IDENTIFYING JUSTIFICATIONS IN WRITTEN DIALOG

Or Biran
Department of Computer Science
Columbia University
orb@cs.columbia.edu

Owen Rambow
CCLS
Columbia University
rambow@ccls.columbia.edu

IEEE-ICSC
September 19th, 2011

OUTLINE

- CONTEXT
- METHOD
- EVALUATION
- SUMMARY

CONTEXT

- **Written dialog – blogs, discussion forums**
- **Persuasion task: identify when a participant actively attempts to persuade others of an opinion**
  - More specifically, when a participant explicitly expresses an opinion or belief (makes a *claim*) and gives support for his or her claim
  - Different types of support, but most common one (>92%) is *justification*
  - Justification defined as an argument in support of an expressed claim
- **Justification task: for a pair of sentences, the first marked a claim, decide if the second is a justification of the claim**

CONTEXT

- **The corpus: LiveJournal blog threads**
  - 309 threads, annotated for claims and justifications
  - Wide range of topics:
    - Discussion of current events
    - Film / book reviews
    - Personal diary-like blogs
    - How cool that internet meme is (zomg!!!!!11)
  - Non-standard language, low register, bad punctuation, sometimes ungrammatical

EXAMPLES

CLAIM: This is a great, great record.

JUSTIFICATION: I'm hesitant to say that kind of thing because I'm not a critic; but it is certainly in a league with Robyn's very best work. The Venus 3 come together as a band in a way I don't think they really did on O' Tarantula, and it just touches me very deeply.

CLAIM: I don't think Wilf will die.

JUSTIFICATION: Wilf's going to have to kill Ten to save Donna or something, 'cause of the whole 'you've never killed a man' thing that TV woman said.

OUTLINE

- **CONTEXT**
- **METHOD**
    - INTUITION AND THEORY
    - INDICATORS
    - WORD PAIR EXTRACTION
    - SUPERVISED LEARNING
- **EVALUATION**
- **SUMMARY**

OBSERVATION

- **Justifications tend to be complex**
  - The justification itself is likely to contain some argumentation/explanation, and therefore discourse relations

- **Instead of**

  „I am in pain, ***I broke my arm***“,

  „I am in pain, ***I slipped on a banana peel and broke my arm***“ [cause]

RHETORICAL STRUCTURE THEORY (RST)

- Mann and Thompson (1987)
- Defines and characterizes a set of discourse relations
- RST Treebank

DISCOVERING RST RELATIONS - INDICATORS

- **Some relations typically contain a connector word or phrase – such as *but* for contrast. But..**
  - Sometimes it is omitted
  - Can be replaced with a paraphrase (*on the other hand*)
  - *But* is too common and ambiguous to be reliable, anyway

WORD PAIRS

- **Appearance of certain word combinations can imply a particular relation**
- **Marcu and Echihabi (2002)**
  - Used frequency of word appearance in text spans participating in fixed patterns to detect discourse relations
- **Blair-Goldensohn et al. (2007)**
  - Further developed method for **cause** and **contrast**.

WORD PAIR EXAMPLES

- **„Its *easy* to flatter people, but its *difficult* to tell the truth and say something honest that might sound mean“**

  (easy, difficult: contrast)

- **„While slavery was a *horrible* thing, we just can't act like it never *happened*“**

  (horrible, happened: concession)

- **„Canon provides an overall better *photography* system, from body to *sensor* to optics (canon Lseries lenses are something out of this world).“**

  (photography, sensor: elaboration)

WORD PAIRS

- **The words in the examples are all *content words*..**
  - Should be relevant independent of linguistic style and grammaticality
  - Can be applied to a variety of corpora, specifically the online discussions we are interested in

WORD PAIRS

- We are not interested in identifying the particular relation – many relations may contribute to argumentation...
- Instead of using fixed patterns with few standard indicators (but, because..), relax the patterns and allow many indicators
- First step: get a list of indicators

OUTLINE

- **CONTEXT**
- **METHOD**
    - INTUITION AND THEORY
    - INDICATORS
    - WORD PAIR EXTRACTION
    - SUPERVISED LEARNING
- **EVALUATION**
- **SUMMARY**

