ACL 2016
AUGUST 7 – 12 | BERLIN, GERMANY

Context-aware Argumentative Relation Mining

Huy Nguyen[1] Diane Litman[1,2]

[1]Department of Computer Science | [2]Learning Research and Development Center, University of Pittsburgh, PA, USA

Contact: Huy Nguyen, hvn3@pitt.edu

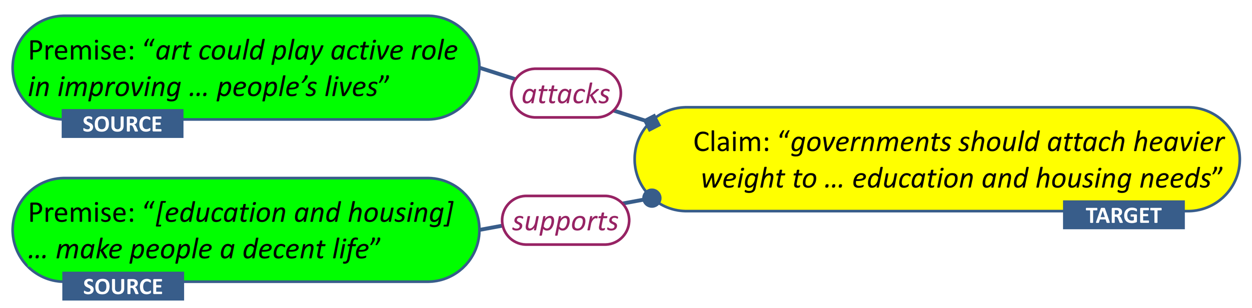
This research is supported by NSF/1122504

Introduction

- An increasing need of automatically identifying and validating **natural language argumentation** in large scale

Argumentation mining in text involves automatically:
- Identifying **argument components**, e.g., premises, claims,
- Classifying **argumentative relations**, e.g., support, attack, between source and target components

To conclude, (1)art could play an active role in improving the quality of people's lives[Premise], but I think that (2)governments should attach heavier weight to other social issues such as education and housing needs[Claim] because (3)those are the most essential ways enable to make people a decent life[Premise].

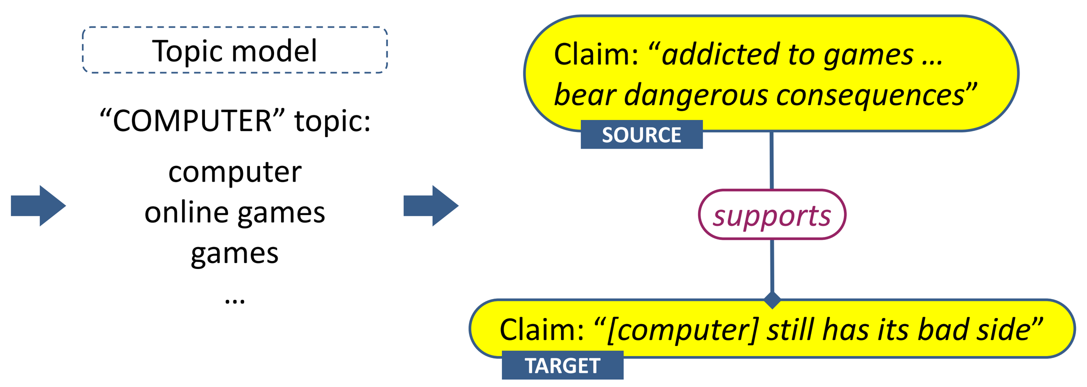


Context-aware Argumentative Relation Mining
- Our study proposes novel **topic-context** and **window-context features** for improving argumentative relation mining

Topic-context

Essay 24. Topic: computer has negative effects to children.

…(1)People who are addicted to games, especially online games, can eventually bear dangerous consequences[Claim]. (2)Although it is undeniable that computer is a crucial part of human life[Premise], it still has its bad side[MajorClaim].



