Temporality and Modality in Entailment Graph Learning

Liane Guillou

with Sander Bijl de Vroe, Mohammad Javad Hosseini, Miloš Stanojević, Nick McKenna,
Mark Johnson, Mark Steedman

Edinburgh
University of Edinburgh
Natural Language Processing
NLP

Table of contents

Introduction

The SEMANTAX Project

- Aim: Learn entailments between predicates from raw text

Example: buy $\rightarrow$ own

Google **bought** YouTube for $1.65 billion.
Google **owns** YouTube and it has proven to be an amazingly successful purchase.

- Use entailment information in downstream applications:
  - Question Answering
  - Knowledge Graph population

Recognising Textual Entailment and Question Answering

- Question: Did Arsenal play Man United last night?

Recognising Textual Entailment and Question Answering

- Question: Did Arsenal play Man United last night?

Match Report

"Arsenal beat Man United 1-0"

- Question: Did Arsenal play Man United last night?

Match Report

"Arsenal beat Man United 1-0"

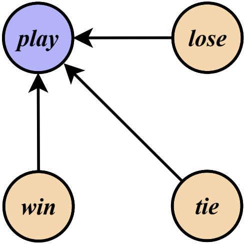
- To answer, we must know the entailment relation:

beat $\rightarrow$ play

TeamA beats TeamB $\rightarrow$ TeamA plays TeamB

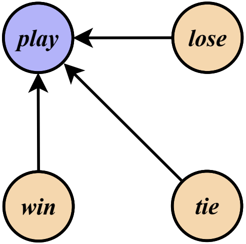
Learning Entailment Graphs

Entailment Graphs



- Nodes: predicates (e.g. play, win, lose)
- Edges: entailment relations
- For multiple type pairs, e.g. ORG-ORG for sports teams

Entailment Graphs



- Nodes: predicates (e.g. play, win, lose)
- Edges: entailment relations
- For multiple type pairs, e.g. ORG-ORG for sports teams
- Learned from large corpora of multi-source news text
  - Authors use different language to describe the *same* event
- **Unsupervised method** of Hosseini et al. (2018)

Learning Entailment Relations

- Learning signal: Distributional Inclusion Hypothesis (Dagan et al., 1999; Geffet and Dagan, 2005):

  A predicate ***p*** entails another predicate ***q*** if for any context in which ***p*** can be used, ***q*** may be used in its place

Example: co-occurrences with argument pairs

	Arsenal-Man U.	Arsenal-Chelsea	Chelsea-Spurs	...
win	2	1	0	...
play	3	3	2	...

- Distribution of **win** is *included* in distribution of **play**
- Compute similarity between **win** and **play**

Challenge: Spurious Entailments

- This can fail for some highly correlated, contradictory relations: win $\rightarrow$ lose etc.

Example: co-occurrences with argument pairs

	Arsenal-Man U.	Arsenal-Chelsea	Chelsea-Spurs	...
lose	2	1	1	...
win	2	1	0	...
play	3	3	2	...

- How can we avoid learning spurious entailment relations?

Challenge: Spurious Entailments

- This can fail for some highly correlated, contradictory relations: win $\rightarrow$ lose etc.

Example: co-occurrences with argument pairs

	Arsenal-Man U.	Arsenal-Chelsea	Chelsea-Spurs	...
lose	2	1	1	...
win	2	1	0	...
play	3	3	2	...

- How can we avoid learning spurious entailment relations?

1. Temporality: Compare eventualities that happen *at the same time*

Challenge: Spurious Entailments

- This can fail for some highly correlated, contradictory relations: win $\rightarrow$ lose etc.

Example: co-occurrences with argument pairs

	Arsenal-Man U.	Arsenal-Chelsea	Chelsea-Spurs	...
lose	2	1	1	...
win	2	1	0	...
play	3	3	2	...

- How can we avoid learning spurious entailment relations?

