Towards Computing Contextual Lexical Contrast: Cont2Lex Corpus, Recognition benchmarks, and Preliminary Analyses

Wenqiang Lei*, Yisong Miao*, Runpeng Xie,
Bonnie Webber, Meichun Liu, Tat-Seng Chua, Nancy Chen

Towards Computing <u>Contextual Lexical Contrast:</u> Cont2Lex Corpus, Recognition benchmarks, and Preliminary Analyses

1st
Half

Wenqiang Lei, Yisong Miao, Runpeng Xie, Bonnie Webber,
Meichun Liu, Tat-Seng Chua, Nancy Chen

Towards Computing Contextual Lexical Contrast: <u>Cont2Lex Corpus, Recognition benchmarks, and Preliminary Analyses</u>

2nd
Half

Wenqiang Lei, Yisong Miao, Runpeng Xie, Bonnie Webber, Meichun Liu, Tat-Seng Chua, Nancy Chen

Lexical Contrast

- Wet — Dry
- Fast — Slowly
- Catch — Escape
- City — Farm

Lexical Contrast exist between a pair word, they have some non-zero degree of binary incompatibility and/or have some non-zero difference across a dimension of meaning. (Mohammad et al. 2013)

Lexical Contrast — Meaningful

- Contradiction detection:

Kyoto has a predominantly **wet** climate / It is mostly **dry** in Kyoto

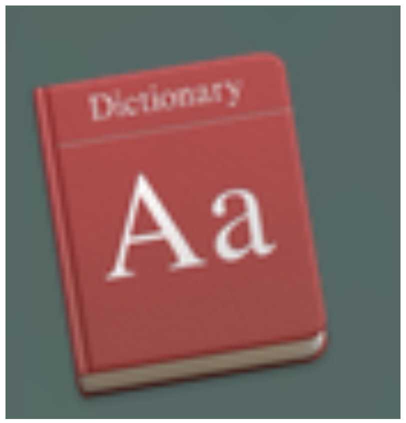
- Paraphrase generation:

The dementors **caught** Sirius Black / Black could not **escape** the dementors

- Discourse relation detection:

During December, Shenzhen is **hot**, Harbin is **cold.**

<u>Contextual</u> Lexical Contrast



What will happen if we directly look-up dictionary?

✅ • Context 1: During the industrial revolution, air quality worsened in the **<u>city</u>**, but remained the same in the **<u>farm</u>**. (being understood as contrast in the context)

❌ • Context 2: They had a packed trip: they visited the **<u>farm</u>**, the beach, the **<u>city</u>**, the zoo, and the museum over the weekend. (related, but not understood as contrast)

❌ • Context 3: After John left New York **<u>City</u>**, he attended the Indiana Small **<u>Farm</u>** Conference. (co-occur just by chance)

Contextual Lexical Contrast

Task Description:

- Input: A pair of words w+ and w– , we denote their corresponding context as c = w1 , w2 , ..., wn (note that w+ and w– are included in c and we do not specially mark their position).

- Output: A model needs to predict if contextual contrast holds between w+ and w-.

Cont2Lex Corpus

- Corpus: Wikipedia and WSJ Corpus.
- Lexicon: ConceptNet
- Limit: Max 3 times per word pair.
- Total #: 11,279

	Noun	Verb	Adj.	Adv.	Total/ Avg.
# instance	4665	2283	3579	407	11279
pr	27%	29%	44%	39%	33%

Corpus Statistics: the number of instance of each part-of-speech and the positive instance ratio (pr), e.g., there are 4,665 Noun instances and 27% are positive instances.

