Neural dialogue act recognition with transformer pre-training

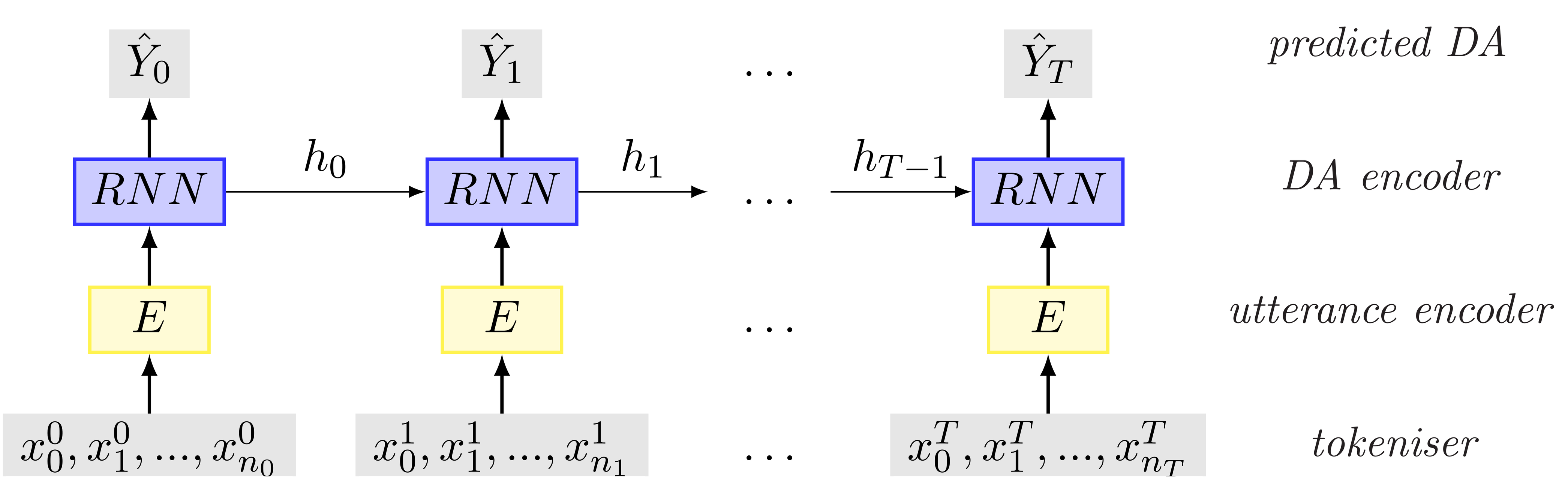
GÖTEBORGS UNIVERSITET

Bill Noble and Vladislav Maraev · Centre for linguistic theory and studies in probability (CLASP)
Department of Philisophy, Linguistics and Theory of Science · University of Gothenburg

BERT SOLVES DIALOGUE #LAUGHTER

Recently, multi-layer neural language models pre-trained on massive amounts unlabeled data have been used to provide context-sensitive word vectors and sentence-level distributional representations. One such model, BERT, uses an attention-based transformer architecture to achieve state of the art results on a variety of NLP tasks (Devlin et al., 2018). However, given that BERT is pre-trained on book and encyclopedia data, there is no guarantee it will improve performance on *dialogue-specific tasks*. To assess BERT's potential for dialogue applications, we propose a series of *dialogue act recognition* (DAR) experiments with various utterance encoders, including BERT.

Neural dialogue act recognition model



Where is laughter (or its absense) most helpful?

It might be the case that dialogue-specific features, such as discourse markers, disfluencies, and laughter are useful for DAR. We train models on Switchboard *dialogues with laughter* (**L**) and *with laughters removed* (**NL**) to test if this is the case.

		accuracy	macro-average recall
GloVe	L	.702	.198
	NL	.698	.209
BERT	L	.779	.400
	NL	.772	.408

Top 5 dialogue acts by the impact of laughter

for GloVe:	increase in accuracy
continuer (+)	.158
non-verbal (x)	.108
appreciation (ba/fe)	.101
statement-opinion (sv/fx)	.092
conventional-closing (fc)	.067

for BERT [CLS]:	increase in accuracy
or-clause (qrr)	.488
open-question (qo)	.170
statement-opinion (sv/fx)	.163
response-acknowledgement (bk)	.124
rhetorical-questions (qh)	.122