1. **Temporality**: Compare eventualities that happen *at the same time*
2. **Modality**: Exclude eventualities that are *uncertain* to happen

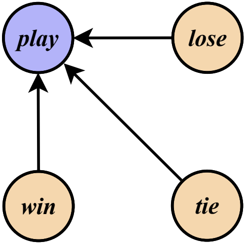
Temporality

Overview: Temporality

- **Problem:** observed *spurious entailments* between disjunctive outcomes e.g. win $\rightarrow$ lose
- **Approach:** use information about when events took place to refine the learning process
- Initial experiments focused on sports domain
- Later experiments applied the technique to the general domain

Play-win-lose-tie Scenario

Arsenal - *played* and *lost against* - **Man United** 1-3 (25/01/2019)
Arsenal - *played* and *beat* - **Man United** 2-0 (10/03/2018)
Arsenal - *played* and *tied with* - **Man United** 1-1 (30/09/2019)



Aim: Learn entailments: win/lose $\rightarrow$ play
Avoid learning *spurious* entailments: win $\rightarrow$ lose

Guillou et al. (2020)

Adding Temporal Information

- Extract binary relations with eventuality **start/end times**:

arg1 -	*predicate*	- **arg2**	time interval
Arsenal -	*tied with*	- **Man United**	(30/09/19, 30/09/2019)

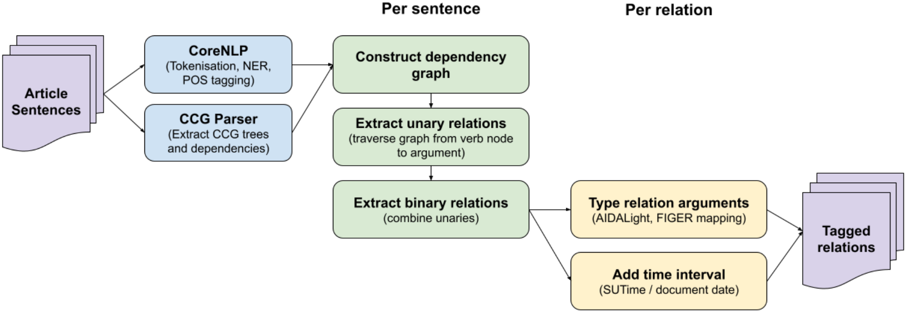
- Two temporal information sources:
  - Document creation date
  - Automatically resolve temporal expressions in the text, e.g.

Manchester United vs. Arsenal | 30th September 2019

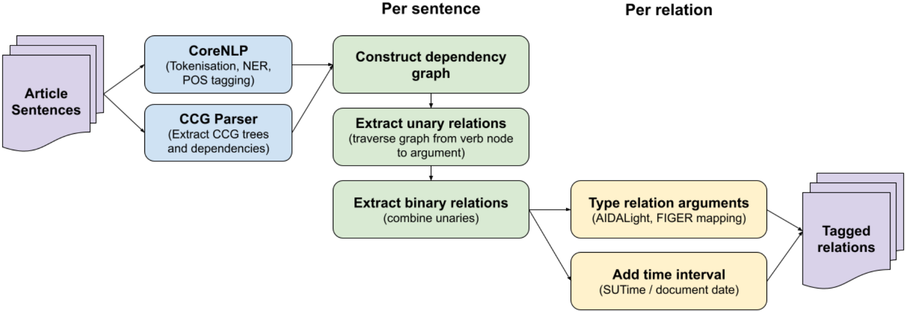
Manchester United and Arsenal played to a 1-1 draw in a sloppy, rain-soaked match at Old Trafford **on Monday** night.

Guillou et al. (2020)

Relation Extraction (with MoNTEE)



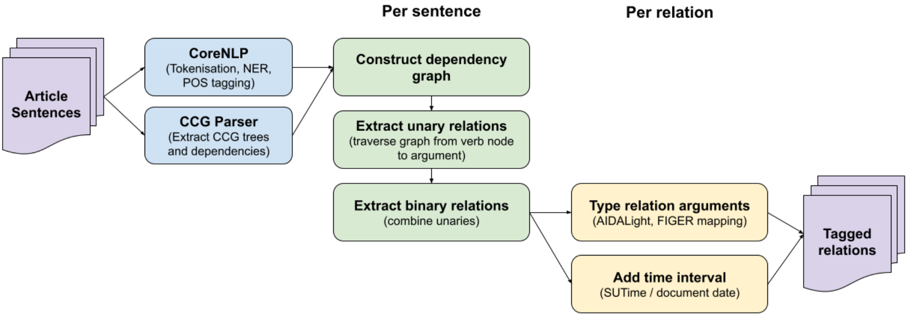
Relation Extraction (with MoNTEE)



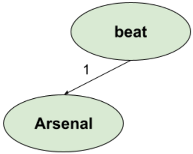
Arsenal beat Man United last night.

