある
先に関係の有無を当ててから、関係の
種類を予測する方法もある



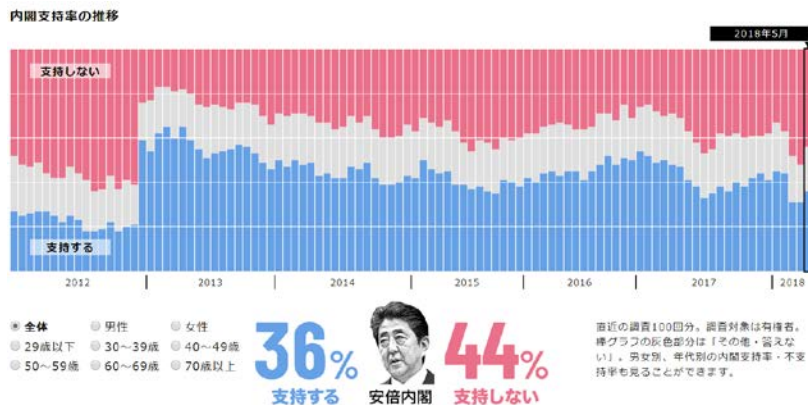
議論マイニングの応用 (※未実現の応用も含む)



ディベート・裁判の自動化・支援
<https://www.youtube.com/watch?v=7q59PJxbGhY>



見聞の広い市民の養成 (※)
<https://www.procon.org/>



世論調査 (※)

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

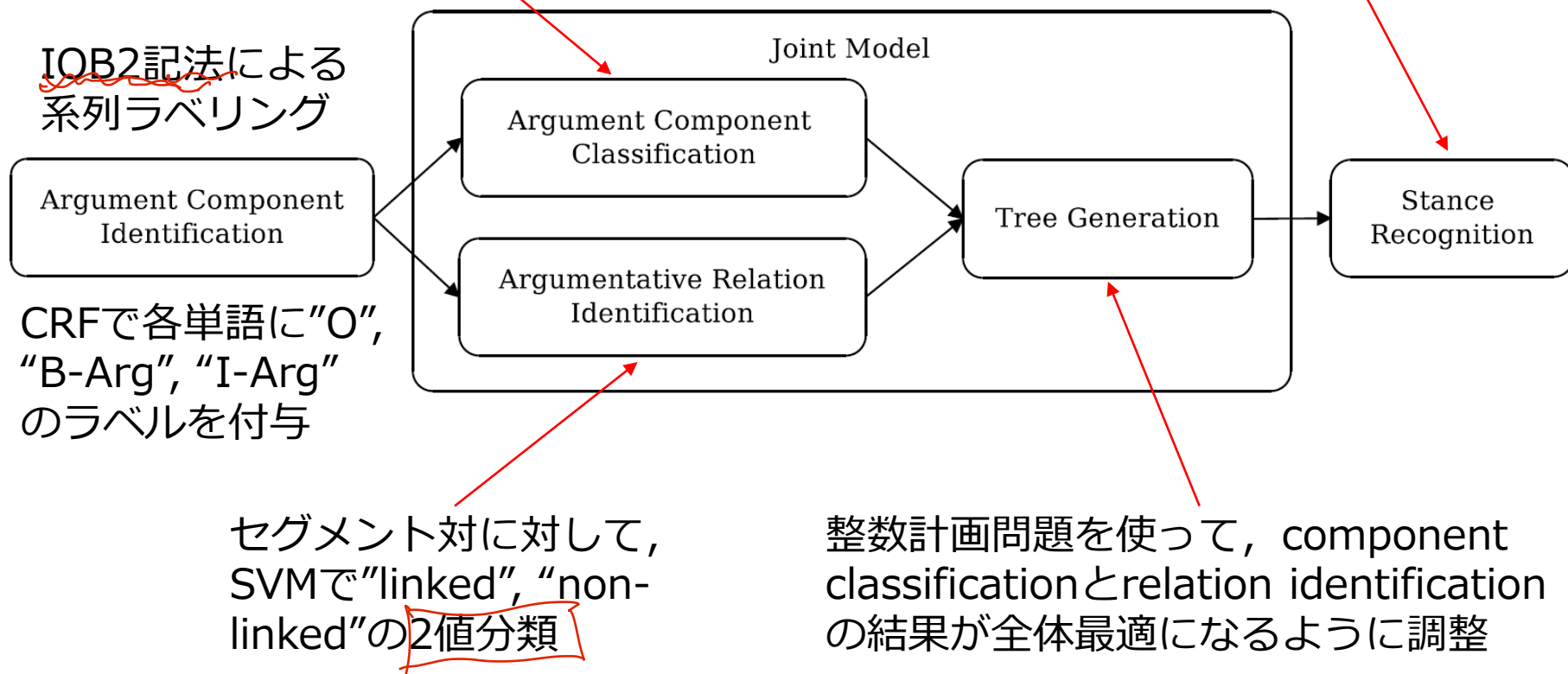


フェイクニュースの検出
<http://www.fakenewschallenge.org/>

典型的な手法 (Stab+ 2017)

