Improving Intent Detection Accuracy Through Token Level Labeling

Michał Lew, SentiOne
Aleksander Obuchowski, SentiOne, Gdańsk University of Technology
Monika Kutyła, SentiOne

Project funded by the EU Research Grant

European Funds Smart Growth · Republic of Poland · NCBR National Centre for Research and Development · European Union European Regional Development Fund

Abstract

Intent detection is traditionally modeled as a sequence classification task where the role of the models is to map the users' utterances to their class. In this paper, however, we show that the classification accuracy can be improved with the use of token level intent annotations and introducing new annotation guidelines for labeling sentences in the intent detection task. What is more, we introduce a method for training the network to predict joint sentence level and token level annotations. We also test the effects of different annotation schemes (BIO, binary, sentence intent) on the model's accuracy.

Dataset

- For the purpose of evaluation of our method we created our own dataset of computer mediated customer-agent helpline conversations in the banking domain.
- This dataset contains real human-human conversations of customers with customer service agents on Facebook's Messenger in the Polish language.
- The list of labels and their respected number of examples is shown in the table

Intent	train	test
300	26	7
unblocking access	97	24
deposit machine fee	29	7
double charge	34	9
payment confirmation	22	6
canceling an application	30	8
application malfunction	18	5
trusted profile	10	3
card malfunction	69	17
contact request	17	4
server malfunction	26	6
sessions	21	5
sms	31	8
application status	29	7
cdm funds posting	23	6
application processing time	32	8
cash withdrawal	12	3
IBAN/BIC/SWIFT	35	9
blocking card documents	21	5
helpline waiting time	27	7
change of personal data	42	10
card delivery time	25	6
change of phone number	49	12
thanks	14	3
Sum	**739**	**185**

Annotation

- Each statement is assigned exactly one intention e.g. *how long do I have to wait for the application?* [application_processing_time]
- The chosen intention concerns the main topic of the conversation
- The scope of a tag covers the part of the statement that is specific to the intention. If the statement is complex and the client describes the reason for making contact in a few sentences then, unless otherwise impossible, the sentences were annotated in a way that helped to indicate the intentions in their context, e.g. *Hello, I would like to order an activation package. I created an account, I received an activation package via text, valid for 48 hours, but I was not able to activate it within 48 hours, hence the need to receive a new activation package. How can I order it?* [unlocking_access]

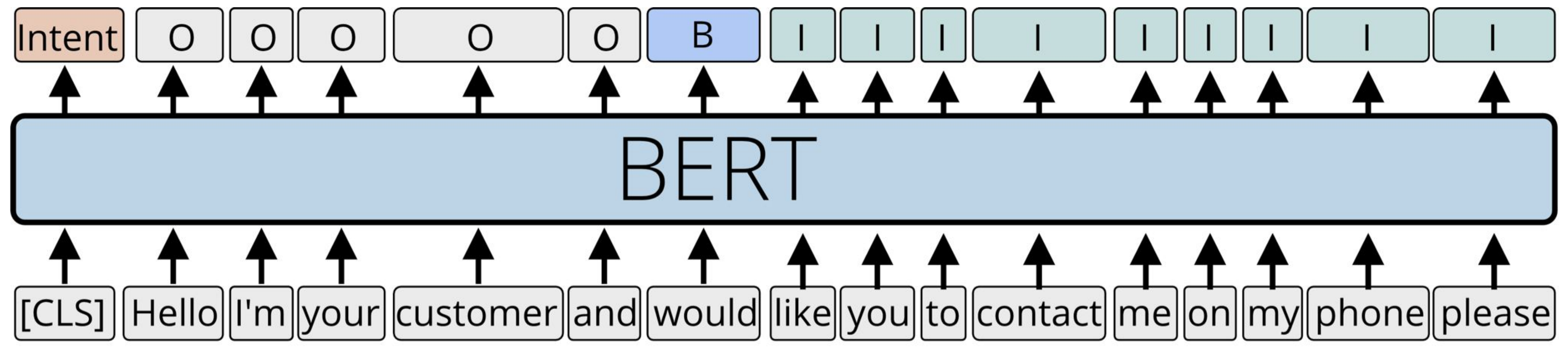
Can I cancel my application? Unfortunately, the examination is taking too long and I cannot wait this long.
canceling_an_application