Bijl de Vroe et al. (2021)

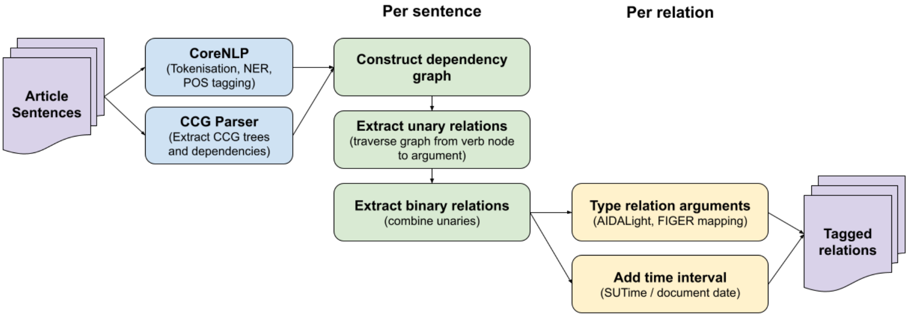
Relation Extraction (with MoNTEE)



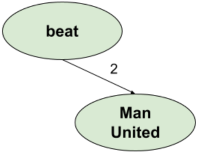
Arsenal beat Man United last night.



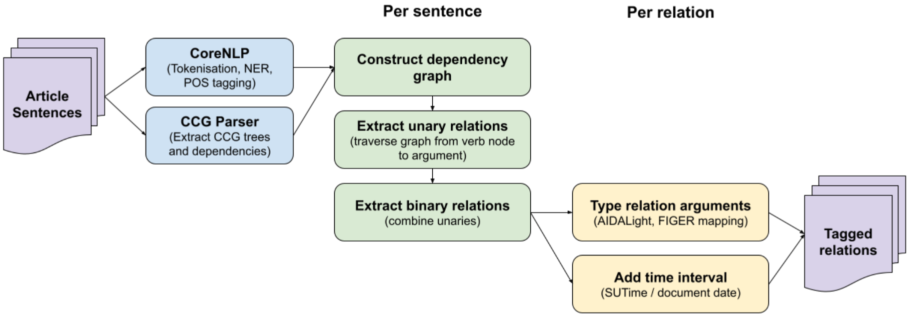
Relation Extraction (with MoNTEE)



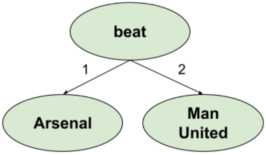
Arsenal beat Man United last night.



Relation Extraction (with MoNTEE)

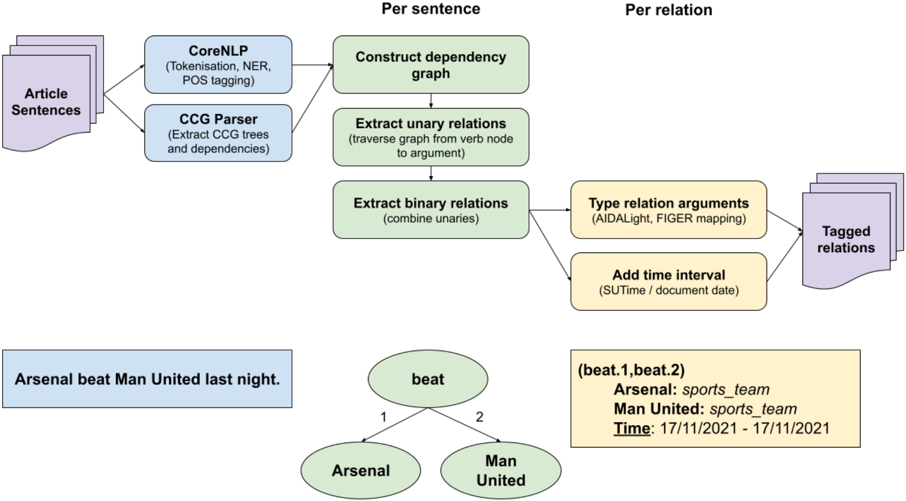


Arsenal beat Man United last night.



