Adding Semantics to Data-Driven Paraphrasing

Ellie Pavlick, Johan Bos, Malvina Nissim, Charley Beller, Benjamin Van Durme, Chris Callison-Burch

橋本捷人（東京大・相澤研）

2015-08-30 (Sun)

論文概要

- 大規模言い換え DB の PPDB を整理
- PPDB に意味関係のラベルを付与
- 含意関係認識 (RTE) タスクの性能向上に寄与

言い換えデータベースの構築

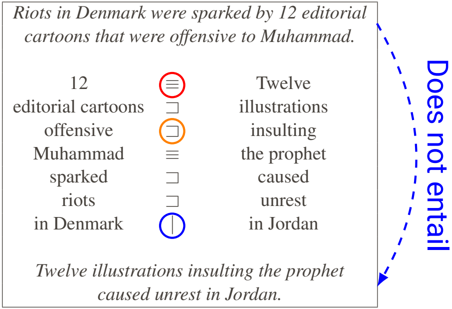
言い換えとは？
→ 双方向に含意関係のあるフレーズの組

言い換えデータベースの構築法

- 単言語コーパスから (Lin & Pantel, 2001)
  - 文脈を共有するフレーズを言い換えとする
  - 逆の意味を表すフレーズ対も得られてしまう
- 対訳コーパスから (Ganitkevitch+, 2013)
  - 別言語への翻訳を元に言い換えを得る
  - 上位語・下位語の組や関係ない組が得られることが多い

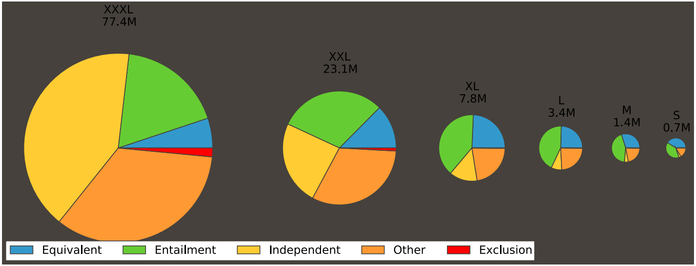
含意関係認識 (RTE)

含意関係認識タスクを解くには等価関係以外の関係も必要



ほとんどの RTE システムは WordNet を使用
→ 限られたリソース

PPDB (Ganitkevitch+, 2013)



- Equivalent: *distant/remote*
- Entailment: *girl/little girl*
- Exclusion: *nobody/somebody*
- Other (related): *swim/water*
- Independent (not related): *car/family*

言い換えの選択

- RTE データに現れるような言い換えに的を絞る
  - 最終的にはすべてのデータにラベル付けしたデータを公開する（予定？）
- SICK データセット (SemEval-2014) (Marelli+, 2014)
  - 10,000 文（訓練・テスト）
  - Entailment (29%), Contradiction (15%), Neutral (56%)
- PPDB のペア (p_1, p_2) の中で，SICK データセットで p_1 が T に p_2 が H に現れるもの
  - ただし p_1, p_2 は 3 語以下
- 9,600 ペア（訓練・テストで半々）
  - 分類器の訓練と評価はこのデータセットに対して

含意関係

MaxCartney の分類 (MacCartney, 2009)

- Equivalence: $P \equiv Q \iff \forall x.[P(x) \leftrightarrow Q(x)]$
- Forward entailment: $P \sqsubset Q \iff \forall x.[P(x) \rightarrow Q(x)]$
- Reverse entailment: $P \sqsupset Q \iff \forall x.[Q(x) \rightarrow P(x)]$
- Negation: $P \hat{} Q \iff \forall x.[P(x) \leftrightarrow \neg Q(x)]$
- Alternation: $P | Q \iff \forall x.\neg[P(x) \wedge Q(x)]$
- Cover: $P \smile Q \iff \forall x.[P(x) \vee Q(x)]$
- Independence: $P \# Q$ all other cases

アノテーション

- Amazon Mechanical Turk を使ってアノテーション
- いくつかの分類を変更
- Control questions (WordNet)
  - Accuracy = 82%, κ = 0.56