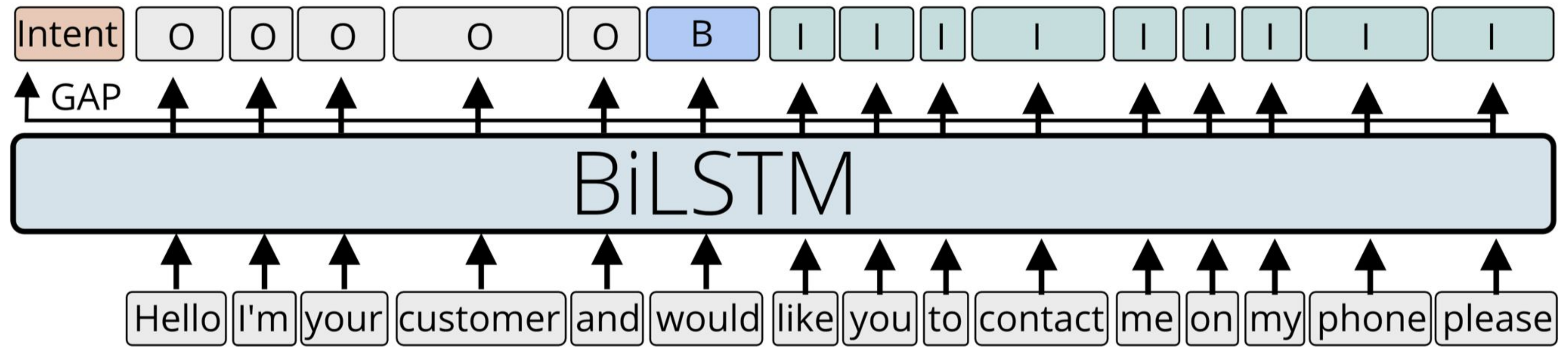
Well, **my card has been rejected** 2 times when paying contactless and 2 times during a transaction with a card reader while using the correct pin
card_malfunction

Hello, I'm your customer and **would like you to contact me on my phone please**
Contact_request

Hello. **I deposited money into the cash deposit machine** because I have to make an urgent transfer. **When will the money be on my account?**
Cdm_funds_posting
Cdm_funds_posting

Models





- For BERT implementation, we chose the base multilingual model. In our experiments we fine-tuned the model for both sequence labeling and the classification task. During the training, each token was labeled in a corresponding format. We also used BERT's special [CLS] token for labeling the entire sentence. Token level embeddings were mapped to their labels using a fully connected layer with softmax activation function.
- For BiLSTM network we used Word2Vec embeddings pre-trained on the NKJP corpus. These inputs were inputted into the bidirectional LSTM layer with a hidden state size of 300 neurons. Subsequently, for the token level classification we used a fully-connected layer with a softmax activation function. The sentence level labels were predicted based on LSTM cells output pooled with global average pooling, on top of which another fully connected layer with softmax activation function has been added.

Training

- Both networks were trained using categorical cross entropy loss function. This loss was calculated between predicted token-level predictions and their true labels, as well as between sentence level intent prediction and its true intent.
- The loss function is shown in the equation, where T is the number of tokens in the sentence, C_t is number of token classes dependant on the annotation style, C_s is the number of intents, t_i, p_j represent the correct token level class and prediction, and t_k and p_k represent sentence leave prediction and true class.

$$L = -\sum_{i}^{T}\sum_{j}^{C_t} t_i log(p_j) - \sum_{k}^{C_s} t_k log(p_k)$$

Results

- no token labeled - not using token level annotations
- all tokens labeled - using sentence intent as label for all the tokens
- binary labels - tokens labeled as either relevant or irrelevant to the sentence intention
- BIO labels - tokens labeled with the BIO scheme
- intent labels - tokens relevant to the sentence intention labels with its intent
- We also compared our solution with baseline Support Vector Machines (SVM) model trained on the whole sentences without additional token labels.

Annotation scheme	BERT	BiLSTM	SVM
no tokens labeled	0.918	0.859	0.837
all tokens labeled	0.913	0.859	-
binary labels	0.918	0.864	-
BIO labels	0.929	0.864	-
intent labels	0.929	0.875	-

Future work

- Testing the influence of token level intent annotation on the accuracy of joint intent detection and slot filling models
- More sophisticated annotation scheme - action-object model