Bijl de Vroe et al. (2021)

Relation Extraction (with MoNTEE)



Bijl de Vroe et al. (2021)

Temporal Filtering

- **Filter out** co-occurrence counts where there is **no temporal overlap** between events

Filtering Algorithm

```
for a in argPairs:
  for p in predicates:
    for q in predicates:
      count_{p,q} += co-occur(a,p,q);
      filteredCount_{p,q} += temporal_overlap(a,p,q);
```

co-occur: count of predicate ***p***, given that ***q*** also occurs with argPair ***a***

temporal_overlap: number of events of ***p*** that overlap with any event of ***q*** (for argPair ***a***)

Guillou et al. (2020)

Worked Example

Two matches between **Arsenal** and **Man United**:

Arsenal *played against* and *beat* **Man United** (10/03/2018)

Arsenal *played against* and *lost to* **Man United** (25/01/2019)

predicate	**count**
play against	2
beat	1
lose to	1

Guillou et al. (2020)

Worked Example

Two matches between **Arsenal** and **Man United**:

Arsenal *played against* and *beat* **Man United** (10/03/2018)

Arsenal *played against* and *lost to* **Man United** (25/01/2019)

predicate	**count**
play against	2
beat	1
lose to	1

filter ⇒

	count	
entailment pair	**regular**	**filtered**
beat → play against	1	1
lose to → play against	1	1
beat → lose to	1	0
lose to → beat	1	0

Guillou et al. (2020)

Temporal Similarity Measures

- *Similarity measures* determine whether predicates in the graph entail each other
- Temporal measures inspired by BInc (Szpektor and Dagan, 2008)
  - Directional component: Weed's precision
  - Symmetrical component: Lin's similarity

Measure	Directional	Symmetrical
T. BInc Binary	✓	✓
T. BInc Ratio	✓	✓
T. Weed's Precision	✓	✗

- Baseline: BInc (atemporal)

Evaluation

- Previous work evaluated on Levy/Holt (Levy and Dagan, 2016; Holt, 2018)
  - Does not test for antonymous non-entailments e.g. win $\not\rightarrow$ lose
  - Poorly balanced: many paraphrases, few directional examples
- Semi-automatically constructed two new datasets:
  - Sports - sports domain
  - ANT - general domain

Sports Entailment Dataset

- Manually construct paraphrase clusters: win, lose, tie, play from predicates in the training data
- Automatically construct entailment pairs according to patterns:

	win	lose	tie	play
win	1	0	0	1
lose	0	1	0	1
tie	0	0	1	1
play	0	0	0	1

1 = entailment

0 = non-entailment

Category	Examples	Size
directional entailment 1	defeat $\rightarrow$ face	272
antonym 0	beat $\not\rightarrow$ fall to	446
directional non-entailment 0	play $\not\rightarrow$ win	272
paraphrase 1	defeat $\leftrightarrow$ outplay	322
		1,312

Guillou et al. (2020)

Ant

- Extract *antonymous predicate pairs* and *synonyms* from **Wordnet**
- For each antonym pair (A1, A2), identify a set of predicates (E) entailed by all elements in U(A1, A2)
- Automatically construct entailment pairs according to patterns:

	A1	A2	E
A1	1	0	1
A2	0	1	1
E	0	0	1

1 = entailment

0 = non-entailment

Category	Examples	Size
directional entailment 1	acquitted $\rightarrow$ accused	1,465
antonym 0	acquitted $\not\rightarrow$ convicted	1,800
directional non-entailment 0	accused $\not\rightarrow$ convicted	1,465
paraphrase 1	acquitted $\leftrightarrow$ absolved	1,570
		6,300

Experiments

Data:

- NewsSpike: multi-source news corpus, 0.5M articles, spanning ~6 weeks (Zhang and Weld, 2013)
- Extract relation triples. Approx. 19% time-stamped with SUTime

Experiments:

1. Sports: Temporal info source: doc date / time expressions / both
2. Sports: Add a *uniform* temporal window: N days
3. General: Add a *dynamic* per-predicate window with TacoLM (Zhou et al., 2020)

Evaluation:

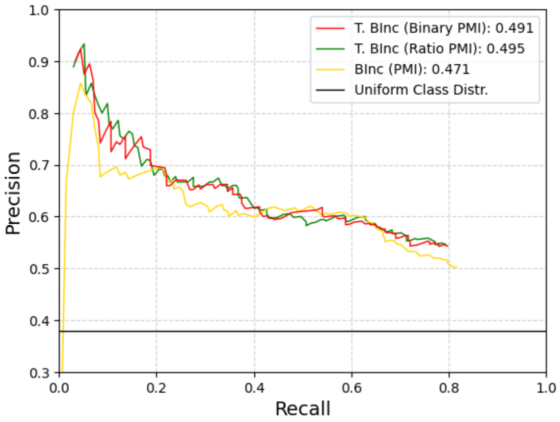
- Compare using **AUC** score: area under precision-recall curve
- Points on the curve = different entailment score thresholds

