

Preface

This book presents entropy guided transformation learning (ETL), a machine learning algorithm for classification tasks. ETL generalizes transformation based learning (TBL) by automatically solving the TBL bottleneck: the construction of good template sets. ETL uses the Information Gain measure, through Decision Trees induction, in order to select the feature combinations that provide good template sets. This book also details ETL Committee, an ensemble method that uses ETL as the base learner.

The main advantage of ETL is its easy applicability to natural language processing (NLP) tasks. Its modeling phase is quick and simple. It only requires a training set and a naive initial classifier. Moreover, ETL inherits the TBL flexibility to work with diverse feature types. We also show that ETL can use the template evolution strategy to accelerate transformation learning.

The book also details the application of ETL to four language independent NLP tasks: part-of-speech tagging, phrase chunking, named entity recognition and semantic role labeling. Overall, we apply it to thirteen different corpora in six different languages: Dutch, English, German, Hindi, Portuguese and Spanish. Our extensive experimental results demonstrate that ETL is an effective way to learn accurate transformation rules. Using a common parameter setting, ETL shows better results than TBL with handcrafted templates for the four tasks. For the Portuguese language, ETL obtains state-of-the-art results for all tested corpora. Our experimental results also show that ETL Committee improves the effectiveness of ETL classifiers. Using the ETL Committee approach, we obtain state-of-the-art competitive performance results in the thirteen corpus-driven tasks. We believe that by avoiding the use of handcrafted templates, ETL enables the use of transformation rules to a greater range of NLP tasks.

The text provides a comprehensive introduction to ETL and its NLP applications. It is suitable for advanced undergraduate or graduate courses in Machine Learning and Natural Language Processing.

Rio de Janeiro, January 2012

Ruy L. Milidiú