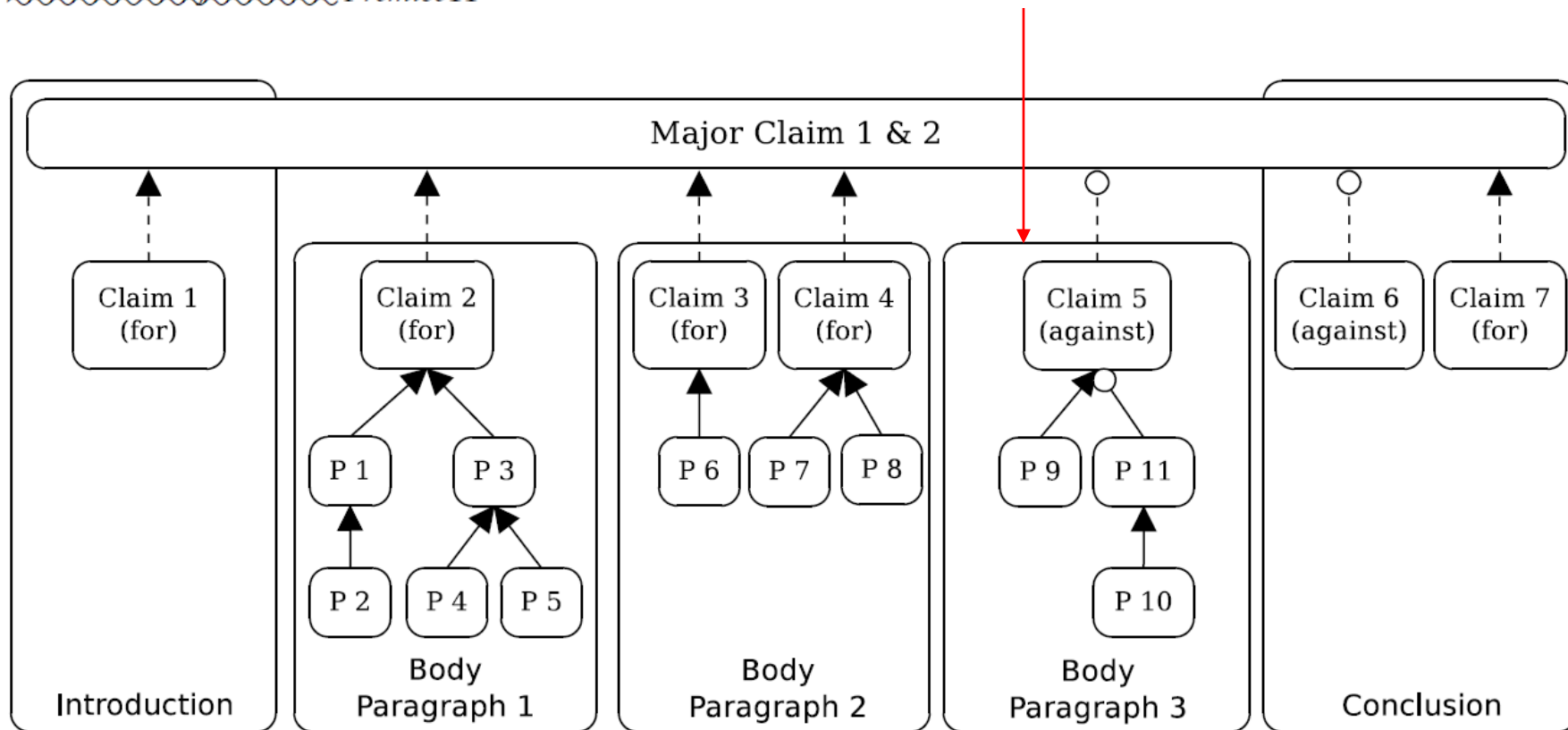
セグメントから特徴量を取り出し、SVMで“major claim”, “claim”, “premise”の3値分類

“linked”と判定されたセグメント対の関係をSVMで“support”と“attack”に2値分類



入力と出力の例 (Stab+ 2017)

Admittedly, [cloning could be misused for military purposes]_{Claim5}. For example, [it could be used to manipulate human genes in order to create obedient soldiers with extraordinary abilities]_{Premise9}. However, because [moral and ethical values are internationally shared]_{Premise10}, [it is very unlikely that cloning will be misused for militant objectives]_{Premise11}.



論説文章に対する性能 (Stab+ 2017)

Table 8

Model assessment of argument component identification (\dagger = significant improvement over baseline heuristic).

	F1	P	R	F1 Arg-B	F1 Arg-I	F1 O
Human upper bound	0.886	0.887	0.885	0.821	0.941	0.892
Baseline majority	0.259	0.212	0.333	0	0.778	0
Baseline heuristic	0.642	0.664	0.621	0.364	0.867	0.677
CRF all features	\dagger 0.867	\dagger 0.873	\dagger 0.861	\dagger 0.809	\dagger 0.934	\dagger 0.857

Table 12

Model assessment on persuasive essays (\dagger = significant improvement over baseline heuristic; \ddagger = significant improvement over base classifier).

