A Appendix

A.1 Related work

We supplement some work that is very different from classic classification models.

One line of work is based on label-word joint embedding, such as LEAM (Wang et al., 2018), MTLE (Zhang et al., 2017) and EXAM (Du et al., 2019). It poses extral requirements on label information. However in our case, many class labels are close to each other, and not well-defined. Though we can present the class label with one of user's utterances from that class, the label will become not only prolix but also biased.

Another line of interesting work is the joint training of sentence classification and NER, such as (Kruengkrai et al., 2020; Zhang et al., 2020; Hakkani-Tür et al., 2016; Liu and Lane, 2016; Goo et al., 2018). For example, If both the city and date information are recognized, then they are good indicators that this query might be booking an ticket.

Relation to the label embedding based LEAM

LEAM shows its superiority over TextCNN in all datasets in our experiments, by modeling labeling information.

LEAM does not perform as well as SFCs, since the first intuitive reason is it does not use a pretrained model as its input encoding module; another critical reason is in task-specific chat applications, it is common to have many similar intents, and it becomes hard and even impossible to name each intent with a short clear name to feed into LEAM model. One candidate solution is setting a most standard sample as the label of the class. When using non-pretrained models base LEAM, this is applicable. Yet when using a pretrained model based, as all labels can be hundreds of thousands, then the training can not be accommodated in a poplar 32 G Tesla V100 GPU.

Actually, joint SFC can be kind of understood as a generalization form of LEAM. When there is no clear class labels, the relationship between a sample and a class label is implicitly encoded as that between a sample and another sample from the same class, and this turns into a sentence pair similarity model.