Recognition Benchmarks

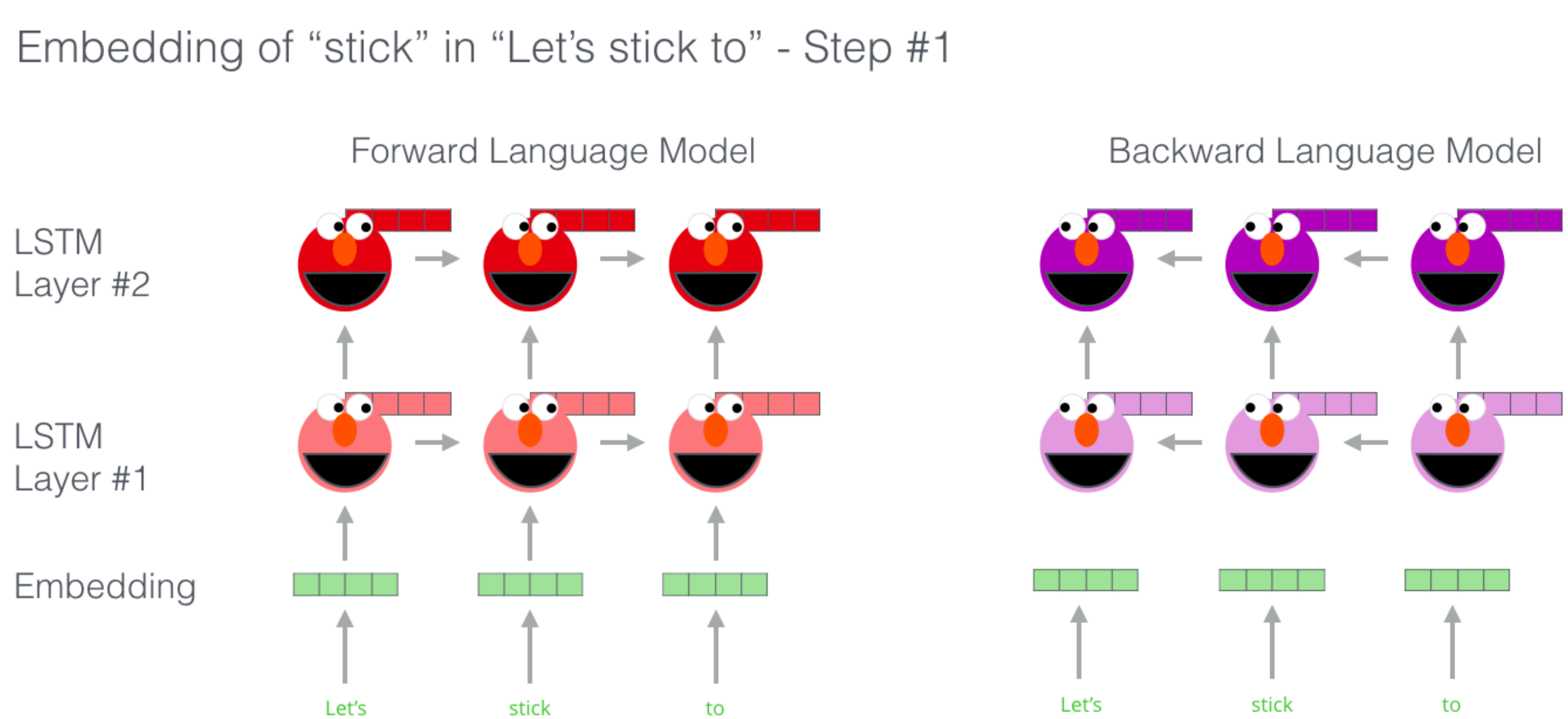
Embed & Encode Framework by Shwartz and Dagan (2019).

- Embedding:
  - Non-Contextual: Glove, Word2Vec, fastText
  - Contextual: BERT, ELMo, GPT
- Encoder:
  - None, Self-Attention, BiLSTM

Embeddings

	1	2	3	4	5	6	7	8	9	10	11
so	0.60308	-0.320240	0.088857	-0.551760	0.531820	0.047069	-0.36246	0.005702	-0.37665	0.225340	-0.13534
them	0.64642	-0.556000	0.470380	-0.820740	0.795120	0.287710	-0.56426	0.146300	-0.52421	0.021607	-0.11266
what	0.45323	0.059811	-0.105770	-0.333000	0.723590	-0.087170	-0.61053	-0.037695	-0.30945	0.218050	-0.43605
him	0.11964	-0.045405	0.051100	-0.828730	0.976650	0.111280	-0.54588	1.156100	-0.68081	0.060207	-0.28765
united	-0.39874	0.071993	-0.069773	0.147060	0.118500	0.147700	-0.84431	0.147600	0.64804	-0.559260	0.50164
during	0.29784	-0.018422	-0.718910	-0.465100	-0.456610	-0.004215	-0.74598	0.346620	-0.51781	-0.587700	0.18398
before	0.30806	-0.296650	-0.257060	-0.587100	0.095135	-0.152110	-0.91478	0.757270	-0.30423	-0.290580	-0.13034
may	0.70480	0.222610	0.086997	-0.212410	-0.089356	0.437420	-0.28170	0.133780	-0.50859	-0.182420	0.49506
since	0.15423	-0.125520	0.022279	-0.067561	-0.359750	0.144090	-1.09020	-0.028693	-0.43147	-0.137810	0.37841
many	0.69790	0.082340	0.041526	-0.507040	-0.158010	0.360480	-1.07450	-0.239270	-0.74704	0.160070	-0.18420

Static Embedding (e.g. Glove, Word2Vec): given a word and "look-up" its embedding.



Contextual Embedding (e.g. ELMo, BERT): given a word and its context, generate its embeddings dynamically.