	Components				Relations			Stance recognition			Avg F1
	F1	F1 MC	F1 CI	F1 Pr	F1	F1 NoLi	F1 Li	F1	F1 Sup	F1 Att	
Human upper bound	0.868	0.926	0.754	0.924	0.854	0.954	0.755	0.844	0.975	0.703	0.855
Baseline majority	0.260	0	0	0.780	0.455	0.910	0	0.478	0.957	0	0.398
Baseline heuristic	0.759	0.759	0.620	0.899	0.700	0.901	0.499	0.562	0.776	0.201	0.674
Base classifier	0.794	\dagger 0.891	0.611	0.879	0.717	0.917	0.508	\dagger 0.680	\dagger 0.947	\dagger 0.413	0.730
ILP joint model	\dagger \ddagger 0.826	\dagger 0.891	\ddagger 0.682	\ddagger 0.903	\dagger 0.751	\dagger 0.918	\ddagger 0.585	\dagger 0.680	\dagger 0.947	\dagger 0.413	0.752

単一文書での解析性能は比較的高いが、意思決定・合意形成の目的を考えると、著者・話者・動機の異なる複数の文書の解析が必要

議論マイニングのボトルネックは世界知識

(Saint-Dizier 2016, Hanawa+ 2017, Moens 2018)

- 世界知識が必要となる例 (cellphoneの知識が要る)

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)



2番目の文は1番目の
文の論拠

- 議論マイニングにおいて **世界知識は必須**
 - 議論単位間の関係予測において, 78%の議論で世界知識が必要 (Saint-Dizier 2016)
 - 賛否分類タスクにおいて, 33.9%の発言に対して世界知識が必要 (Hanawa+ 2017)

議論マイニングに世界知識を導入する試み

- 世界知識に基づく議論マイニング
 - 主張間のギャップ (Boltuzic+ 2016)
 - Argument reasoning comprehension (Habernal+ 2018)
 - 注意機構に基づくスタンス分類 (Hanawa+ 2018)
- 世界知識の自動獲得
 - Wikipediaからの知識獲得 (Hanawa+ 2017)
 - Twitterの投稿からの知識獲得 (Sasaki+ 2017; Sasaki+ 2018)
- 仮説推論

主張間の暗黙のギャップ (Boltuzic+ 2016)

implicit 中々 見えない 関係 ...

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.

ギャップを埋める暗黙の前提

Legalized marijuana can be controlled and regulated by the government.

上の主張と同じ立場であることは人間には分かるが、コンピュータには難しい

主張間のギャップを埋めてみる (Boltuzic+ 2016)

より explicit な形へ

- 主張間のギャップを埋めるデータセットの構築
 - 既存のデータセット (Hasan+ 2014) を利用
 - 500件の主張対に対して, 3人の人間が前提を補完

• 得られた知見

- ギャップの数と主張間の類似度は負の相関がある
- 補完された前提を利用することで, 主張の自動マッチングの精度が向上

● 人間が補完する前提はバラバラ (右図参照)

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 WWII.*

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.*

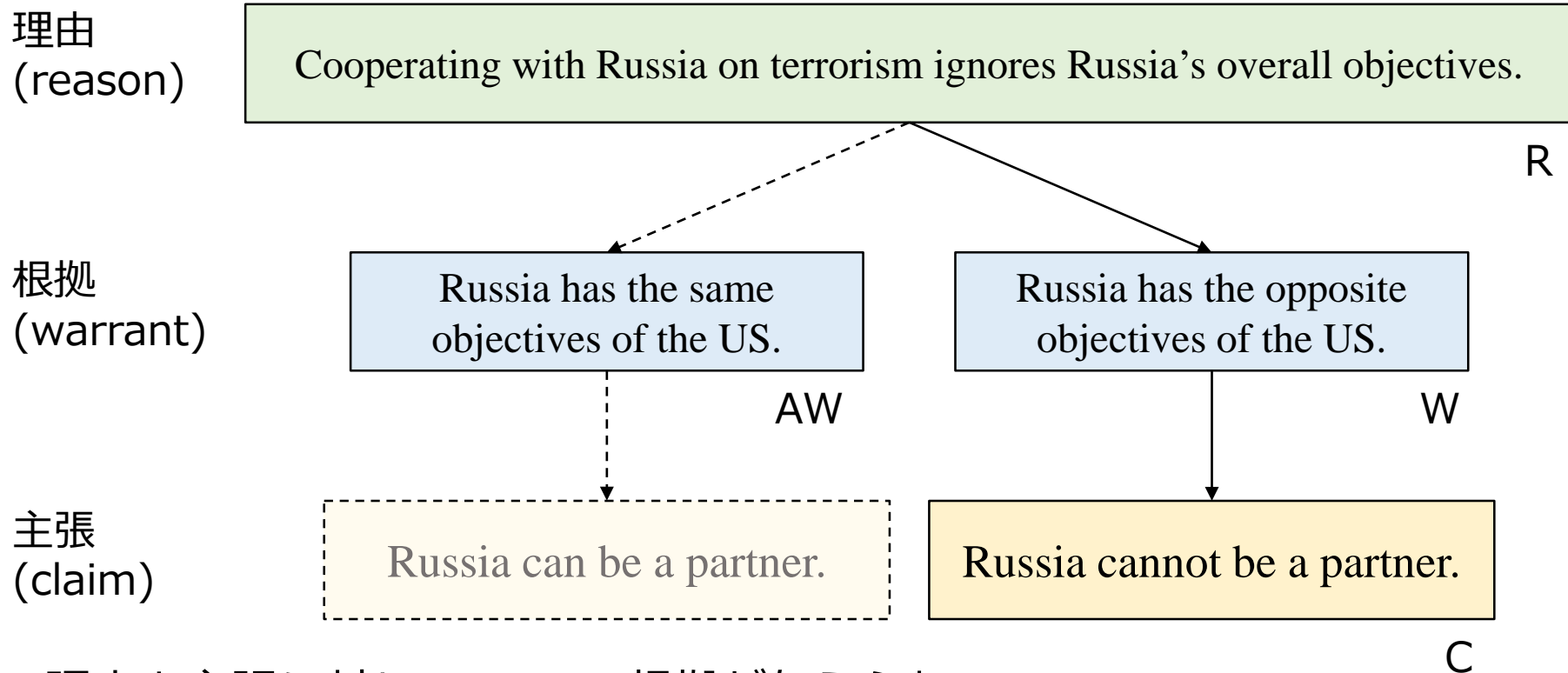
A1 Premise 5: *Aggression is a crime.*

A1 Premise 6: *"WWII" stands for the Third World War.*

A3 Premise 1: *Marijuana leads to irrational paranoia which can lead to committing a crime.*

