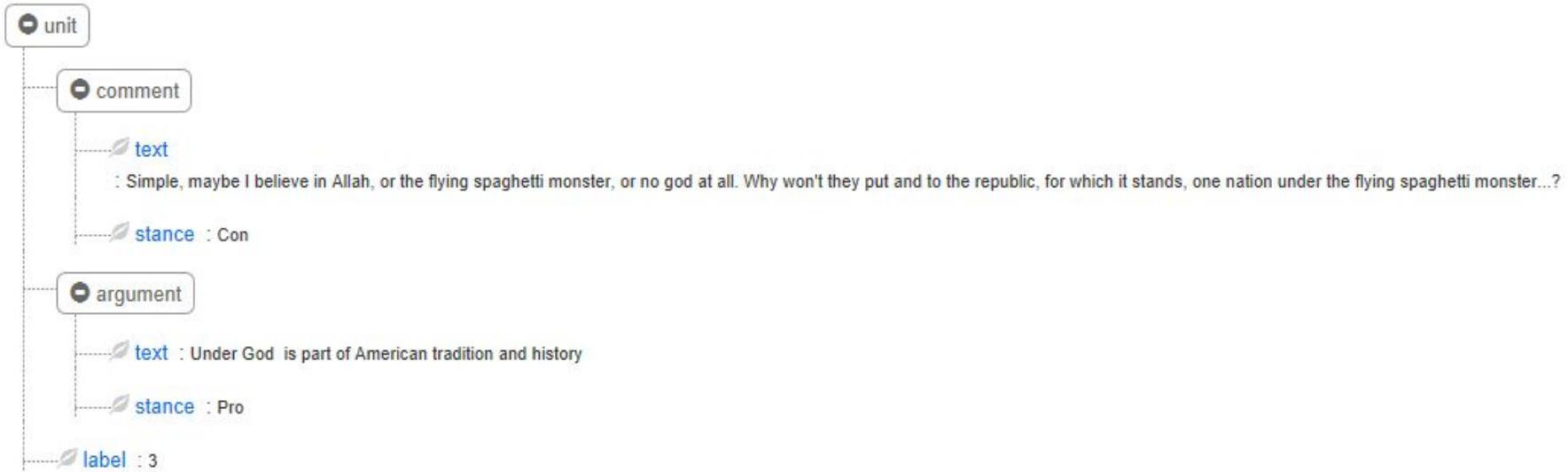
Corpora

- ComArg - Pairs of Comments / Arguments of two different debate topics each labeled for stance and for entanglement relationship between the pair.
- Sara - Sentences from livejournal and wikipedia with claim annotation and intra-sentence subjective/objective annotations.
- MPQA arguing - Subset of Multi-Perspective Questioning extended for arguing opinion private state annotations. The MPQA annotation scheme builds upon the private state notion which describes mental states including opinions, emotions, speculations and beliefs among others. The annotation scheme strives to represent the private states in terms of their functional components (i.e. experiencer holding an attitude towards a target). It consists of frames (direct subjective, expressive subjective element, objective speech event, agent, attitude, and target frames) with slots representing various attributes and properties (e.g.intensity, nested source) of the private states.

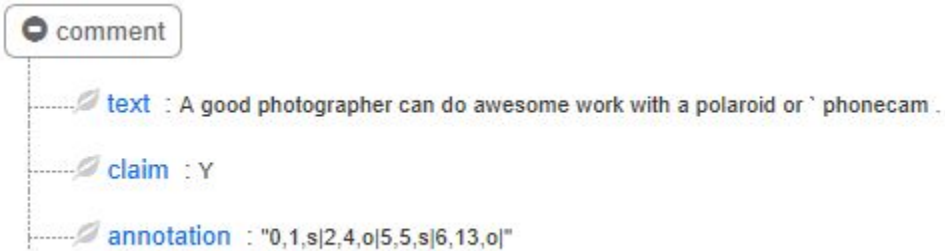
Examples

comArg



Examples

Sara



Examples

Arguing Corpus

Texts from blog posts on ObamaCare labeled for arguments and relations.