Results: Temporal Information Source

Similarity measure	timexOnly		docDateOnly	timexAndDocDate
	rec < 0.1	< 0.75	< 0.75	< 0.75
BInc	**0.072**	**0.471**	0.471	0.471
T. BInc Ratio (PMI)	0.051	0.051	**0.493**	**0.495**
T. BInc Binary (PMI)	0.058	0.081	0.489	0.491
Weed's Pr (Count)	0.061	0.440	0.440	0.440
T. Weed's Pr (Count)	0.067	0.120	0.449	0.455

- **Sports** subset: BASE (directional entailment 1 + antonym 0)
- Uniform temporal window size: 5 days
- r = recall threshold reached by all similarity measures

Exp 1: Sports BASE Subset

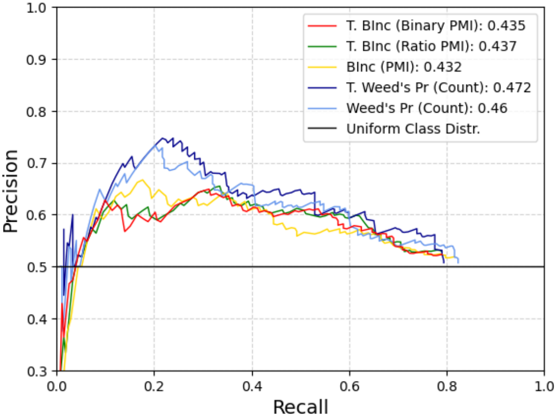


Settings: timexAndDocDate, 5 day window, evaluate on BASE subset

BASE: directional entailment 1 + antonym 0

Conclusion: Temporal filtering is **beneficial** in separating out events

Exp 1: Sports Directional Subset

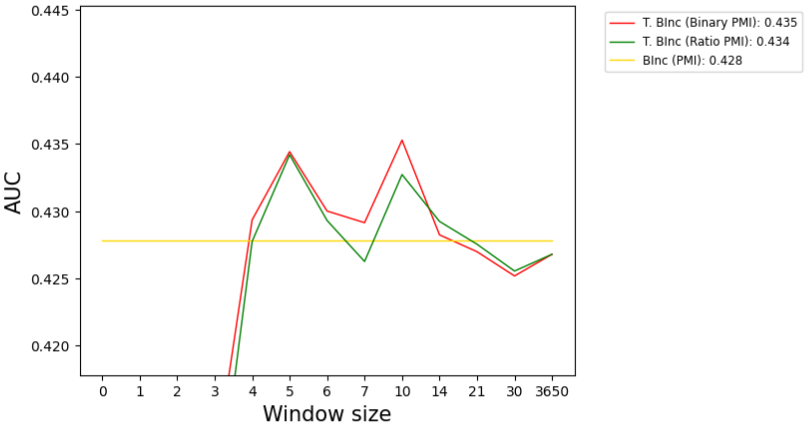


Settings: timexAndDocDate, 5 day window, evaluate on Directional

Directional: dir. entailment 1 + dir. non-entailment 0

Conclusion: Temporal info **helps** us learn directional entailments

Exp 2: Uniform Temporal Window Size



Settings: timexAndDocDate, evaluate on Sports BASE subset

Conclusion: Window size is important

Question: Why *two peaks* for each temporal similarity measure?