Argument reasoning comprehension

(Habernal+ 2018) (ギャップを埋めるタスクを選択式とした)



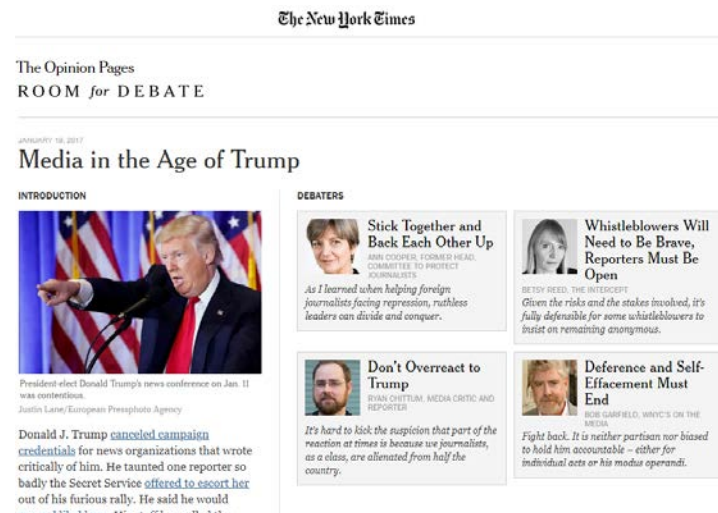
- 理由と主張に対して, 2つの根拠が与えられる
- どちらの根拠が理由と主張を繋ぐのにふさわしいか選ぶ
- もう一方の根拠は, 理由から「反対の主張」を導くとして用意された
 - この研究では反対の根拠 (AW: alternative warrant) と呼んでいる

Argument reasoning comprehension

(Habernal+ 2018)

- データセットの作成

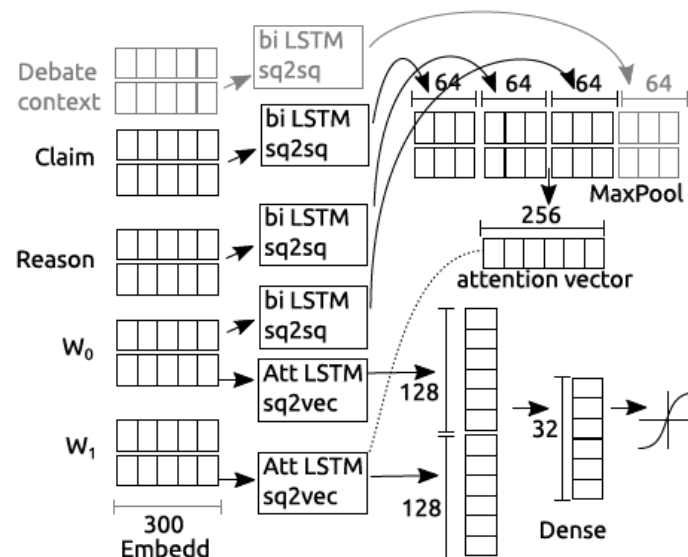
- New York TimesのRoom for Debateを題材として用いる (右図)
- クラウドソーシングで8個のマイクロタスクを実行し、データセットを構築
- 1,970件の(C, R, W, AW)を収録
- 人間の正解率 (AWではなくWを正しく選ぶ率) は79.8% (平均的な作業者), および90.9% (訓練した作業者)



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

- WとAWの自動識別実験

- アテンション付きのエンコーダ・デコーダモデルによる分類器 (右図)
- 正解率は56.0% (人間よりかなり低い)
- (私見) 多様なトピックを右図のようなモデルに与えると、過学習しやすい (キーワードを覚えこもうとする)



我々の取り組み

共同研究者



佐々木 彬
リクルートテクノロジーズ
(元・東北大学)






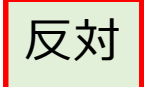
埴 一晃
東北大学



乾 健太郎
東北大学
理研AIP

タスク: 賛否分類 (Mohammad+ 2016)

トピック: 憲法96条の改正

- 憲法96条の改正は当然の流れだ  
- 絶対に憲法96条の改正は阻止する  
- トピックに対するテキストの書き手の賛否を推定
 - 賛成 (support, agree, for, pro)
 - 反対 (attack, disagree, against, con)
 - 中立 (neutral, none, other)
- 議論の的であるトピックに関して, ツイートなどの普通の投稿から世の中の意見を探ることができる

賛否分析でも世界知識が必要

トピック: 憲法96条の改正

全トピックに関して賛否の学習
データを作るのは不可能

総議員の2/3以上の賛成で発議 というのはおかしい

改正前

単純過半数で改正できないのは重要なこと

批判

同意

議員の過半数で発動 というのは危うすぎる

批判

改正後

賛成

反対

反対

予測モデル



トピック
固有の知識



トピック
非依存の知識

テーマ固有の知識を新聞
記事やウェブ等から自動
獲得し、議論分析に活用

憲法改正などの国民全体の議論を分析・支援できる技術確立したい

賛否分類コーパスの構築

(Hanawa+ 2018, submitted)