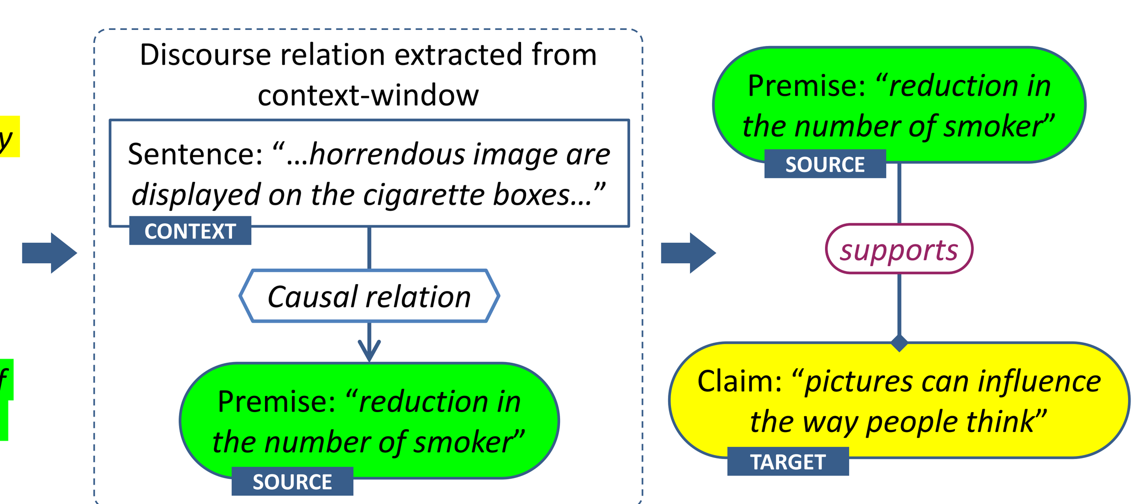
Topic context
- Knowing which words are topically related may help determine relations across components

Argument and domain word extraction algorithm (Nguyen & Litman, 2016)
- Essay prompts are used to supervise the argument/domain word separation from **LDA output**
- **Argument word**: indicators of argumentative content, e.g., *think*, *believe*
- **Domain word**: specific terminologies commonly used within the topic, e.g., *computer*, *game*

Window-context

Essay 73. Topic: is image more powerful than the written word?

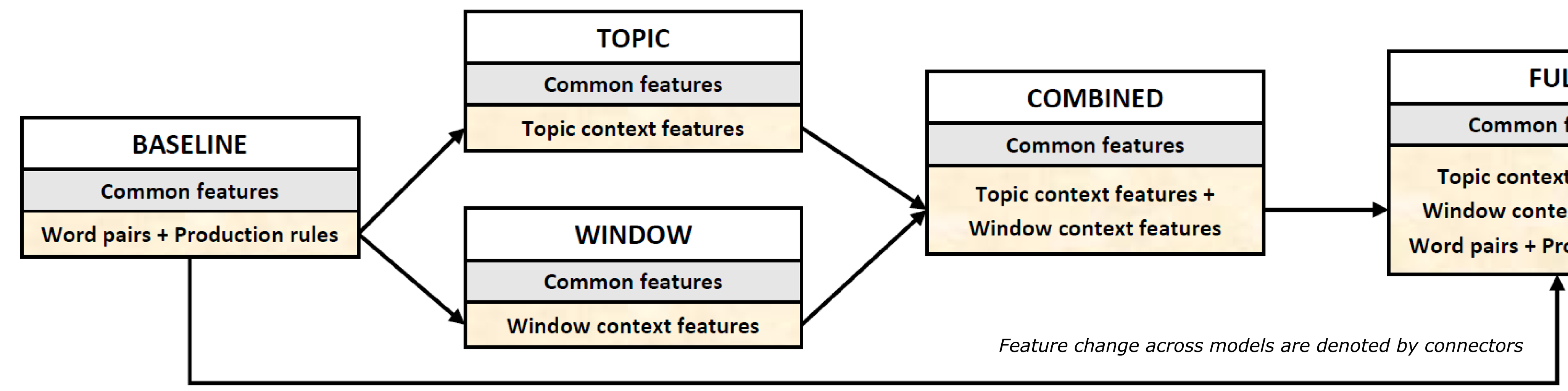
…(2)Firstly, pictures can influence the way people think[Claim]. (3)For example, nowadays horrendous images are displayed on the cigarette boxes to illustrate the consequences of smoking[Premise]. (4)As a result, statistics show a slight reduction in the number of smokers, indicating that they realize the effects of the negative habit[Premise]…



