DeepPavlov: An Open Source Library for Conversational AI

Mikhail Burtsev **Alexander Seliverstov** **Rafael Airapetyan** **Mikhail Arkhipov**

Dilyara Baymurzina **Eugene Botvinovsky** **Nickolay Bushkov** **Olga Gureenkova**

Andrey Kamenev **Vasily Konovalov** **Yuri Kuratov** **Denis Kuznetsov**

Alexey Litinsky **Alexey Lymar** **Maxim Petrov** **Leonid Pugachev** **Alexey Sorokin**

Maria Vikhreva

Moscow Institute of Physics and Technology
9 Institutskiy per., Dolgoprudny, Russian Federation
`burtcev.ms@mipt.ru`

Abstract

An open-source DeepPavlov library is specifically tailored for development of dialogue systems. The library prioritizes efficiency, modularity, and extensibility with the goal to make it easier to create dialogue systems from scratch with limited data available. It supports modular as well as end-to-end approaches to implementation of conversational agents. In DeepPavlov framework an agent consists of skills and every skill can be decomposed into components. Components are usually trainable models which solve typical NLP tasks such as intent classification, named entity recognition, sentiment analysis or pre-trained encoders for word or sentence level embeddings. Sequence-to-sequence chit-chat, question answering or task-oriented skills can be assembled from components provided in the library. ML models implemented in DeepPavlov have performance on par with current state of the art in the field [1].

1 Introduction

Textual and voice communication are becoming more and more widespread. As a result, interest in technology for conversational interfaces has grown significantly over the last years. At the same time, current progress in application of deep neural networks to solve natural language processing tasks demonstrates high potential for improvement of existing methods as well as for emergence of new approaches to development of conversational agents.

Today, complex dialogue systems organize different functional blocks into conversational skills and skills into a dialogue agent. Such agents usually have a hybrid architecture that combines rule-based parts with trainable ML models. Active research and development in the field produce rich code base of such components. This makes development easier, but high diversity of frameworks and implementations results in challenge to integrate all parts together.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

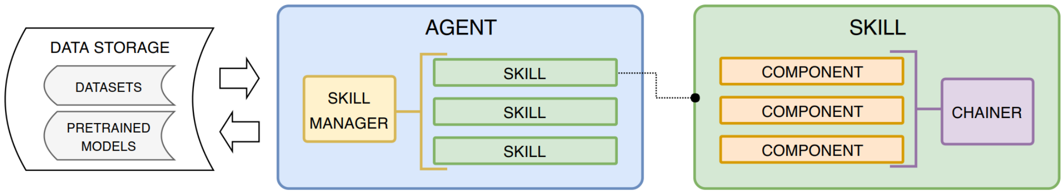

Figure 1: High level architecture of DeepPavlov library.

DeepPavlov combines a library of building blocks for NLP functions with a framework for assembling dialog agents. This makes possible to address a problem with current fragmentation of the code base for conversational systems. It also allows implementation of wide range of dialogue systems from task-oriented to chit-chat in a unified way.

2 DeepPavlov Overview

Two major applications areas of DeepPavlov library are:

- development of production-ready chatbots and complex conversational systems;
- NLP and dialog systems research.

The goal is to enable AI-application developers and researchers with:

- a set of pre-trained NLP models, pre-defined dialogue system components (ML/DL/Rule-based) and pipeline templates;
- a framework for implementing and testing their own dialogue models;
- tools for integration of applications with adjacent infrastructure (messengers, helpdesk software etc.);
- a benchmarking environment for conversational models and uniform access to relevant datasets.

High level architecture of the dialogue agent in the library is presented on the figure 1. Key concepts of DeepPavlov library are following.

- `Agent` is a conversational agent communicating with users in natural language (text).
- `Skill` fulfills user's goal in some domain. Typically, this is accomplished by presenting information or completing transaction (e.g. answer question by FAQ, booking tickets etc.). However, for some tasks a success of interaction is defined as continuous engagement (e.g. chit-chat).
- `Component` is a reusable functional part of Skill.
- `Skill Manager` performs selection of the Skill to generate response.
- `Chainer` builds an agent/component pipeline from heterogeneous components (Rule-based/ML/DL). It allows to train and infer models in a pipeline as a whole.

3 Interactive Demo

An interactive demo covers main functional components implemented in the library and is available at

`http://demo.ipavlov.ai/`

The demo showcases performance of pretrained models by reading user input and returning system response. It includes following sections (fig. 2).

