Temporal and Aspectual Entailment

Thomas Kober, Sander Bijl de Vroe
and Mark Steedman

University of Edinburgh

Tense, Aspect & Entailment

Entailments become more involved with Tense...

Jane has arrived in London $\rightarrow$ Jane is in London (now)

Jane will arrive in London $\not\rightarrow$ Jane is in London (now)

Outline

- Tense, Aspect & Entailment
- Tense and Aspect with Distributional Models?
- Lets have some **TEA**: A new Entailment dataset
- Temporal Entailment with Distributional Models?
- Conclusion

Outline

- **Tense, Aspect & Entailment**
- Tense and Aspect with Distributional Models?
- Lets have some **TEA**: A new Entailment dataset
- Temporal Entailment with Distributional Models?
- Conclusion

Tense, Aspect & Entailment

In entailment rule mining, entailment typically applies to *verb lemmas*

(e.g. Szpektor and Dagan, 2008, Berant et al., 2011…)

Arrive in <LOC> → be in <LOC>

Visit <LOC> → arrive in <LOC>

Visit <LOC> → leave <LOC>

Tense and Aspect typically ignored

Tense, Aspect & Entailment

However, Entailment heavily interacts with Tense & Aspect (and also Mood)

Jane has sold the house $\rightarrow$ Jane owned the house

Jane has sold the house $\nrightarrow$ Jane will own the house

Jane has bought the house $\rightarrow$ Jane owns the house

Jane has bought the house $\nrightarrow$ Jane owned the house

Tense, Aspect & Entailment

Consequences of the present perfect hold in the present

Elizabeth has gone to Meryton $\rightarrow$ Elizabeth is in Meryton

Elizabeth went to Meryton $\nrightarrow$ Elizabeth is in Meryton

Elizabeth went to Meryton $\rightarrow$ Elizabeth was in Meryton

Tense, Aspect & Entailment

Aspect also gives rise to the Imperfective Paradox (Dowty, 1986)

Mary was walking in the woods $\rightarrow$ Mary walked in the woods

Jane was reaching London $\nrightarrow$ Jane reached London

(resolved through lexical aspect types and type coercion, Moens and Steedman (1988))

Outline

- Tense, Aspect & Entailment
- **Tense and Aspect with Distributional Models?**
- Lets have some **TEA**: A new Entailment dataset
- Temporal Entailment with Distributional Models?
- Conclusion

Tense and Aspect with Distributional Models?

Test whether morphosyntactic tense is encoded in vector spaces

Two Experiments

Auxiliary-Verb Agreement: Can we detect whether an inflected verb is combined with a correct or incorrect auxiliary? (inspired by Linzen et al. (2016))

Translation Operation: Can we learn a translation operation between different tenses? (inspired by Bolukbasi et al. (2016) and Shoemark et al. (2017))

Tense and Aspect with Distributional Models?

5 models:

- Word2vec (Mikolov et al., 2013)
- fastText (Mikolov et al., 2018)
- BERT (Devlin et al., 2018)
- APTs (Weir et al., 2016)
- ELMo (Peters et al., 2018)

Pointwise addition as composition function for word2vec & fastText

Dependency based composition function for APTs

ELMo, BERT, encoder for phrases and sentences

Tense and Aspect with Distributional Models?
Auxiliary-Verb Agreement

Create representations for verb-auxiliary pairs, such as
will visit, *will visiting
is arriving, *has arriving

Simple binary classification task (correct combination vs. incorrect combination)

Extract verbs from OBWB, generate pairs, post-processing -> 36k aux-verb pairs,

Train logistic regression classifier. If it distinguishes correct from incorrect combinations, the embeddings encode some agreement information

Tense and Aspect with Distributional Models?

Auxiliary-Verb Agreement

Model	Averaged Accuracy
Majority class	0.61
Word2vec	0.71
APTs	0.92
fastText	0.71
ELMo	0.59
BERT	0.91

Tense and Aspect with Distributional Models?
Auxiliary-Verb Agreement

All models decent except ELMo:

Worse than majority class baseline. We also tried encoding full sentences instead of just auxiliary-verb phrases, but weren't able to improve performance

Word2vec and fastText:

some combinations work better than others, but overall substantially weaker than BERT and APTs

BERT and APTs:

very strong, APTs in particular benefit from sparse representation

Tense and Aspect with Distributional Models?
Translation Operation

Averaged offset: Learn offset vector for each tense:
Randomly choose 10 verbs; subtract lemma vector from inflected form vector; average the differences to obtain offset vector.
Prediction: add offset vector to unseen lemma to predict inflected verb vector
Translation matrix:
2.3k verbs per tense from OBWB, learn a neural network for each tense that given lemma vector predicts the inflected form vector
Evaluate: using MRR, the rank of the correct inflected verb in the neighbour list of the predicted verb

Tense and Aspect with Distributional Models?