Window context
- Consider surrounding sentences (i.e., context window) of the source and target components
- Discourse relation in context windows help characterize the argumentative relation

Both **PDTB** and **RST discourse relations** are extracted from context windows to derive Window-context features

Prediction Features in Different Models



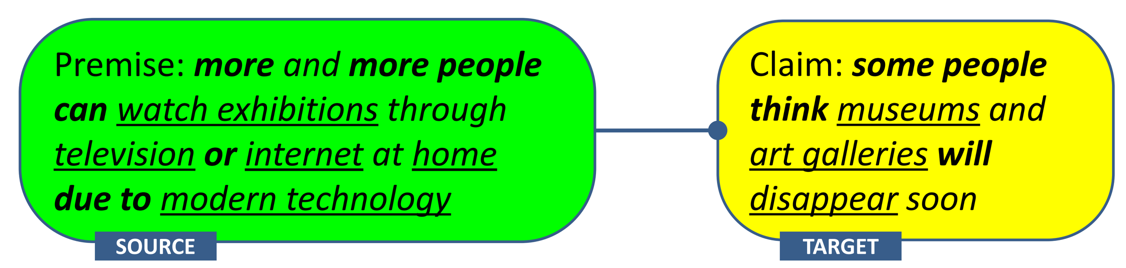
Common features (features in common among models)
- Structural (word count, sentence position, component position)
- Lexical (pairs of first words, discourse connectives)
- Predicted labels of argument components (by using Nguyen & Litman's model, 2016)

Baseline model is adapted and improved from (Stab & Gurevych 2014b)
- Common features + word pairs + production rules (e.g., S → NP VP)

TOPIC, **WINDOW** and **COMBINED** models are for evaluating Topic-context and Window-context features in isolation and combination

FULL model takes all features together

Topic-context Features



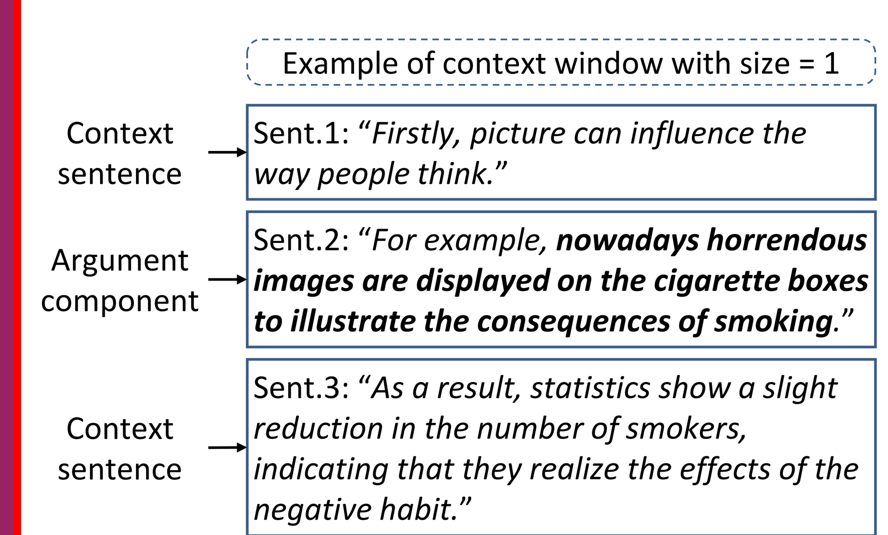
Argument & domain word extraction (Nguyen & Litman 2016)
- Development data of 6794 un-annotated essays
- Extracted 263 **argument words**, 1806 domain words

