Using Relational Features to Detect Opinion Targets

- Thesis with AI researchers from Swisscom
- AI project using cutting edge machine learning methods
- Can lead to a scientific publication and/or an offer from us

Description

Mining the opinions people express has applications in numerous fields, from customer relations to policy creation. While in certain cases knowing the overall sentiment in a document has value, in general that analysis level is not enough. A more informative view is provided by attaching opinions to specific targets. These more fine grained opinions, known as aspect-based opinions, require the extraction of the relevant aspects. For example in the phrase "This thesis will be awesome" the aspect is "thesis".

Multiple aspect modelling approaches have been proposed. In this thesis, you will focus on the influence of relational features on aspect extraction. Relational features include syntactic relations within phrases, based on the phrase parse trees. You can read more about parse trees and syntactic relations
here: https://spacy.io/blog/german-model

This year Google released its free parser Parsey McParseface (pre trained English parser), along with 40 cousins for other languages. Spacy also created a model for German, so the choice of models increased dramatically from the past, when the default choice were the Stanford models. This is an opportunity for you to acquire experience with state of the art industrial parsers.

These relations have recently been employed to improve the extraction of opinion targets
(e.g. http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00141 or *https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/download/11973/12051).*
Unlike many existing rule based approaches, we will learn the importance of the syntactic relations for the aspect retrieval task. These will be features that will be combined with others basic features linked to the vocabulary. You will then perform classification tasks using all the features and discriminate opinion aspects from regular words.

Your Tasks

In this Thesis you will create a method for extracting opinion targets by mixing word embeddings, vocabulary and syntax features.

- You will use multiple parsers to generate the relational features
  - Syntaxnet
  - Stanford parser
  - Spacy
- You will combine these features with word embeddings (e.g. obtained using Google's Word2vec)
  - You will use the resulting feature vectors in a classification task to find opinion aspects. You will train the classifiers with publicly available corpora, like the ones from Semeval ABSA (e.g. semeval 2016, task5)
- You will compare the results obtained with each parser with the baseline that does not use relational features.
- You will work with at least two languages - German and English.

Ideal candidate

- **Experienced with Python and a good coder in general.**
- Familiar with machine learning and AI and passionate about the topic. You should have taken at least a ML course.
- Good English is a must. Knowledge of French and German helps.

Contact

Internal Supervisor Prof. Boi Faltings (boi.faltings@epfl.ch)

External Supervisor Dr. Musat Claudiu (Claudiu.Musat@swisscom.com)