We experimentally investigate the behavior of several AL strategies for sequence labeling tasks (in a partially-labeled scenario) tailored on Partially-Labeled Conditional Random Fields, on four sequence labeling tasks: phrase chunking, part-of-speech tagging, named-entity recognition, and bio-entity recognition.

**INTRODUCTION**

One of the main problems of machine learning approaches lies in their need of large human-annotated training data. The process of active learning (AL) asks human annotators to label new samples which are supposed to be the most informative in the creation of a new classifier. In this work we focus on AL strategies for partially labeled sequences adopting the single token, instead of the entire sequence, as annotation unit and Partially-Labeled CRFs (PL-CRFs) as learning algorithm.

**Partially-Labeled CRFs**

PL-CRFs allow to learn a CRF model using partially-labeled sequences.

**Active Learning Strategies**

We have presented several AL strategies tailored for PL-CRFs in a pool-based scenario. We have tested the proposed strategies on four different datasets for four different sequence labeling tasks. Differently from other similar work in the field of AL, in this study we have shown that margin-based strategies constantly achieve good performance on four tasks with very different data characteristics.

** REFERENCES**