- 7トピックに言及しているツイートを収集し、トピックに賛成／反対／中立なのかを人手で付与した
- 40.2%の意見表明の賛否を推定するには、トピックに関する知識が必要であることが分かった



「消費促進」なんて目的決められてしまっっては、それはもう「休み」ではないよ。

プレミアムフライデーに否定的な発言例

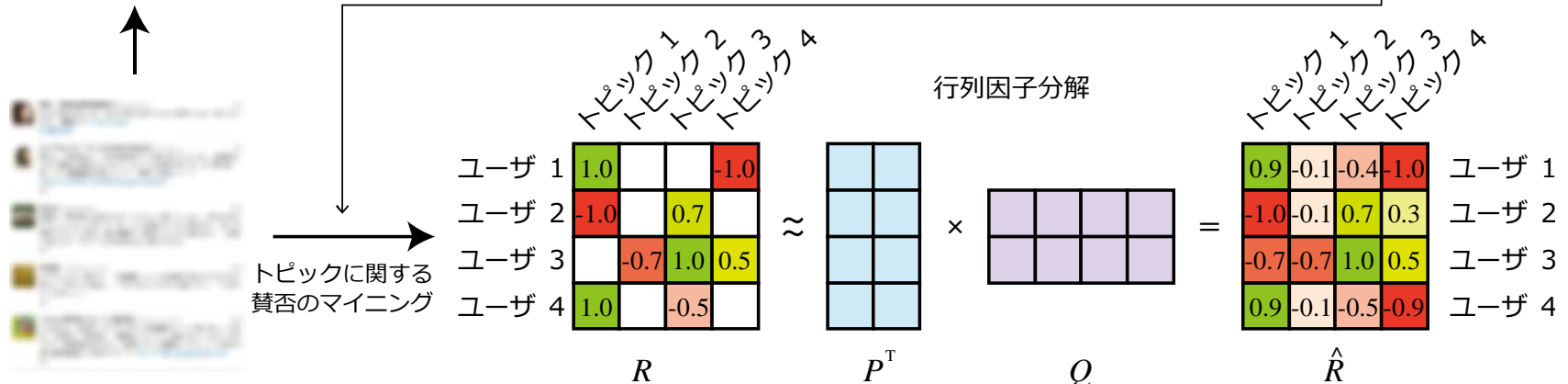
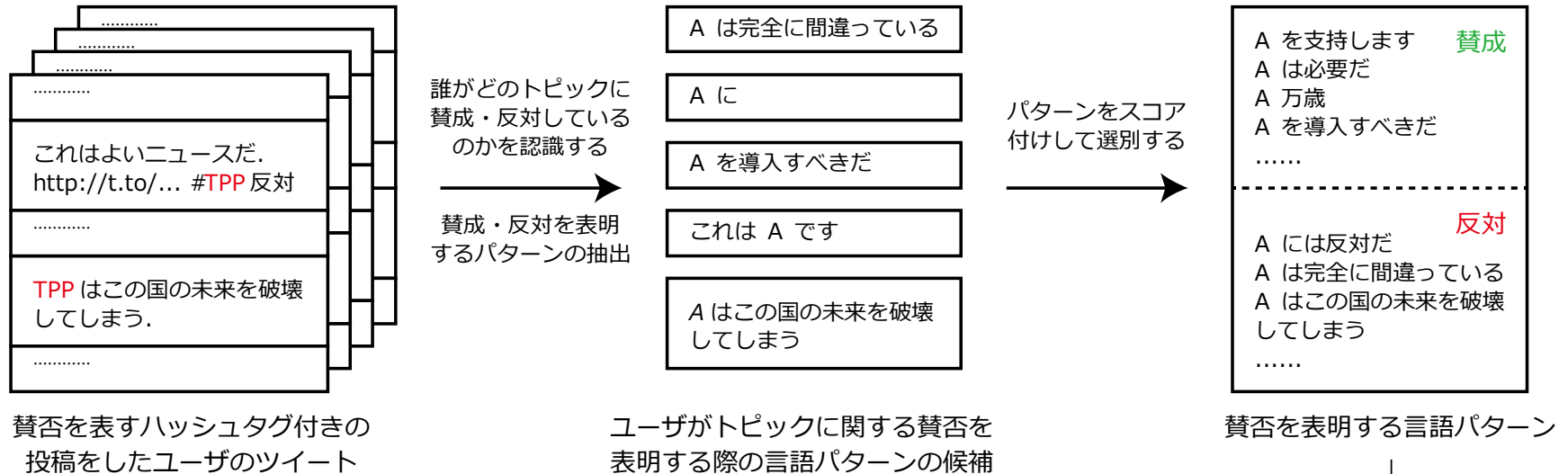
トピック	賛成	反対	中立
大阪都構想	239	259	380
安保法案	168	352	262
プレミアムフライデー	153	744	218
TPP	53	802	230
原発	47	783	202
集団的自衛権	160	468	196
共謀罪	86	592	308
合計	906	4000	1795