Nat. Log.	This work	MTurk description
$\equiv$	$\equiv$	X is the same as Y
$\sqsubset$	$\sqsubset$	X is more specific than/is a type of Y
$\sqsupset$	$\sqsupset$	X is more general than/encompasses Y
$\hat{}$	$\neg$	X is the opposite of Y
$\mid$		X is mutually exclusive with Y
$\#$	$\sim$	X is related in some other way to Y
	$\#$	X is not related to Y

分類器の構築

素性

- Lexical features
- WordNet features
- Monolingual features
  - Path features (Snow+, 2004)
    - ▷ フレーズペアが同時に現れる表現 e.g., more X than Y
  - Distributional features (Lin & Pantel, 2001)
    - ▷ Dependency relations を文脈としてそれらに対する指標
- Bilingual features
  - Paraphrase features
    - ▷ PPDB に付与された情報
  - Translation features
    - ▷ 別言語への対訳語の情報

ロジスティック回帰

素性詳細

Binary	p_1 is a substring of p_2
Binary	p_2 is a substring of p_1
Binary	fine-POS(p_1) == fine-POS(p_2)
Binary	coarse-POS(p_1) == coarse-POS(p_2)
Binary	Both p_1 and p_2 are lexical
Binary	Either p_1 or p_2 is phrasal
Binary (sparse)	all words in p_1, position unspecified
Binary (sparse)	all words in p_2, position unspecified
Binary (sparse)	all words in p_1, noted as p_1
Binary (sparse)	all words in p_2, noted as p_2
Binary (sparse)	all POS tags in p_1, noted as p_1
Binary (sparse)	all POS tags in p_2, noted as p_2
Real-valued	Number of words in p_1
Real-valued	Number of words in p_2
Real-valued	Number of shared POS tags
Real-valued	levenstein(p_1, p_2)
Real-valued	jaccard(p_1, p_2)
Real-valued	hamming(p_1, p_2)

Cosine Similarity		Monolingual (symmetric)		Monolingual (asymmetric)		Bilingual	
$\sqsupset$	shades/the shade	$\neg$	large/small	$\sqsupset$	boy/little boy	$\equiv$	dad/father
$\sqsupset$	yard/backyard	$\equiv$	few/several	$\sqsupset$	man/two men	$\sqsupset$	some kid/child
$\#$	each other/man	$\neg$	different/same	$\sqsupset$	child/three children	$\equiv$	a lot of/many
$\sqsupset$	picture/drawing	$\neg$	other/same	$\equiv$	is playing/play	$\equiv$	female/woman
$\sim$	practice/target	$\neg$	put/take	$\sqsupset$	side/both sides	$\equiv$	male/man

分類器の性能

	Freq.	Precision	Recall	F score
$\#$	39%	84.22	87.55	85.85
$\equiv$	8%	70.36	83.07	76.19
$\sqsupset$	26%	79.81	76.00	77.85
$\neg$	7%	73.73	73.33	73.53
$\sim$	19%	70.57	63.70	66.96

- 全体で 79% accuracy

素性の寄与

		Δ F1 when excluding					
	All	Lex.	Dist.	Path	Para.	Tran.	WN
#	79	-2.0	-0.2	-1.2	-1.7	-0.2	-0.1
≡	57	-3.5	+0.2	-0.7	-2.4	-3.7	+0.5
⊐	68	-4.6	-0.3	-0.8	-0.8	-0.7	-1.6
¬	49	-4.0	-0.8	-2.9	+0.3	-0.0	-2.2
∼	51	-4.9	-0.5	-0.7	-1.2	-0.9	-0.3

Predicted label
(using monolingual features)

True label	≡	⊐	¬	#	∼
≡	58%	20%	4%	15%	3%
⊐	20%	51%	3%	18%	7%
¬	26%	14%	37%	17%	6%
#	8%	13%	2%	71%	6%
∼	15%	21%	5%	36%	23%

Predicted label
(using bilingual features)

True label	≡	⊐	¬	#	∼
≡	62%	21%	5%	4%	8%
⊐	27%	5%	7%	7%	54%
¬	6%	14%	30%	36%	14%
#	1%	7%	6%	78%	8%
∼	8%	19%	9%	30%	35%