Translation Operation

Results

MRR (Averaged Offset)

	w2v	APT	fT	ELMo	BERT
will	0.97	0.13	0.93	0.01	0.09
is	0.46	0.01	0.95	0.01	0.05
was	0.47	0.01	0.86	0.01	0.03
gerund	0.38	0.01	0.46	0.01	0.09
has	0.46	0.01	0.92	0.01	0.05
had	0.46	0.05	0.93	0.01	0.04
past	0.37	0.06	0.48	0.02	0.08

MRR (Neural Network)

	w2v	APT	fT	ELMo	BERT
will	1.00	0.01	1.00	0.01	0.01
is	0.83	0.01	0.75	0.01	0.01
was	0.83	0.01	0.76	0.01	0.01
gerund	0.56	0.01	0.66	0.01	0.01
has	0.84	0.01	0.76	0.01	0.01
had	0.82	0.01	0.75	0.01	0.01
past	0.60	0.01	0.69	0.01	0.01

ELMo BERT have contextualised embeddings, so they are at a disadvantage in comparison to word2vec & fasttext who have static embeddings. APTs also have static embeddings, but suffer from sparsity

Tense and Aspect with Distributional Models?

Actually works fairly well

Most models don't excel on both tasks, but morphosyntactic relations seem to be somehow encoded in each space

	Aux-Verb Agreement	Translation Operation
w2v		✅
fastText	✅	✅
ELMo	✅	
BERT	✅	
APT	✅	

Outline

- Tense, Aspect & Entailment
- Tense and Aspect with Distributional Models?
- **Lets have some TEA: A new Entailment dataset**
- Temporal Entailment with Distributional Models?
- Conclusion

TEA: A new Entailment dataset

There are a number of recently released large scale entailment datasets, e.g. SNLI (Bowman et al., 2015; Williams et al., 2017; Poliak et al., 2018; etc)

However, neither focuses on or includes a substantial proportion of examples focusing on Temporal Entailment

(Very limited number in e.g. FraCas, but nothing of larger scale)

Therefore, we created **TEA** (**T**emporal **E**ntailment **A**ssessment)

Available from here: https://github.com/tttthomasssss/iwcs2019

TEA: A new Entailment dataset

TEA contains 11,138 premise-hypothesis sentence pairs

The only differences between premise and hypothesis are the

- Main verb
- Morphosyntactic tense of the main verb
- (a preposition or particle to make the sentence grammatical)

Labelled by 2 human annotators (Thomas & Sander)

TEA: A new Entailment dataset

Candidate entailment pairs were mined from previous entailment datasets and other lexical resources such as the WordNet (Fellbaum, 1998) verb entailment graph or the before-after category of VerbOcean (Chlovski and Pantel, 2004)

Arguments were manually added to form full sentences

    The arguments also resolved *some* ambiguity

    We tried to avoid habitual readings

All sentences follow the same syntactic structure

TEA: A new Entailment dataset

Examples:

John is visiting London	John has arrived in London	1
John is visiting London	John will leave London	1
John is visiting London	John has left London	0
George is acquiring the house	George owns the house	0

Outline

- Tense, Aspect & Entailment
- Tense and Aspect with Distributional Models?
- Lets have some **TEA**: A new Entailment dataset
- **Temporal Entailment with Distributional Models?**
- Conclusion

Temporal Entailment with Distributional Models?

Evaluation:

Binary labels (entailment : no-entailment) with a class distribution 22 : 78

Average Precision (a.k.a. Area under the Precision/Recall curve), Accuracy & F1-Score

a) Run word2vec, APTs, fastText, ELMo and BERT
b) Run biLSTMs pre-trained on SNLI or DNC
c) Compare to two simple baselines
    i) Majority class
    ii) Majority class per Tense pair

Temporal Entailment with Distributional Models?

Model	Avg. Precision	Model	Avg. Precision	Accuracy	F1-Score
word2vec	0.31	biLSTM-DNC	0.22	0.58	0.49
APT	0.28	biLSTM-SNLI	0.21	0.51	0.47
fastText	0.30	Majority class	0.22	0.78	0.44
ELMo	0.21	Maj. Class / Tense pair	**0.35**	0.80	0.66
BERT	0.27				

Temporal Entailment with Distributional Models?

No model beats the majority class / tense pair baseline

All models achieve low scores - despite the simplicity of the sentences

Temporal Entailment is a difficult problem

Outline

- Tense, Aspect & Entailment
- Tense and Aspect with Distributional Models?
- Lets have some **TEA**: A new Entailment dataset
- Temporal Entailment with Distributional Models?
- **Conclusion**

Conclusion

Highlighted the importance of Tense & Aspect for Natural Language Inference

Showed that the morphosyntax of Tense & Aspect is encoded in vector space (to a reasonable degree at least)

Introduced TEA - a dataset for Temporal Entailment (https://github.com/tttthomasssss/iwcs2019/)

But model performance is bad across the board on a semantic task

Distributional model governed by contextual similarity, not ideal for doing inference that depend on Tense & Aspect

Questions/Discussion

tkober@inf.ed.ac.uk

sbdv@ed.ac.uk