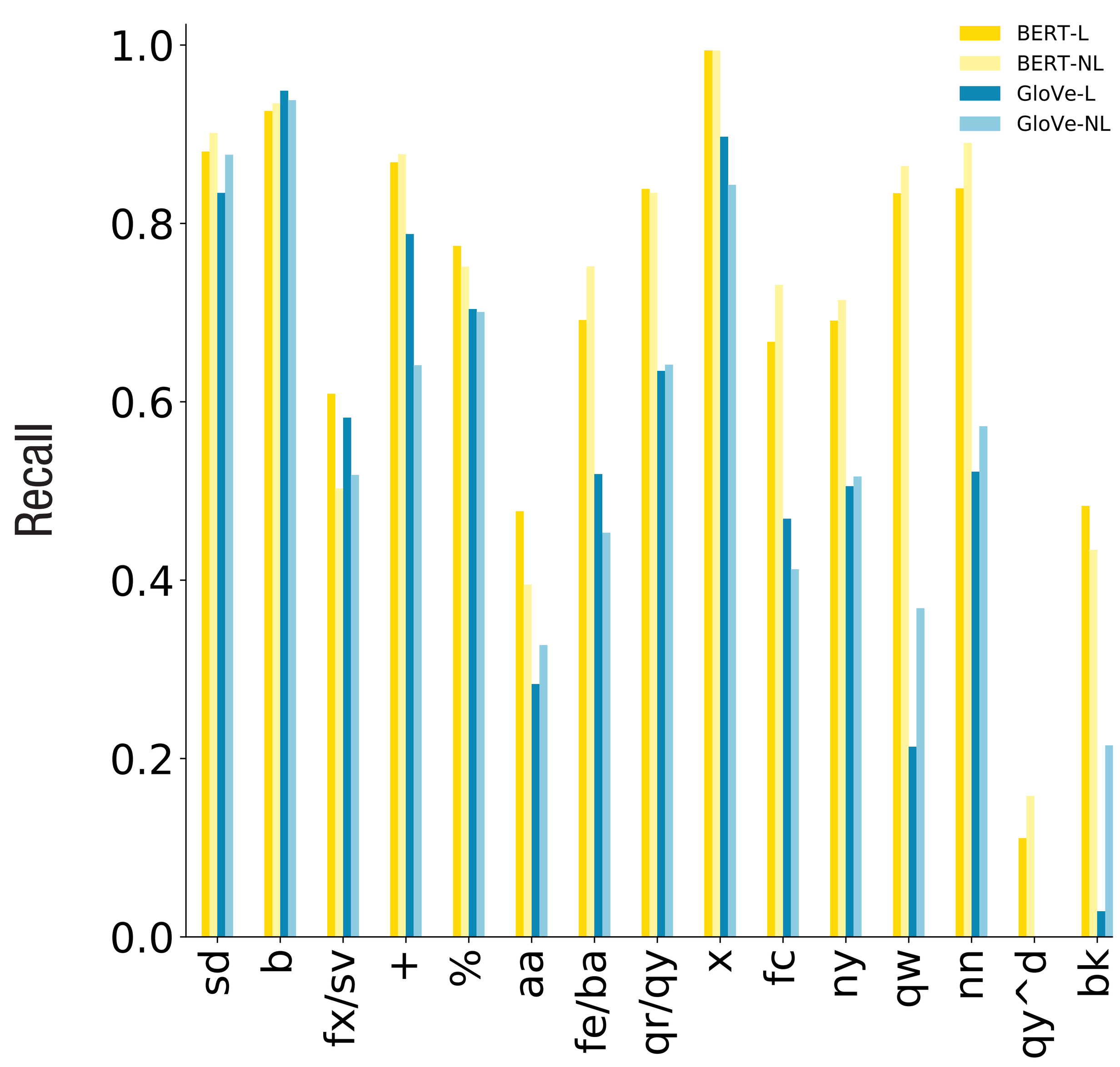
Switchboard dialogue act tags

Dialogue acts represent the meaning of an utterance by the speech act it carries out (Austin and Urmson, 2009). **Dialogue act recognition (DAR)** is the task of automatically labeling utterances with tags from a dialogue act schema such as DAMSL (Core and Allen, 1997).

Classes: 44 (including continuers and padding)
Dialogues: 808 / 115 / 232 (train/val/test)
Utterances: 156441 / 20825 / 44350 (train/val/test)

RNN for representing discourse

Many DAR strategies attempt to model discourse context in addition to the content of the utterance in question. Stolcke et al. (2000), for example, use a HMM to tag dialogue acts. The hidden state of neural sequence models can also represent discourse context (e.g. Kalchbrenner and Blunsom, 2013; Tran et al., 2017).



		prev.	accuracy on utterances containing laughter	next
GloVe	L	.703	**.703**	.650
	NL	.689	**.679**	.643
BERT	L	.768	**.768**	.761
	NL	.752	**.759**	.737

References

John L. Austin and James O. Urmson. 2009. How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955, 2. ed., [repr.] edition. Harvard Univ. Press.

Mark G Core and James F Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. In Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines: 28–35.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent Convolutional Neural Networks for Discourse Compositionality. In Proceedings of the Workshop on Continuous Vector Space Models and Their Compositionality, pages 119–126.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. 26(3):339–373.

Quan Hung Tran, Ingrid Zukerman, and Gholamreza Haffari. 2017. Preserving Distributional Information in Dialogue Act Classification. In Proceedings of EMNLP-2017, pages 2151–2156.

CLASP centre for linguistic theory and studies in probability

The work was supported by a grant from the Swedish Research Council (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.