賛否分類において必要な知識	%	意見の例
一般的な賛否表明	56.3	原発は絶対必要（原発: 賛成）
トピックの因果関係 (∈ Wikipedia)	26.3	関税が機能するべき（TPP: 反対）
トピックの因果関係 (∉ Wikipedia)	13.9	遺伝子組み換え食品とか心配（TPP: 反対）
その他の知識	2.5	治安維持法を復活させたいのか（安保法案: 反対）

Twitterからのトピック間選好知識の獲得

(Sasaki+ 2017)

- 個人の意見表明から政治的スペクトル（イデオロギー）を抽出・分析
- 商品推薦の手法に着想を得て「〇〇に賛成する人は××にも賛成する」を予測



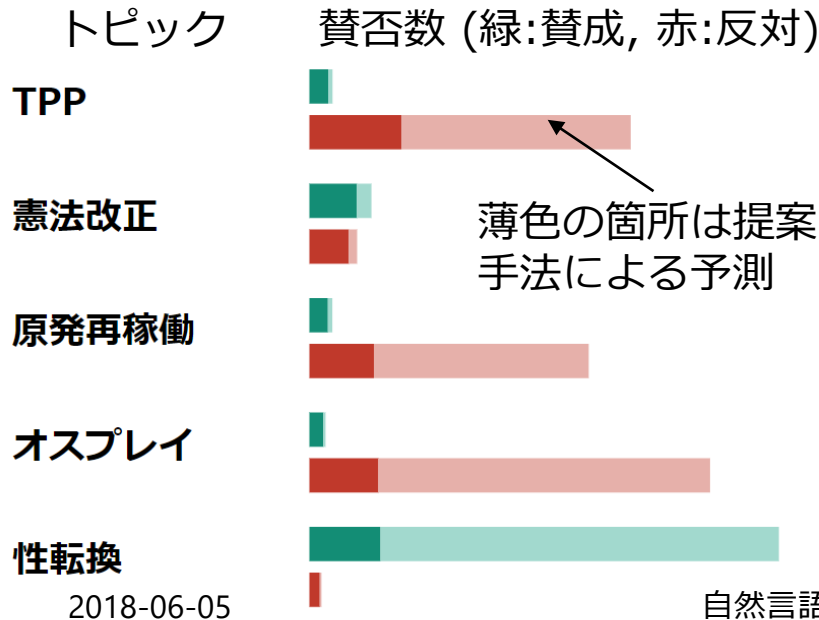
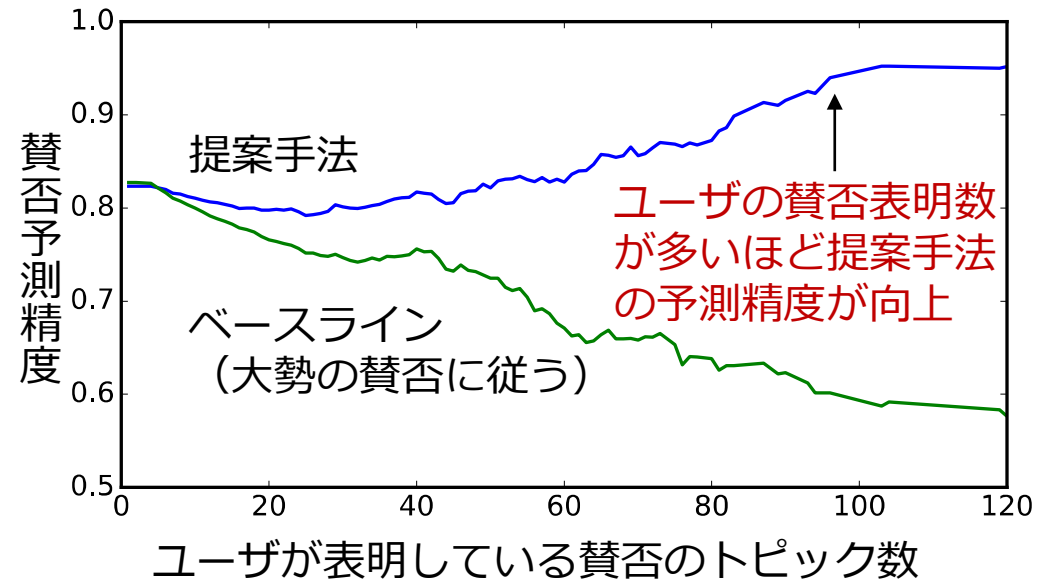
コーパス（ツイート）

ユーザ・トピック行列 ユーザベクトル トピックベクトル 低ランク近似（欠損値補完）

Twitterからのトピック間選好知識の獲得

(Sasaki+ 2017)

- ユーザが明示的に表明していない賛否を82-95%の精度で予測
- 学習結果の副産物として、一緒に賛成／反対されるトピックを抽出できる (右下)
- サイレント・マジョリティの発言や行動からトピックへの賛否を予測する手法に拡張可能



一緒に賛成／反対される別のトピック

社会保障改悪/ヘリパッド建設/平和憲法改悪/盗聴法改悪/刑事罰

オリンピック/特定秘密保護法施行/9条改定/原発再起動/憲法9

再稼働/集団的自衛権閣議決定/原発再稼働/すべての原発再稼働

戦争する国づくり/辺野古新基地建設/沖縄新基地/伊方原発/柏崎

一妻多夫制/刹処分/国際結婚/資金援助/移住/人材派遣/大阪万博

トピック間選好知識を活用した賛否分類

(Sasaki+ 2018)