Exp 3: Dynamic Temporal Window

Evaluate on: ANT dataset

	ANT Base		ANT Directional	
Window Method	Uni.	Dyn.	Uni.	Dyn.
Similarity measures:				
Weed's Pr (Count)	**0.181**	**0.181**	**0.199**	**0.199**
T. Weed's Pr (Count)	0.164	0.180	0.177	0.198
BInc (PMI)	0.161	0.161	0.178	0.178
T. BInc (Ratio PMI)	0.144	0.161	0.157	0.178

Conclusions:

- Adding a **dynamic per-predicate window doesn't help**, but brings performance in line with the atemporal method
- The **atemporal formulation of the DIH** is appropriate for the general domain

Analysis

- Effect of temporal filtering is greater for the sports domain (than the general domain):
    - Antonym pairs are a) **observed** and b) **temporally disjoint** more often in Sports
    - Some areas in the general domain (e.g. legal news) could benefit from temporal filtering

- SUTime is not enough: limited number of time expressions + partial time information

- Speculation about events:
    - **Conditionals** (e.g. "If Arsenal win")
    - **Modals** ("I still expect Arsenal...")
    - Incorrect future **predictions** ("Arsenal will win")
    - **Counterfactuals** ("Had Arsenal won,...")

  is especially common in the sports domain
  and can result in conflicting evidence e.g. if Arsenal actually *lost*

Conclusions and Future Work

- Results (Exp 1) are promising, but we rely heavily on document creation date - temporal expressions are sparse
  ➥ We need an accurate way to temporally locate all eventualities
- Essential to add a temporal window around time intervals (Exp 2)
- Adding temporality is beneficial in the Sports domain (Exp 3)
  - Especially for directional entailments
  ➥ Reinterpret the DIH to include time
- The atemporal formulation of the DIH is appropriate for the general domain (Exp 3)

Modality

Linguistic Modality

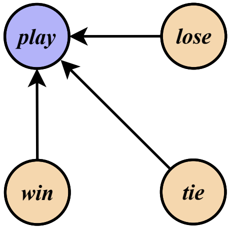
Category	Example
Modal operator	Protesters **may** have attacked the police
Conditional	**If** protesters attack the police...
Counterfactual	**Had** protesters attacked the police...
Propositional attitude	Journalists **said** that protesters attacked the police

Essential for downstream tasks: Question Answering and Knowledge Graph population

$\rightarrow$ Also useful for Entailment Graph Learning?

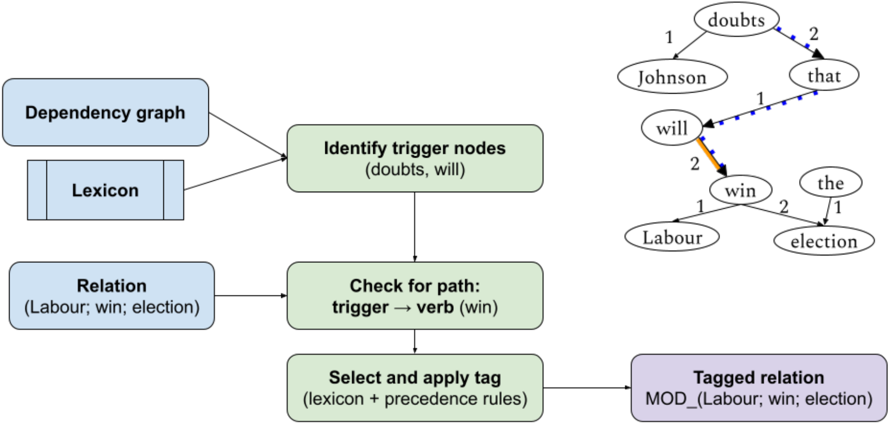
Method

- Learn entailment graphs from different training sets:
  - Only predications *asserted* as actually happening?
  - A mixture of *asserted* and *modalised* predications?
- Extract binary relations using MoNTEE (Bijl de Vroe et al., 2021)
  - Binary relations: **arg1**-predicate-**arg2** e.g. **Spurs**-beat-**Arsenal**
  - Tag binary relations as: modal operator, conditional, counterfactual, propositional attitude



Guillou et al. (2021)

MoNTEE: Modality Tagging



Lexicon

530 entries composed from:

- Modality Lexicon (Baker et al., 2010)
- Reporting verbs (Fay, 1990)
- Conditionals (Somasundaran et al., 2007)
- Conflicting event outcomes (Guillou et al., 2020)
- WordNet synonyms / antonyms (Miller, 1995)

Lemma	Category	POS-tag
shall	MOD	MD
conceivably	MOD	RB
impossible	MOD	JJ
as long as	COND	RB
reckon	ATT_THINK	VB
	...	

