Speech Function Classification

Task Description	Dialogue management is a challenging task in Conversational AI, especially when it comes to *casual conversation* (aka chit-chat). Casual conversations are not motivated by a clear pragmatic purpose. They are informal, can have humor, and be lengthy in most cases. One state-of-the-art approach to manage dialogues is based on the use of speech act classification. **Speech Acts** work at the utterance level: hearer interprets speaker's intentions. Besides that, there is another classification type that considers **Speech Functions** that are similar to Speech Acts, but they produce utterance through its role in discourse: speaker shows his intentions in a dialogue. The taxonomy of speech function classification by Eggins& Slade was taken as a basis. It consists of more than 40 classes, but you'll be dealing with only 10 high-level classes.
Data	As data, we used dialogues representing face-to-face conversations, which were annotated manually. Each of the dialogues consists of approximately 500 moves. Data includes text, speaker, and label for each move. https://raw.githubusercontent.com/deepmipt/interns2021_sfc/main/test_data.csv https://raw.githubusercontent.com/deepmipt/interns2021_sfc/main/train_data.csv
Challenges	The number of annotated data is limited. Moreover, this is a huge data imbalance because some of the classes are less used in casual conversations.
Task	- to classify moves in dialogues using the DeepPavlov model for sequence tagging. Instead of word embeddings should be used sentence ones. - to make predictions of classes for next moves in dialogues considering statistics