Multi-task Learning for Text Normalization, Parsing and Machine Translation

Recent progress in Natural Language Processing (NLP) are conditioned to two key elements. First, they require relatively large amount of annotated data for the given task. Second, the resulting system will perform very well only if the test data is somehow not too different from the training data. In other words, state-of-the-art NLP systems remain constrained in the domain, language level and genre on which the model has been trained. The path that will be taken to tackle this challenge is multi-task learning. By bringing together various tasks such as machine translation, syntactic parsing and text normalisation, and by designing algorithms able to efficiently make use of their complementarity, we aim to build more robust NLP systems. The context of this research is twofold: the SoSweet project, which aims at modeling socio-linguistic aspects of social media linked to language variability, and the Parsiti project, which aims at developing more robust and context aware models for parsing and machine translation, specifically for user-generated content as found on social media.