A thorough argument structure is provided

Pull the plug, Barack: It's a mistake to try to ram through health care reform
Sunday, March 14th 2010, 4:00 AM
Health reform: Down the stretch it comes. President Obama has gone so far as to postpone an overseas trip in a last-ditch push to get a comprehensive reform bill to his desk.

He shouldn't waste his energy. Not with unemployment justifiably the nation's top concern and the possibility of a double-dip recession still looming. Remember how Obama said, in his State of the Union reboot, "Jobs must be our No. 1 focus in 2010"? Well, apparently he doesn't. Not with the American people abandoning the President's prescription in huge numbers. Just one in four voters supports the reform bill as written; half want Congress to start over.

Compare that with the popular support other major pieces of social legislation enjoyed before passage, like welfare reform (68%), Medicare (63%) and civil rights (60%). Not with health care costs having risen 73% over the last decade - with Medicaid growing at 21% a year - and showing no signs of coming down to Earth.

Controlling costs is the absolute, unconditional, a-blind-man-could-see-it prerequisite for expanding coverage. Not with American health care quality actually quite impressive in many respects, including world-leading rates for patients surviving cancer and among the shortest wait times for hard-to-find treatments and surgeries. Not with a half-dozen accounting gimmicks built into the legislation - including the fact that federal budget projections are based on 10 years worth of tax collections and just six years of spending increases. Not with the Senate, lacking even a single Republican vote, having to resort to reconciliation, a little-used parliamentary maneuver, to get it through. Sure, it's been used before - but not on anything that has such limited public support. Not after all the cynical back room deals - deals Obama has committed to removing - having frayed the public trust. With legislation of this magnitude, that is not a renewable resource.

Obama clearly thought he was doing the right and necessary and bold thing by tackling health reform. His conviction has been admirable. But reality has come a-knocking, quite persistently, and it's saying: Enough already.

Fix what's broken, then move on. He should assemble a modest consensus package of two or three smart ideas, such as outlawing insurance company abuses - like discrimination based on preexisting conditions - speeding the adoption of electronic medical records and reining in the medical malpractice market. Only a targeted approach will begin to tackle the central problem plaguing American health care, which is how terribly much it all costs.

And only a targeted approach will carry the promise of bipartisan action that will not set off a new civil war in Washington.

And it will enable Obama and the Democrats - who run Washington - to get back to creating jobs. Which happens to be the best way for them to keep their own jobs secure.

Type	Set	Start	End	Id	Features
rel		186	201	2	{id=r1}
arg		361	471	3	{polarity=positive, ss-label=unemployment_higher_priority_than_healthcare, type=label}
rel		485	569	4	{ss-label=unemployment_higher_priority_than_healthcare, type=label}
arg		609	684	5	{polarity=negative, ss-label=unpopular_among_public, type=label}
rel		686	782	6	{ss-label=unpopular_among_public, type=label}
arg		1108	1219	7	{incomplete=true, polarity=positive, ss-label=does_not_control_healthcare_costs, type=label, unclear=true}
arg		1228	1300	8	{group=self-group, polarity=positive, prospect=gain, ss-label=healthcare_quality_good, type=label}
rel		1312	1438	9	{ss-label=healthcare_quality_good, type=label}
arg		1448	1507	10	{incomplete=true, polarity=negative, ss-label=increases_govt_spending, type=label, unclear=true}
rel		1534	1649	11	{ss-label=increases_govt_spending, type=else}
arg		1659	1801	12	{polarity=negative, ss-label=not_bipartisan, type=label}

reminder:

- CAPS words
- >2x Punct

General info

	comArg (GM + UGIP)	sara (livejournal + wikipedia)	arguing (obamacare)
words	192717 + 110955	28341 + 31291	68530
unique words	3446 + 2473	4785 + 4346	7522
lexical diversity	0.1788 + 0.223	0.1688 +0.1389	0.1098
punctuation			

Annotation Comparison

	comArg (GM + UGIP)	Arguing
anti	97 + 180	47
pro	101 + 170	37
total	198 + 350	84

	sara (livejournal + wikipedia)
claim	1198 + 1282
non claim	790 + 715
total	1988 + 1997

Related work

comArg

Boltuzic, Filip, and Jan Snajder. 2014. **"Back up Your Stance : Recognizing Arguments in Online Discussions."** Proceedings of the First Workshop on Argumentation Mining, 49–58. http://aclweb.org/anthology/W/W14/W14-2107.pdf.