LISTS OF INDICATORS

- **RST Treebank (Carlson et al,. 2003) – a subset of the Wall Street Journal part of the Penn Treebank, annotated with RST discourse relations**
- **We chose 12 relations which are likely to participate in an attempt to make the reader accept a previously made claim:**

  - Analogy
  - Antithesis
  - Cause
  - Concession
  - Consequence
  - Contrast
  - Evidence
  - Example
  - Explanation-argumentation
  - Purpose
  - Reason
  - Result

CREATING LIST OF INDICATORS

- **First, create an ordered list of likely indicators:**
  - For each relation (e.g. **cause**) collect all text from the RST Treebank which participates in it.
  - Extract n-grams (1,2,3,4 and 5-grams)
  - Compute idf for each n-gram
  - Compute the tf variant: $tf^* = \frac{l_{ij}}{\sum_k l_{ik}}$
    - l_{ik} is the number of relation instances where the n-gram k appears at least once.
  - Sort n-grams for each relation by tf*-idf

CREATING LIST OF INDICATORS

- **Discard entries with a score less than 0.004**
- **Finally, go over the list manually from the top and remove irrelevant, ambiguous and domain-specific n-grams**
  - The RST Treebank has a relatively narrow domain: *in New York* is the second-highest ranking trigram for the **evidence** relation...
- **The result: 69 indicators, some of which are shared among multiple relations**

LISTS OF INDICATORS

Relation	Nb	Sample Indicators
Analogy	15	as a, just as, comes from the same
Antithesis	18	although, even while, on the other hand
Cause	14	because, as a result, which in turn
Concession	19	despite, regardless of, even if
Consequence	15	because, largely because of, as a result of
Contrast	8	but the, on the other hand, but it is the
Evidence	7	attests, this year, according to
Example	9	including, for instance, among the
Explanation-argumentation	7	because, in addition, to comment on the
Purpose	30	trying to, in order to, so as to see
Reason	13	because, because it is, to find a way
Result	23	resulting, because of, as a result of

OUTLINE

- **CONTEXT**
- **METHOD**
    - INTUITION AND THEORY
    - INDICATORS
    - WORD PAIR EXTRACTION
    - SUPERVISED LEARNING
- **EVALUATION**
- **SUMMARY**

EXTRACTING WORD PAIRS

- **The list of indicators is used to extract word pairs from English Wikipedia which co-occur with an indicator in the same sentence**
- **Two lists:**
  - *Sides* - the first word occurs on the left and the second word on the right of the indicator.

    447,149,688 pairs

  - *Anywhere* – words may occur anywhere in the sentence (in order).

    1,017,190,824 pairs

- **No stop words – we only want content words**
- **Pairs which appear less than 20 times are removed, reducing list sizes to 334,925 (*sides*) and 719,439 (*anywhere*)**

OUTLINE

- **CONTEXT**
- **METHOD**
    - INTUITION AND THEORY
    - INDICATORS
    - WORD PAIR EXTRACTION
    - SUPERVISED LEARNING
- **EVALUATION**
- **SUMMARY**

USING THE INFORMATION

Task: for each pair of sentences, the first marked a claim, decide if the second is a justification of the claim

USING THE INFORMATION

- **Disjunctive word-pair feature for each indicator:**

$$\phi_j = \begin{cases} 1 & if\ the\ candidate\ sentence\ contains\ any\ pair\ p \in P_j\ with\ some\ constraints \\ 0 & otherwise \end{cases}$$

- **Three constraint variants:**
  - Unigrams – positive if either word appears
  - Unordered – positive if both words appear, in any order
  - Ordered – positive if both words appear in their original order

EXAMPLE

CLAIM: I don't think Wilf will die.

JUSTIFICATION: Wilf's going to have to kill Ten to save Donna or something, 'cause of the whole 'you've never killed a man' thing that TV woman said.