- **Data:** NewsSpike, approx. 0.5M articles (Zhang and Weld, 2013)
- **Models:**

	% Data	Modalised predications present?
ASSERTED	85	✗
BASELINELARGE*	100	✓
BASELINESMALL	85	✓

* Equivalent to (Hosseini et al., 2018)

Evaluation

Datasets:

- Levy/Holt: general domain, 18,407 entailment pairs

medicine *kills* **disease**	$\rightarrow$	**medicine** *treats* **disease**
medicine *treats* **disease**	$\not\rightarrow$	**medicine** *kills* **disease**

- Sports Entailment Dataset: sports, 718 entailment pairs

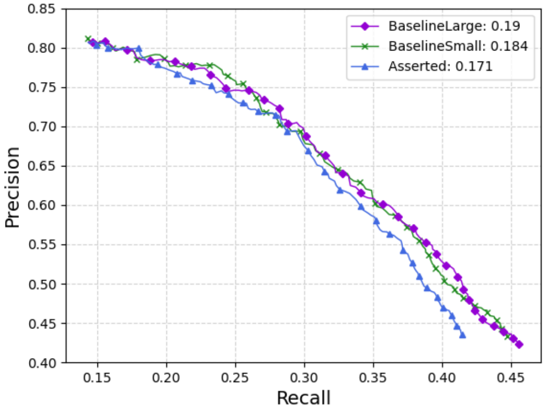
Spurs *beat* **Arsenal**	$\rightarrow$	**Spurs** *play against* **Arsenal**
Spurs *beat* **Arsenal**	$\not\rightarrow$	**Spurs** *lose to* **Arsenal**

Metric:

- AUC score: area under precision-recall curve
- Points on the curve = different entailment score thresholds

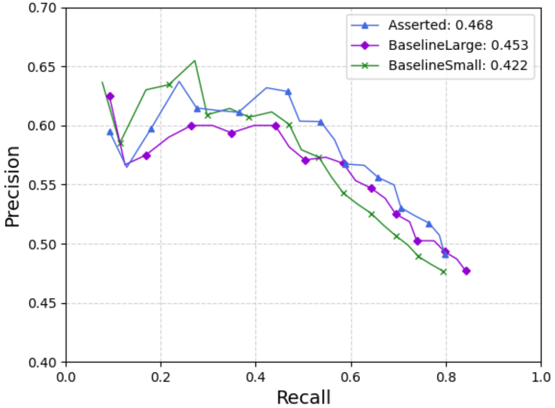
Guillou et al. (2021)

Results: Levy/Holt Dataset (Levy and Dagan, 2016; Holt, 2018)



Conclusion: **Ignoring** modality is beneficial in the general domain

Results: Sports Entailment Dataset (Guillou et al., 2020)



Conclusion: **Removing** modalised predications is beneficial in the sports domain

Guillou et al. (2021)

Analysis: Graph Comparison

Why does BaselineSmall outperform Asserted on Levy/Holt?

- Perhaps this is caused by size and/or coverage of the graph

	Nodes	Edges	% Levy/Holt predicates found	
			all examples	directional
BaselineLarge	334K	72,7M	63.06	70.29
BaselineSmall	277K	58,4M	61.13	69.29
Asserted	254K	46,3M	58.51	67.92

- However, this pattern also holds in the sports domain (where Asserted performs best)
  - ➥ Further investigation required...

Guillou et al. (2021)

Analysis: Examples

- Why is ignoring modality helpful in the general domain?
  - Perhaps modals are often used when the prior probability of the main predicate is high

Acquisition of Dell by Michael Dell

Feb 5th 2013: "...founder and CEO Michael Dell and investment firm Silver Lake Partners will buy Dell."
Feb 6th 2013: "So Michael Dell and a private equity group have bought Dell and taken it private."

- Why does removing modalised predications help in the sports domain?
  - Match outcomes are widely speculated upon, but highly uncertain

Seattle vs. Atlanta

Jan 10th 2013: "The popular opinion on this game seems to be Seattle beating Atlanta because..."
Jan 14th 2013: "Falcons come back to beat Seahawks."