- Sasaki+ (2017) は, トピック間の選好知識をモデル化しただけで, ユーザの投稿を活用できてはいない
- Matrix Factorizationの代わりにFactorization Machinesを用いて, ユーザの投稿とトピック間知識を同時に考慮しながら, ユーザの賛否を予測する手法に拡張

	ユーザ				トピック				ユーザの投稿内容 (単語)										
	ユーザ1	ユーザ2	ユーザ3	ユーザ4	大阪都構想	原子力発電所	安保法案	集団的自衛権	放射能	知事	自衛隊	戦争法案	ベクレル	日報	強行採決	...			
レコード1																			反対
レコード2																			反対
レコード3																			賛成
レコード4																			賛成

トピック間選好知識を活用した賛否分類

(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

- 各ユーザのトピックに対する賛否を予測できるか？
 - ユーザの投稿を利用することで賛否予測の正解率が向上
 - ユーザがより多くのトピックに対して賛否を言及するほど、賛否予測の性能は向上していく
- 一度も賛否に言及していないユーザ（全体の7割くらい）に対する賛否予測の正解率は65%程度と推測される

Wikipediaから獲得した促進・抑制関係に基づく賛否分類 (Hanawa+ 2018, submitted)

テキスト

セールに出かける金銭的余裕がない

否定的

どうにかして個人消費を伸ばさないといけない

肯定的

プレミアムフライデーに
対する賛否

→ 反対

→ 賛成

知識に基づく賛否分類

プレミアムフライデー

プレミアムフライデーは、2017年（平成29年）に日本国政府と経済界が提唱した個人消費喚起キャンペーンである。略称はプレ金^{[1][2][3]}。プレミアムフライデーにセールやキャンペーンを実施している企業のうち、来店者増加は約7割で売り上げ増加は5割越えであり、イベントとして当初の目的である消費の拡大には貢献している^[4]。

Wikipedia 記事（知識獲得源）

(原子力発電所 , sup, 環境)
...
(大阪都構想 , pro, 大阪府)
...
(プレミアムフライデー , pro, 個人消費)
(プレミアムフライデー , pro, キャンペーン)
(プレミアムフライデー , pro, プレ金)
(プレミアムフライデー , pro, セール)
(プレミアムフライデー , pro, 来店者増加)
...

Pro
ProBy
Sup
SupBy

促進・抑制の関係知識

Wikipediaから促進・抑制関係知識を獲得

(Hanawa+ 2017)

- 社会問題, 災害, 病気や症状, イノベーション, 政策, 金融, エネルギー技術, 生体分子, 栄養の9つのカテゴリに属する1,494記事について, 促進・抑制などの因果関係知識を付与したコーパスをクラウドソーシングで構築
- コーパスを深層学習の訓練事例として用い, 因果関係の自動抽出器を構築
- 構築したコーパスを公開中: http://www.cl.ecei.tohoku.ac.jp/wikipedia_pro_sup/

前へ 次へ

PRO:  SUP:  PRO_BY:  SUP_BY: 