Abstract

In online discussions, users often back up their stance with arguments. Their arguments are often vague, implicit, and poorly worded, yet they provide valuable insights into reasons underpinning users' opinions. In this paper, we make a first step towards *argument-based opinion mining* from online discussions and introduce a new task of *argument recognition*. We match user-created comments to a set of predefined topic-based arguments, which can be either attacked or supported in the comment. We present a manually-annotated corpus for argument recognition in online discussions. We describe a supervised model based on comment-argument similarity and entailment features. Depending on problem formulation, model performance ranges from 70.5% to 81.8% F1-score, and decreases only marginally when applied to an unseen topic.

Related work

sara

Rosenthal, Sara, and Kathleen McKeown. 2012. **“Detecting Opinionated Claims in Online Discussions.”** In Proceedings - IEEE 6th International Conference on Semantic Computing, ICSC 2012, 30–37. doi:10.1109/ICSC.2012.59.

***Abstract*—This paper explores the automatic detection of sentences that are opinionated claims, in which the author expresses a belief. We use a machine learning based approach, investigating the impact of features such as sentiment and the output of a system that determines committed belief. We train and test our approach on social media, where people often try to convince others of the validity of their opinions. We experiment with two different types of data, drawn from LiveJournal weblogs and Wikipedia discussion forums. Our experiments show that sentiment analysis is more important in LiveJournal, while committed belief is more helpful for Wikipedia. In both corpora, n-grams and part-of-speech features also account for significantly better accuracy. We discuss the ramifications behind these differences.**