Guillou et al. (2021)

Conclusions

- Overall, uncertain predications constitute a valuable learning signal for Entailment Graphs
- Removing modalised predications can help in specific domains, e.g. sports

Future Directions

- Contextualised modality / uncertainty detection
- How can we use modal information to retain what is beneficial vs. not?
  - Identify specific sub-domains
  - Retain data under different epistemic strengths e.g. "undoubtedly" vs "unlikely"
- Explore entailments between modal predicates (+temporality?)
  - if *beat* $\rightarrow$ *play*, then also *play* $\rightarrow$ *MOD_beat* (precondition)
  - if *buy* $\rightarrow$ *own*, then also *MOD_buy* $\rightarrow$ *MOD_own* (consequence)

Guillou et al. (2021)

Summary

Summary

- Experimented with:
  - Using temporal information to avoid learning spurious entailments such as win $\rightarrow$ lose
  - Ignoring modality vs. removing modalised predications
- **Conclusion:** Temporality and modality can provide a benefit in Entailment Graph learning
  - ➥ But we should pay attention to the domain
- Future Directions:
  - Contextualised modality / uncertainty detection
  - Robust temporal location of eventualities, within document and cross-document

Other Directions

Other projects within the group:

- Multivalent Entailment Graphs (McKenna et al., 2021)
- Cross-lingual Entailment Graphs (English + Chinese) (Li et al., 2022)
- Smoothing Entailment Graphs with Language Models (McKenna and Steedman, 2022)
- Incorporating Entailment Graphs for link prediction in Knowledge Graphs (Hosseini et al., 2021)

Questions?

References

K. Baker, M. Bloodgood, B. Dorr, N. W. Filardo, L. Levin, and C. Piatko. A modality lexicon and its use in automatic tagging. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 1402–1407, Valletta, Malta, May 2010. European Language Resources Association.

S. Bijl de Vroe, L. Guillou, M. Stanojević, N. McKenna, and M. Steedman. Modality and negation in event extraction. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, pages 31–42, Online, Aug. 2021. Association for Computational Linguistics.

I. Dagan, L. Lee, and F. C. N. Pereira. Similarity-based models of word cooccurrence probabilities. *Machine Learning*, 34:43–69, 1999.

R. Fay, editor. *Collins Cobuild English Grammar*. Collins, United Kingdom, 1990.

M. Geffet and I. Dagan. The distributional inclusion hypotheses and lexical entailment. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 107–114, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

L. Guillou, S. Bijl de Vroe, M. J. Hosseini, M. Johnson, and M. Steedman. Incorporating temporal information in entailment graph mining. In *Proceedings of the Graph-based Methods for Natural Language Processing (TextGraphs)*, pages 60–71, 2020.

L. Guillou, S. Bijl de Vroe, M. Johnson, and M. Steedman. Blindness to modality helps entailment graph mining. *arXiv preprint arXiv:2109.10227*, 2021.

X. Holt. Probabilistic models of relational implication. Master's thesis, Macquarie University, 2018.

M. J. Hosseini, N. Chambers, S. Reddy, X. R. Holt, S. B. Cohen, M. Johnson, and M. Steedman. Learning typed entailment graphs with global soft constraints. *Transactions of the Association for Computational Linguistics*, 6:703–717, 2018.

M. J. Hosseini, S. B. Cohen, M. Johnson, and M. Steedman. Open-domain contextual link prediction and its complementarity with entailment graphs. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2790–2802, Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics.

O. Levy and I. Dagan. Annotating relation inference in context via question answering. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 249–255, Berlin, Germany, Aug. 2016. Association for Computational Linguistics.

T. Li, S. Weber, M. J. Hosseini, L. Guillou, and M. Steedman. Cross-lingual inference with a Chinese entailment graph. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1214–1233, Dublin, Ireland, May 2022. Association for Computational Linguistics.

N. McKenna and M. Steedman. Smoothing entailment graphs with language models, 2022. URL *https://arxiv.org/abs/2208.00318*.

N. McKenna, L. Guillou, M. J. Hosseini, S. B. de Vroe, and M. Steedman. Multivalent entailment graphs for question answering. *arXiv preprint arXiv:2104.07846*, 2021.

G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

S. Somasundaran, J. Ruppenhofer, and J. Wiebe. Detecting arguing and sentiment in meetings. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 26–34, 2007.

I. Szpektor and I. Dagan. Learning entailment rules for unary templates. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 849–856, Manchester, UK, Aug. 2008. Coling 2008 Organizing Committee.

C. Zhang and D. S. Weld. Harvesting parallel news streams to generate paraphrases of event relations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1776–1786, 2013.

B. Zhou, Q. Ning, D. Khashabi, and D. Roth. Temporal common sense acquisition with minimal supervision. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7579–7589, Online, July 2020. Association for Computational Linguistics.