Predicted label
(using all features)

True label	≡	⊐	¬	#	∼
≡	83%	10%	0%	2%	4%
⊐	6%	76%	2%	7%	8%
¬	2%	8%	73%	13%	3%
#	1%	4%	2%	88%	6%
∼	5%	10%	3%	18%	64%

RTE タスクによる評価

- Nutcracker
  - State-of-the-art formal semantics based RTE system
- Baseline
  - MFC: 常に neutral に分類
  - NC alone: 外部リソースなし（lemma が同一なら等価）
  - +WN: WordNet の関係を使用
  - +PPDB-XL: PPDB-XL をすべて等価関係として使用
- PPDB+
  - 分類されたラベルを使用
- PPDB-H
  - 人手でアノテーションされたラベルを使用

RTE タスクによる評価

	Acc.	# Proofs	Coverage
MFC	56.4	0	0%
NC alone	74.3	878	17.8%
+ WN	77.5	1,051	21.3%
+ PPDB-XL	77.5	1,091	22.1%
+ PPDB+	78.0	1,197	24.3%
+ WN, PPDB+	**78.4**	**1,230**	**25.0%**
+ WN, PPDB-H	*78.6*	*1,232*	*25.0%*

- アノテーションを使ったときとそれほど変わらない精度

まとめ

- PPDB にラベルを付けた
- RTE の性能向上に寄与した
- アノテーションされた PPDB を配布（PPDB2.0 として）
  - **http://www.seas.upenn.edu/~epavlick/data.html**
  - Preliminary release だそう
  - “The full 2.0 release should be posted by the first week of August.”

分類誤りの例

True	Pred.	N	Example misclassifications
$\sim$	#	169	boy/little, an empy/the air
#	$\sim$	114	little/toy, color/hair
$\sqsupset$	$\sim$	108	drink/juice, ocean/surf
$\sqsupset$	#	97	in front of/the face of, vehicle/horse
$\sqsupset$	$\equiv$	83	cat/kitten, pavement/sidewalk
$\equiv$	$\sqsupset$	46	big/grand, a girl/a young lady
$\sqsupset$	$\neg$	29	kid/teenager, no small/a large
$\neg$	$\sqsupset$	29	old man/young man, a car/a window
#	$\equiv$	15	a person/one, a crowd/a large
$\equiv$	#	9	he is/man is, photo/still
$\equiv$	$\neg$	1	girl is/she is

分類誤りの例

	# 38%	≡ 8%	⊐ 26%	¬ 7%	∼ 18%
# 40%	1730 (clear,very) (exhibit,hold) (walk,woman)	9 (cover,front) (photo,still) (woman who,woman with)	97 (hand,male) (man,police) (mountain,side)	49 (drive,park) (female,man) (flag,ship)	169 (child,park) (crowded,many) (note,write)
≡ 10%	15 (a big,very) (a lot,long) (face a,front of)	368 (a small,the little) (away,out) (block,slab)	83 (a gun,a weapon) (a weapon,gun) (legs,leg)	9 (another man,one man) (bike,biking) (young girl,young woman)	48 (a child,kid in) (and hold,and take) (his arms,his hands)
⊐ 24%	82 (device,guy) (something,talk) (the man,the phone)	46 (a call,phone call) (a group,bunch of) (another man,man)	1004 (camera,webcam) (kid,other child) (kid,the daughter)	29 (a car,a window) (a female,a man) (arms,his hands)	97 (a lady,girl) (field,playing) (girl,the lady)
¬ 7%	35 (a ball,a man) (a boy,little) (number,woman)	1 (girl is,she is)	29 (a boy,a teenager) (a kid,daughter) (kid,little girl)	275 (cat,dog) (morning,night) (type,write)	33 (dog,owner) (ground,water) (hat,vest)
∼ 17%	114 (leg,soccer) (perform,run) (sail,water)	19 (chef,cook) (fight,match) (race,ride)	108 (cut,saw) (face,hair) (the kid,the little)	13 (a boat,sail) (dress,suit) (light,the dark)	609 (ice,rink) (snow,snowy) (study by,study the)