Insult detection
USER INPUT: an utterance for insult detection.
SYSTEM OTPUT: a classification result, i.e. 'insult'/'not insult'. The model is CNN trained on Kaggle dataset "Detecting insults in social commentary" [2].

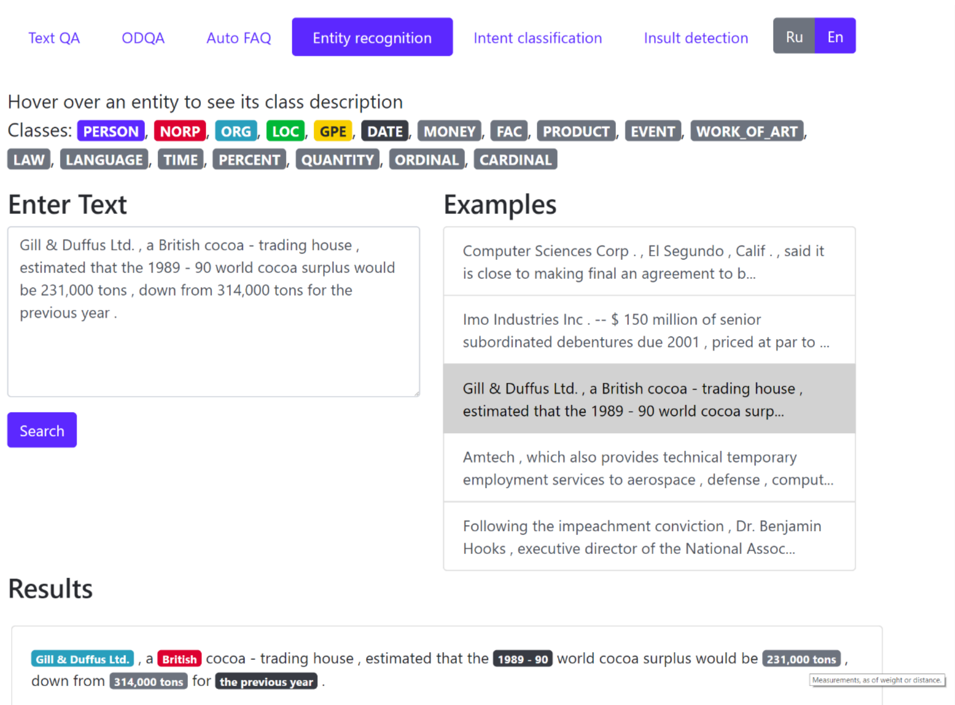

Figure 2: Web interface of DeepPavlov demo. From the top menu user selects component for demonstration. Then input text can be provided by the user or selected from the examples on the right. Output of the system displayed below in the *Results* section.

Intent classification
USER INPUT: utterance for intent classification.
SYSTEM OTPUT: one of seven intents. The model is CNN trained on the SNIPS.ai dataset [3].

Entity recognition
USER INPUT: text for entity recognition.
SYSTEM OTPUT: text with tokens marked up with 18 categories. The model is Bi-LSTM CRF trained on the OntoNotes 5 dataset [4].

Auto FAQ
USER INPUT: question from the insurance dataset.
SYSTEM OTPUT: answer with the highest similarity match from predefined FAQ. The model is Bi-LSTM trained on the Insurance QA dataset [5].

Text QA
USER INPUT: some text and question related to it.
SYSTEM OTPUT: snippet from the user provided text that might be an answer. The model is R-Net [6] trained on the SQuAD dataset [7].

ODQA.
USER INPUT: open domain question.
SYSTEM OTPUT: snippet from the Wikipedia that might be an answer. The solution combines document search with text QA.

Acknowledgments

The work was supported by National Technology Initiative and PAO Sberbank project ID 0000000007417F630002.

References

[1] Burtsev, Mikhail, et al. "DeepPavlov: Open-Source Library for Dialogue Systems." Proceedings of ACL 2018, System Demonstrations (2018): 122-127.

[2] https://www.kaggle.com/c/detecting-insults-in-social-commentary/data

[3] https://github.com/snipsco/nlu-benchmark/tree/master/2017-06-custom-intent-engines

[4] https://github.com/ontonotes/conll-formatted-ontonotes-5.0

[5] Feng, Minwei, et al. "Applying deep learning to answer selection: A study and an open task." arXiv preprint arXiv:1508.01585 (2015).

[6] Wang, Wenhui, et al. "Gated self-matching networks for reading comprehension and question answering." Proceedings of ACL 2017, Long Papers (2017): 189-198.

[7] https://rajpurkar.github.io/SQuAD-explorer/