Related work

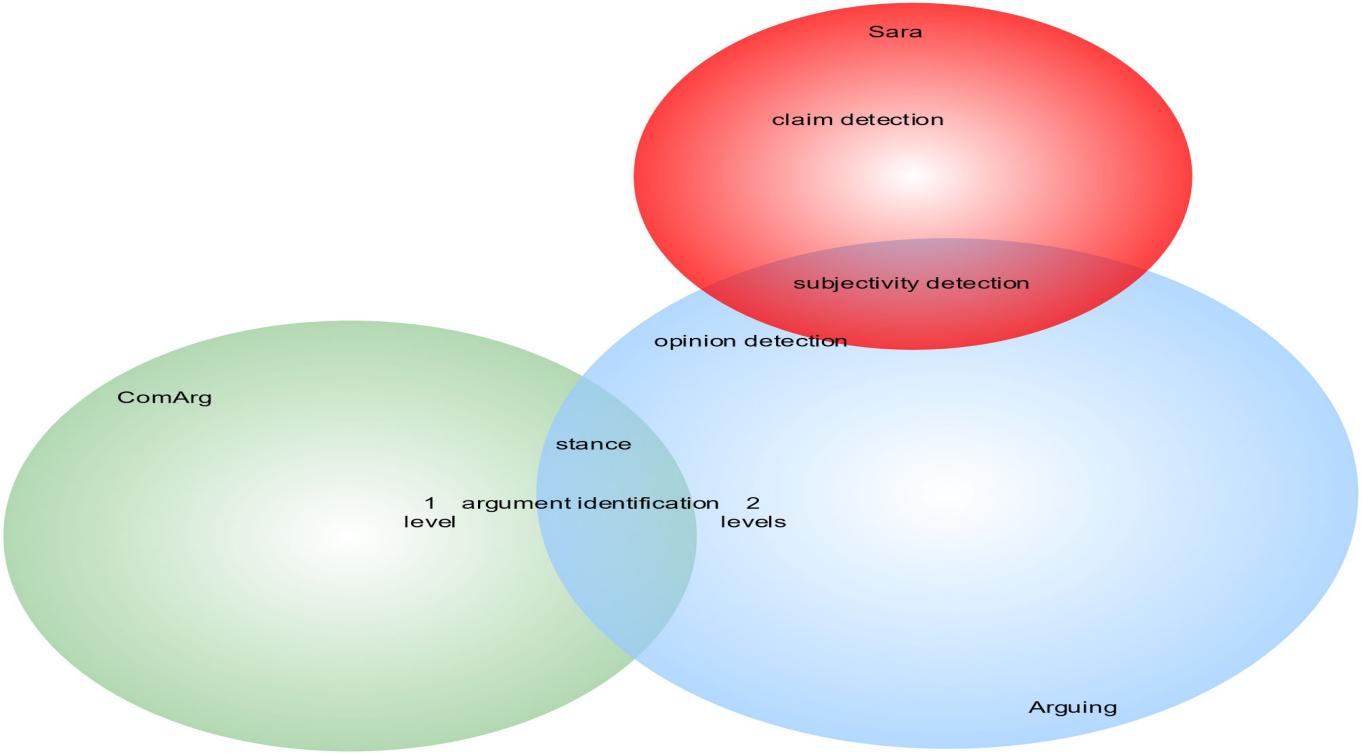
arguing

Conrad, Alexander, Janyce Wiebe, and Rebecca Hwa. 2012. **"Recognizing Arguing Subjectivity and Argument Tags."** Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics, no. July: 80–88. http://dl.acm.org/citation.cfm?id=2392711.

Abstract

In this paper we investigate two distinct tasks. The first task involves detecting arguing subjectivity, a type of linguistic subjectivity on which relatively little work has yet to be done. The second task involves labeling instances of arguing subjectivity with argument tags reflecting the conceptual argument being made. We refer to these two tasks collectively as "recognizing arguments". We develop a new annotation scheme and assemble a new annotated corpus to support our learning efforts. Through our machine learning experiments, we investigate the utility of a sentiment lexicon, discourse parser, and semantic similarity measures with respect to recognizing arguments. By incorporating information gained from these resources, we outperform a unigram baseline by a significant margin. In addition, we explore a two-phase approach to recognizing arguments, with promising results.

Applications



Insights

Both in *ComArg* and *Arguing* we have comments and their relation to central propositions as well as stance towards that proposition which usually allow us to extrapolate the general stance.

These central propositions define a argumentative structure.

I would say that these propositions are not the same as opinionated claims because they possess no sentiment polar charge, neither arguing attitude as they only objectively state the claim (or sub-claim) in an neutral way.

ComArg's argumentative structure is 1 level deep with 3 propositions for each stance and there is no segmentation at the comment.

Arguing argumentative structure is 2 levels deep and is intensively annotated for several private states and relations at the phrase level.

The hypothesis here is that by detecting the overall stance towards the topic we could roll back, identify the facet (central proposition) which would bring us closer to claim identification.

More Insights

Sara’s is a dataset that has several desired capabilities:

- word level subjectivity annotations
- sentence level claim annotations
- comments are from online debate sources (but the comments have no explicit connection between them ex:topic)

Unexplored alternatives

- IBM corpora for claim/evidence detection [Levy et al., 2014, Rinott et al., 2015]

  The benchmark dataset for claim detection.

- Online debates : Biran, Or, and Owen Rambow. 2011. **"Identifying Justifications in Written Dialogs."** In Proceedings - 5th IEEE International Conference on Semantic Computing, ICSC 2011, 162–68. doi:10.1109/ICSC.2011.41.

  309 blog threads from LiveJournal , belonging to users from English-speaking countries. Each thread contains an original entry by the blog owner and a set of comment entries, in a tree structure , by other LJ users as well as the owner. The threads contain annotated claims and their corresponding justifications. A justification can only be made by the same poster who made the original claim, but it may be located in a different entry. All annotated claims have justifications, and a claim may have more than one justification.

  - Text, claim and justification are distinct spans of text vaguely written by the users themselves, not by trained annotators.
  - Lots of different topics, dialogic source (conversation tree)

The cross-road