Overall Model Performance

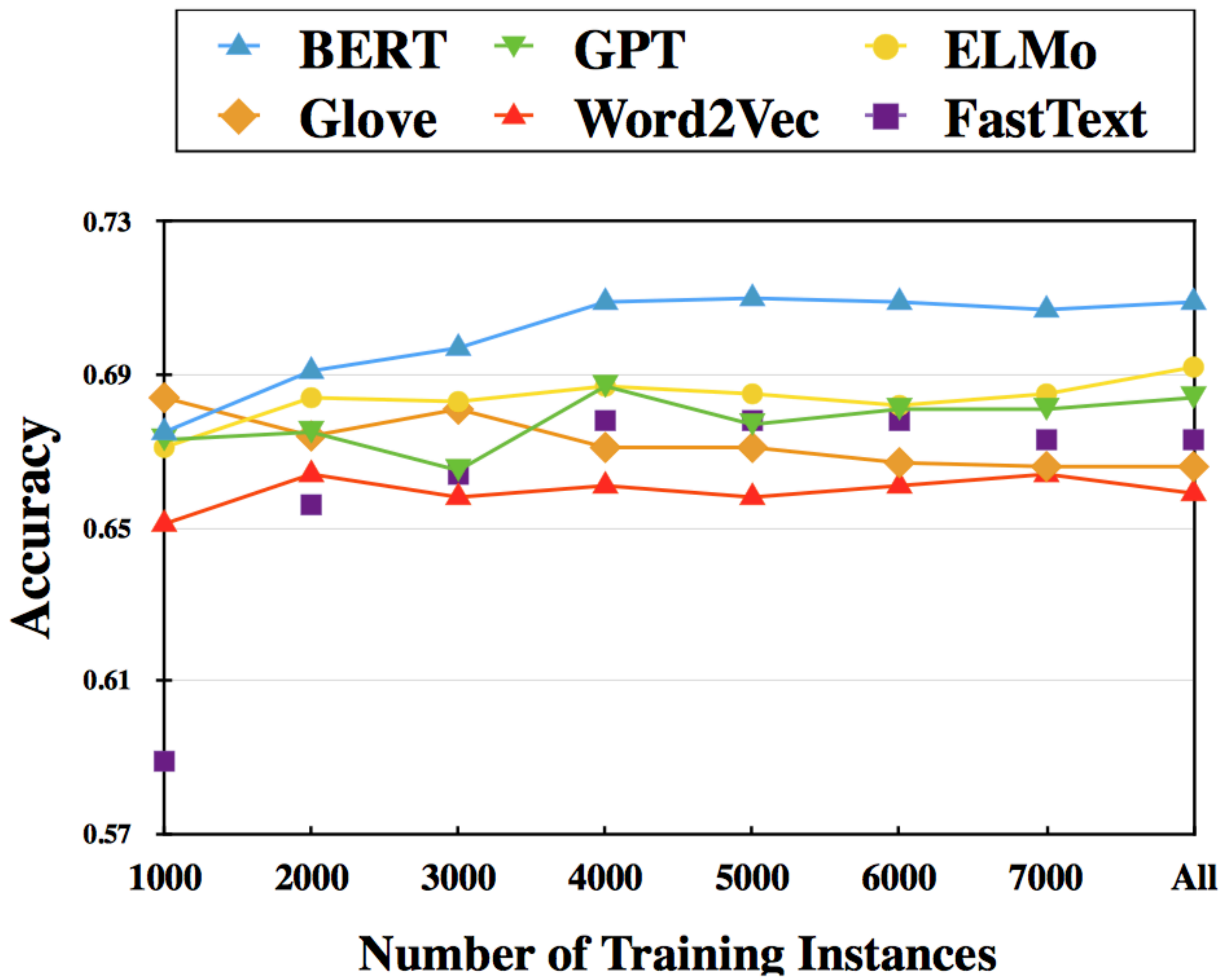
	BiLSTM	Attention	None
Glove	0.666	0.659	0.654
Word2Vec	0.659	0.656	0.651
FastText	0.673	0.665	0.658
ELMo	0.680	0.689	0.692
GPT	0.673	0.683	0.684
BERT	0.703	0.709	0.692
Majority	0.680		

Main Experiment Results: We report the accuracy score of contextual lexical contrast classification by different embedding methods and encoders.

Majority baseline: It is a weak baseline, since it always predicts the random class.

BERT: I am stronger than non-contextual ones, but I am not quite stronger than majority baseline.

Dataset Size Study



The Performance of each embedding methods with respect to the number training instance.

The dataset is sufficient for all models to converge!

Out-of-context Contrast Recognition

Embedding	Glove	Word2Vec	FastText
Acc.	0.875	0.882	0.863

Embedding	ELMo	GPT	BERT
Acc.	0.876	0.896	0.917

The accuracy score of out-of-context Contrast Recognition.
We use the original (w+, w-) as positive sample, and a randomly sampled one (w+, w-‘) as negative sample (POS constrained)

Hah! I am the strongest again! All models can perform much better in out-of-context contrast recognition!

Conclusion

- Contextual lexical contrast recognition is very challenging, even BERT barely beat the majority baseline. We call our community to pay attention to this fundamental NLP phenomenon.

- Contextual embeddings generally outperform non-contextual ones, validating their effectiveness.