- ***because* is an indicator for** cause **and** reason**; *in order to* is an indicator for** purpose
- **(kill, save) appear in Wikipedia around *in order to***
- **(kill, killed) appear in Wikipedia around *because***
- **Both features are positive, and the sentence classified as a justification**

OUTLINE

- CONTEXT
- METHOD
- EVALUATION
- SUMMARY

EVALUATION

- 309 LiveJournal threads, annotated for claims and justifications
- 40 reserved for a test set, 269 used for training. 10-fold cross-validation used for development
- Data set of sentence pairs, composed by using all claims with all candidate justifications. Candidates are all sentences which belong to an entry that is equal or subsequent to that containing the claim and authored by the same participant
- 6636 training data points and 756 test instances. Approximately 10% are positive in both sets

BASELINES

1. **Sentence following claim**
2. **High-recall heuristic: all sentences that are**
   - In the same entry as the claim
   - No more than 4 sentences away from the claim
3. **Hybrid: use the heuristic constraints, then a Naive Bayes classifier with 2 features:**
      I. Candidate length
      II. Position of candidate: before or after the claim
   - Also a post-processing step which picks at most two justifications per claim
4. **Hybrid + Bag-of-words: all words which appear more than 5 times in the data set, as individual features.**

EVALUATED SYSTEMS

- **Tested systems: Hybrid + additional features**
  - Indicators as lexical features
  - Word-pair disjunctions [Unigrams]
  - Word-pair disjunctions [Unordered] *anywhere*
  - Word-pair disjunctions [Ordered] *anywhere*
  - Word-pair disjunctions [Unordered] *sides*
  - Word-pair disjunctions [Ordered] *sides*
  - Word-pair disjunctions (best from above) plus indicators

Word pair lists	Feature constraints
Sides – pairs from both sides of the indicator *Anywhere* – pairs from anywhere the sentence	[Unigrams] – either word appears [Unordered] – both words appear, in any order [Ordered] – both words appear in their original order

EVALUATION

System	CV P	CV R	CV F	Test P	Test R	Test F
next sentence	**46.4**	32.4	38.2	41.7	40	40.8
heuristic baseline	29	**91**	44	27.2	**88.4**	41.6
hybrid baseline	41.5	54.7	47.2	31.7	45.6	40.7
hybrid baseline + bag-of-words	41.4	48.6	44.7	37.5	43.7	40.4
hybrid baseline + indicators	41.5	54.7	47.2	31.7	45.6	40.7
hybrid baseline + unigrams	42.1	56.5	48.3	35.4	46	40
hybrid baseline + *anywhere* with ordering	35.6	20.9	26.3	34.9	17.5	23.3
hybrid baseline + *anywhere* with no ordering	38.2	19.8	26.1	41.7	19.8	26.9
hybrid baseline + *sides* with ordering	42.9	61.6	50.6	**42.6**	53.4	**47.4**
hybrid baseline + *sides* with no ordering	43	61.2	50.5	41.9	52.4	46.6
hybrid baseline + indicators + *sides* with no ordering	**43.1**	**61.8**	**50.8**	**41.9**	**52.4**	**46.6**
hybrid baseline + indicators + *sides-no-stoplist* with no ordering	42.1	58.2	48.8	37.1	47.6	41.7

EVALUATION

- **Another experiment: single sentences, with no claims**
- **8508 training data points, 1197 test**
- **No heuristic. Only word-pair disjunction features: [Unordered] *sides* (from best system)**
- **Baseline is greedy all-positive**

System	CV P	CV R	CV F	Test P	Test R	Test F
baseline	11.7	100	20.9	14.8	100	25.7
sides with no ordering	30.9	48.9	37.8	30.3	40	34.5

OUTLINE

- CONTEXT
- METHOD
- EVALUATION
- SUMMARY

SUMMARY

- **We have proposed a method of identifying justifications in text, particularly suited to written dialog**
- **On the justification-for-claim task, the results of our best system are consistently better than those of 4 baselines and of weaker systems**
- **Without claims, could be used as a general *argumentation detector*, but we did not evaluate (no gold data)**
- **The indicator list used to mine Wikipedia for word pairs is publicly available at**

  http://www.cs.columbia.edu/~orb