Goal

The goal of the project is to explore how the footprint of components for Natural Language Understanding (NLU) can be reduced in a controlled way.

Background

The quality of a conversational system such as a chat-bot or virtual assistant largely depends on the performance of its Natural Language Understanding (NLU) component. In recent decades, data-driven and statistical approaches to NLU have significantly improved performance and is now increasingly being used for conversational AI. This usage is simplified by the fact that various pre-trained models for NLU tasks such as intent classification and entity extraction can be downloaded and used for free. For example, spaCy is a popular package for Python, containing pre-trained convolutional neural networks for word vectors, parsing and entity extraction for various languages.

Problem description

One of the challenges of applying pre-packaged NLU models is their footprint. For example, an NLU component that uses a wide-coverage spaCy model may consume several gigabytes of RAM, which can be problematic e.g. for embedded applications. This problem is partly addressed by the availability of pre-packaged models in different sizes. However, this leaves very limited room for dialogue developers to navigate the trade-off between performance and footprint.

This project will investigate how statistical NLU models can be scaled or pruned in order to reduce footprint in the context of conversational systems. For example, are there domain-specific considerations that could guide the scaling-down of a language model, thereby offering a more fine-grained control over the NLU capacities of a conversational system?

The project is proposed by Talkamatic, a company developing a platform for conversational AI called Talkamatic Dialogue Manager (TDM). TDM supports the use of Rasa NLU and spaCy for NLU, but the project is not limited to these platforms or models.

Recommended knowledge and skills

- NLP (libraries, models, algorithms)
- General machine learning
- Interest in Conversational AI

Supervisors from Talkamatic

- Alex Berman
  alex@talkamatic.se