脳膿瘍

腫脹

PRO: 5人

SUP: 0人

PRO_BY: 3人

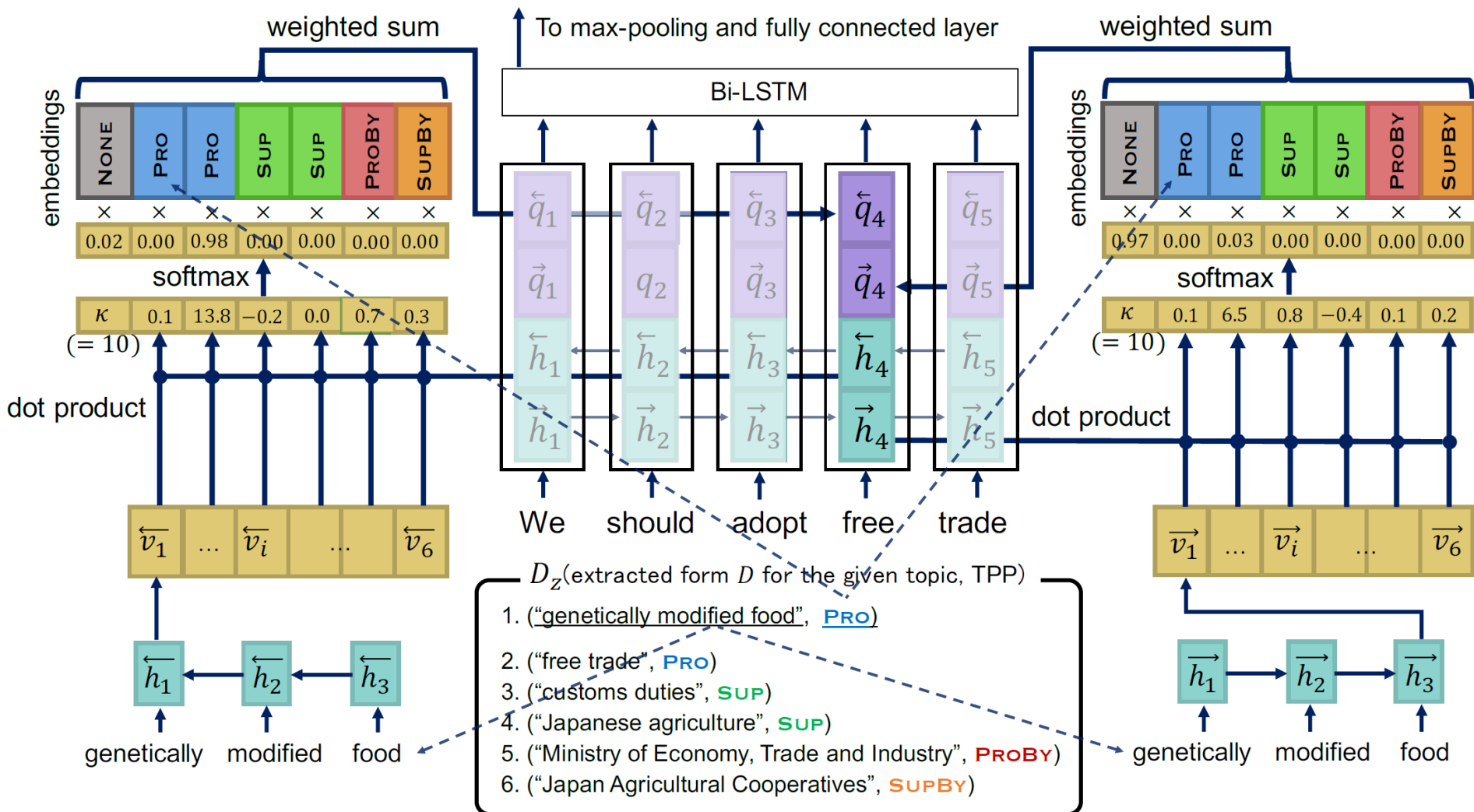
SUP_BY: 0人

脳膿瘍 (のうよう) (cerebral abscess) は、脳の病気のひとつ。
脳の中に細菌が侵入すると、脳組織内の炎症と溜まった膿によって脳が圧迫、占拠された状態をいう。
通常無菌状態の脳に細菌や真菌などの病原体が侵入して感染を起こすと、脳の組織の一部が壊死するとともに、この炎症によって腫脹が起こり、腫瘍は脳に圧力を加えることとなる。

また、体液や破壊された組織細胞、白血球、バクテリアやその死骸などが集まって膿となり、その周囲に形成される。
これらの腫瘍や膿瘍によって頭蓋内の圧力が上昇し、脳に対してさらに損傷と機能不全を引き起こす。

促進・抑制関係知識に基づく賛否分類

(Hanawa+ 2018, submitted)



促進・抑制関係知識に基づく賛否分類

(Hanawa+ 2018, 投稿中)

Model	Knowledge	Ensemble	TPP	PreFri	AntiCons	NPP	OsakaMetro	JapanMil	SelfDef	Overall
Majority baseline			0.375	0.400	0.431	0.425	0.228	0.310	0.363	0.374
Baseline		yes	0.470	0.236	0.466	0.413	0.329	0.525	0.463	0.450
			0.485	0.241	0.469	0.426	0.345	0.533	0.476	0.458
Exact match	topic	yes	0.471	0.238	0.477	0.416	0.330	0.525	0.461	0.453
			0.483	0.244	0.485	0.426	0.344	0.532	0.473	0.460
	automatic	yes	0.474	0.253	0.489	0.437	0.356	0.534	0.457	0.458
			0.479	0.253	0.516	0.428	0.359	0.551	0.462	0.467
	gold	yes	0.490	0.254	0.502	0.435	0.358	0.545	0.458	0.465
			0.506	0.242	0.514	0.462	0.356	0.566	0.460	0.476
Attention match	topic	yes	0.464	0.240	0.471	0.409	0.327	0.524	0.461	0.448
			0.481	0.244	0.473	0.424	0.344	0.535	0.475	0.457
	automatic	yes	0.464	0.242	0.501	0.415	0.342	0.562	0.483	0.458
			0.486	0.230	0.529	0.426	0.351	0.567	0.491	0.471
	gold	yes	0.491	0.241	0.501	0.438	0.355	0.525	0.486	0.473
			0.516	0.246	0.530	0.455	0.356	0.567	0.502	0.491

- 与えられたテキストに促進・抑制関係知識を照合するだけでも、賛否分類の正解率が若干向上
- アテンションを使って柔軟に知識の照合を行うことで、さらに正解率が向上
- 促進・抑制関係の知識を自動で獲得した場合でも、正解率の向上がみられる

まとめと今後の課題

- 議論マイニング
 - 議論に特化した文章の談話構造解析
 - 単一文書に対する解析の精度は比較的よい
 - 複数文書を横断的に解析するには世界知識が必要
- 世界知識を統合した議論マイニングに向けて
 - 新しいデータやタスクが提案されている
 - 自動獲得した知識で議論マイニングの性能は向上
 - 今後の課題: 知識の量・質・活用方法の底上げ
 - 具体的なフィールドでの検証・改良

参考文献

- Filip Boltuzic, Jan Šnajder. 2016. Fill the gap! Analyzing implicit premises between claims from online debates. *Proc. of ArgMining*, pp. 124-133.
- Marie-Francine Moens. 2018. Argumentation mining: How can a machine acquire common sense and world knowledge? *Argument & Computation*, 9:1-14.
- Iryna Gurevych, Chris Reed, Noam Slonim, Benno Stein. NLP approaches to computational argumentation. ACL 2016 tutorial.
- 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 in Wikipedia. *Proc. of PACLIC*.
- Kazi Saidul Hasan, Vincent Ng. 2014. Why are you taking this stance? Identifying and classifying reasons in ideological debates. *Proc. of EMNLP*, pages 751-762.
- 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. 2017. 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.