Domain word counts
- Domain words in common, e.g., 0
- Pairs of two domain words that are in the same topic, e.g., (*exhibitions*, *art*)
- Pairs of two domain words that are not in the same topic, e.g., (*internet*, *art*)
- Absolute difference in number of domain words, e.g., 3

Argument words
- Argument words, e.g., *think*
- Argument word pairs, e.g., (*people*, *people*)
- Argument words in common, e.g., *people*
- Absolute difference in numbers of argument words, e.g., 3

MainVerb-Subject triples
- MainVerb-Subject dependency triples that do not involve domain words, e.g., *nsubj*(*people*, *think*)

Window-context Features



Window-size heuristics
- Given window-size *n*, context window includes at most *n* adjacently preceding and *n* adjacently following sentences of the argument component
- Context sentences must be in the same paragraph with the argument component
- Context windows of the source and target components must not overlap

Best window-size = 3 determined using a development set

Common words
- Number of words in common between the argument component and preceding/following context sentences

Discourse relations
- Between context sentences
- Within covering sentence of the argument component
- Between each pair of source context sentence and target context sentence

Discourse connectives
- Connectives in covering sentence
- Whether connectives are before the covering sentence

Experimental Results

Data (Stab & Gurevych 2014a)
- 90 persuasive essays annotated for argument components in sentences, i.e., **major claim**, **claim**, **premise**, and argumentative relations between components, i.e., **support** and **attack**

Task 1: Support vs. Non-support (80% data for training and 20% data for testing)
- 6330 pairs of argument components in the same paragraph
- 989 (16%) Support vs. 5341 (84%) Non-support pairs (contain 103 attack relations)
- Models are compared to reported results in (Stab & Gurevych 2014b)

Topic-context + Window-context features are more effective than word pairs + production rules
- COMBINED outperforms all other models

Using word pairs and production rules even degrades effectiveness of our context features
- FULL performs worse than COMBINED

Due to imbalanced data, we report only Kappa, F1, and F1:Support (minor class) which are more important metrics

	REPORTED	BASELINE	TOPIC	WINDOW	COMBINED	FULL
Kappa	-	0.445	<u>0.407</u>	0.449	**0.507***	0.481
Macro F1	0.722	0.722	<u>0.703</u>	0.724	**0.753***	0.739
F1:Support	0.519	0.519	<u>0.488</u>	0.533	**0.583***	0.550

*Best values in **BOLD**, significant difference from BASELINE denotes by *, smaller values than BASELINE are <u>underlined</u>*

Task 2: Support vs. Attack (5x10-Fold cross validation)
- 1473 pairs hold argumentative relations: 1312 (89%) Support and 161 (11%) Attack
- All models outperform BASELINE, COMBINED obtains the best performance
- FULL performs significantly worse than TOPIC, WINDOW, and COMBINED

Results further prove the effectiveness of Topic-context and Window-context features
- TOPIC and WINDOW models significantly outperform BASELINE
- COMBINED has significantly higher performance than FULL

Due to imbalanced data, we report only Kappa, F1, and F1:Attack (minor class)

	BASELINE	TOPIC	WINDOW	COMBINED	FULL
Kappa	0.245	0.305*	0.306*	**0.342***	0.274*
Macro F1	0.618	0.651*	0.652*	**0.670***	0.634*
F1:Attack	0.300	0.365*	0.376*	**0.404***	0.330*

Conclusions

- We have proposed novel contextual features for improving argumentative relation mining
- Our proposed features exploit both global (topic-context) and local (window-context) contextual information
- Our combined model significantly outperformed a state-of-the-art baseline