Abstract

This paper presents a method to predict human assessments of machine translation (MT) quality based on a combination of binary classifiers using a coding matrix. The multiclass categorization problem is reduced to a set of binary problems that are solved using standard classification learning algorithms trained on the results of multiple automatic evaluation metrics. Experimental results using a large-scale human-annotated evaluation corpus show that the decomposition into binary classifiers achieves higher classification accuracies than the multiclass categorization problem. In addition, the proposed method achieves a higher correlation with human judgments on the sentence level compared to standard automatic evaluation measures.
Predicting Human Assessment of Machine Translation Quality by Combining Automatic Evaluation Metrics


Index Terms

Computer Science  
Artificial Intelligence

Keywords

Evaluation Metric Combination  
Human Assessment Prediction