

Crowdsourcing syntactic relatedness judgements for opinion mining in the study of information technology adoption

Asad B. Sayeed, Bryan Rusk, Martin Petrov, Hieu Nguyen,
Timothy Meyer, and Amy Weinberg¹

¹University of Maryland, College Park

Overall objectives

- **The application context:** corpus-based social science
 - Identifying opinion-leaders and large scale trends.
 - Our domain: IT innovations, via the IT business press (ComputerWorld, InformationWeek).
- **The approach:** information extraction.
 - **Input:** Corpus of news articles, named entities detected (by e.g. IdentiFinder).
 - **Output:** {source, target, opinion} triples by document.
 - target = mention of opinion “topic”.
 - opinion \in {positive, negative}
 - Standard IR evaluation metrics: precision, recall, F-measure, accuracy.
 - **Goal:** High quality for downstream machine learning applications.

IT corpus and concepts

- Information technology (IT) business journal: InformationWeek
 - Approx 33K articles of varying lengths (news blurbs, full-feature articles).
 - Approx 75K sentences containing IT concept mentions.
 - OpenNLP splitter.
 - Years covered: 1991-2008.
- IT concepts: same 60 including “Enterprise resource planning”, “application service provider”, and abbreviations and variants.

Identifying opinion

- Sentence by sentence.

Example from corpus

Lloyd Hession, chief security officer at BT Radianz in New York, said that virtualization also opens up a slew of potential network access control issues.

- Choose a {*source, target, opinion*} triple: {Lloyd Hession, virtualization, negative}
- What is the evidence that Mr. Hession has a negative opinion about virtualization?

Identifying opinion

- What do we want to know about the target?

Example: information technology business press

*Lloyd Hession, chief security officer at BT Radianz in New York, said that **virtualization** also opens up a slew of potential network access control issues.*

- If we're modeling a business journal, then what matters is the *reader*—the market.
- Opinion is held by “aggregate reader”..
- Need world-knowledge: “pragmatic opinion” (Somasundaran and Wiebe, 2009).

Identifying opinion

- From whose perspective does the sentence contain an opinion?

Example: information technology business press

*Lloyd Hession, chief security officer at BT Radianz in New York, said that **virtualization** also opens up a slew of potential network access control issues.*

- From IT domain:
 - “Virtualization” was important concept.
 - The aggregate reader/market would view this as a potentially negative outcome.

Identifying opinion

- What words give us a negative sentiment?

Example: information technology business press

*Lloyd Hession, chief security officer at BT Radianz in New York, said that **virtualization** also opens up a **slew** of potential network access control **issues**.*

- “slew” and “issues”: convey negative sentiment about “virtualization”.
- How do we know they’re negative in this domain?
- How do we know that they apply to “virtualization”?
- What about words like “update”? Important in IT domain, not mentioned in major polarity lexicon.

Defining opinion

Problem

Not easy to find domain-relevant sentiment words in context.

Task

Use crowdsourcing to construct word-level annotations with links to domain concepts/targets.

- Existing resources (MPQA, JDPA) annotate spans, not individual words.

What we needed

- Experts in IT domain: very expensive.
- Answer: crowdsourcing—unskilled labour from Internet via Mechanical Turk.
- Interface desiderata:
 - It should be no easier to cheat than to answer intelligently.
 - It should be more fun than a multiple-choice test.
 - It should make the “what is an opinion” decision implicit in the task.

Metasubjectivity

Something we need to eliminate: Turns out that “what is an opinion?” is a matter of opinion. . .

Solution

- Drag and drop interface task design:
 - One sentence per task, plus context.
 - Candidate words highlighted.
 - Four boxes: positive, negative, no opinion, can't tell.
 - Each highlighted word must be dragged to one box.
- Boxes are equally “difficult” to drag words to—not much harder to answer intelligently than to cheat.
- “What is an opinion” is not directly asked.

Drag-and-drop interface

New Technologies Along these lines, the fourth strategy is that IT management must let people learn new technologies as they emerge and learn how they apply to business.

"In the world of e-business, it becomes mandatory for IT professionals to understand both business strategy and business processes," says John Finegan, president and CEO of Management Technology Group Inc., a systems integration firm in Englewood, Colo.

"The development of E-commerce software products requires that a business need be translated into accurate application specifications." Thus, people involved in systems design and implementation must not only clearly understand the needs of the customer (relationship management), they must have a passion for getting the product done to the specification of the customer in a timely manner, and be cost-effective.

No effect on
opinion towards IT
concept

Affects opinion
positively

Affects opinion
negatively

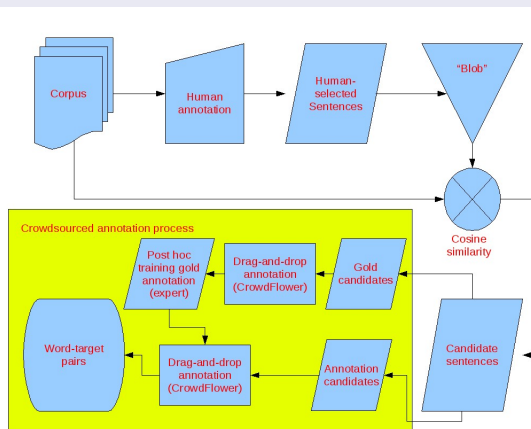
Can't decide

Number of items remaining to classify: (required)

Procedure

- Schematic view:

Drag-and-drop interface strategy



Experiment

- 200 work units with approx 1200 highlighted words
- 3 users/task, 4 cents/task. (Total \$60, incl. CrowdFlower fees.)
- Aggregation: majority vote, with ties dropped.
- Task finished overnight.
- CrowdFlower's task apportionment means that some tasks had 4+ answers.

Post hoc processing

- Baseline: assign, wherever possible, the polarity from a major sentiment lexicon, or none if unavailable.
- Quality control: exclude input from low-quality workers. (Threshold against gold standard.)
- Evaluation: on 30 tasks done by us.
 - Retrieval task: precision, recall
 - Agreement as Cohen's κ —not often used in fine-grained sentiment analysis.

Results

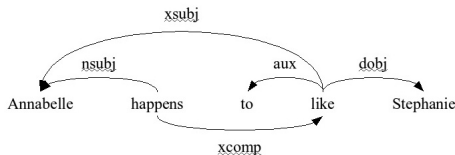
- Experiments:

Excluded	Words lost (of 48)	Prec/Rec/F	Cohen's κ
(prior polarity)	N/A	0.87 / 0.38 / 0.53	-0.26
0	0	0.64 / 0.71 / 0.67	0.48
1	0	0.64 / 0.71 / 0.67	0.48
3	0	0.66 / 0.73 / 0.69	0.51
5	0	0.69 / 0.73 / 0.71	0.53
7	2	0.81 / 0.76 / 0.79	0.65
10	9	0.85 / 0.74 / 0.79	0.54
12	11	0.68 / 0.68 / 0.68	0.20

- Stringent quality control can cause some words to be lost.
- Future work: syntactic phenomena, multiword expressions.

Example downstream application

- Find grammatical links between targets and opinion words.
- Example: Stanford dependency parse for “Annabelle happens to like Stephanie”:



- Would like to be able to use grammatical info for e.g. polarity classification. (Jakob and Gurevych, 2010)

Summary

- Existing resources that contain span-entity links: MPQA, JDPa.
 - Lack of word-level annotation means that entire clauses are treated as opinionated language.
- Resource creation expensive for IT business press domain—created word-level resource inexpensively by crowdsourcing.
- Handled metasubjectivity via UI design, multilayered quality control.