Coping with Limited Training Data in Verb Phrase Ellipsis Detection using Active Learning Approach

Muhammad Ahmad
Universität des Saarlandes
mahmad@lsv.uni-saarland.de

Ingo Reich
Universität des Saarlandes
i.reich@mx.uni-saarland.de

Dietrich Klakow
Universität des Saarlandes
dietrich.klakow@lsv.uni-saarland.de

Verb phrase ellipsis is a very sparse phenomenon which can be detected using machine learning based models but these models require a significant amount of annotated training data. It is costly to obtain annotated training data; to cope with this problem, we present an active learning based approach, using gradient boosting as classification model, to detect the instances of VPE-indicators. We used British National Corpus to evaluate our approach and obtained 3.57% improvement in F1 score compared to state-of-the-art results while using only 36.74% of training data which in results saves 63.26% of annotation effort.
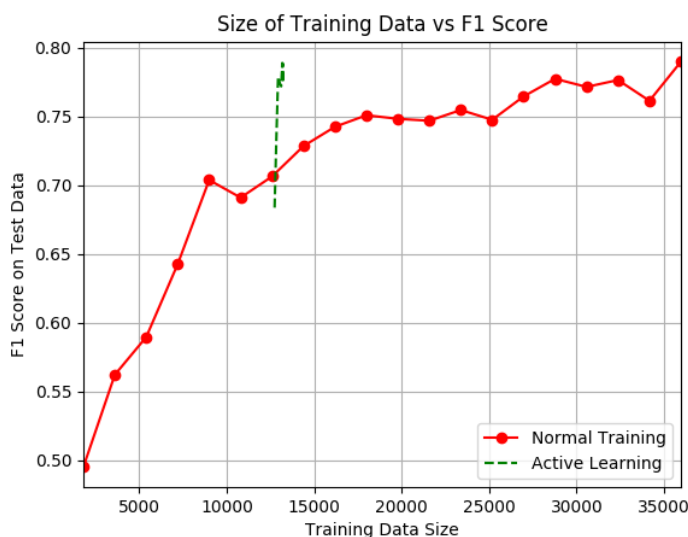
References: • Olsson, F. (2009). A literature survey of active machine learning in the context of natural language processing. SICS Technical Report T2009:06. • Liu, Z., Pellicer, E. G., & Gillick, D. (2016, June). Exploring the steps of Verb Phrase Ellipsis. In CORBON@ HLT-NAACL (pp. 32-40). • Kenyon-Dean, K., Cheung, J. C. K., & Precup, D. (2016). Verb Phrase Ellipsis Resolution Using Discriminative and Margin-Infused Algorithms. In *EMNLP* (pp. 1734-1743).