Analyzing Probability Vectors for Named Entity Statistical Machine Transliteration

Volume 55 - Number 10
Year of Publication: 2012

Authors:
M. L. Dhore
S. K. Dixit
T. D. Sonwalkar

Abstract

Machine transliteration systems are classified as either Rule-based methods or statistical methods. A rule-based method focuses on transliterating names using lots of human-made rules set. These systems are simple to implement but require huge amount of language expertise. In statistical methods, the importance is given in converting transliteration problem into a classification problem and employs a statistical model to solve this classification problem. Though these methods don’t require expert knowledge of Language model, they need large amounts of bilingual data and good algorithm for training. Currently, basic Markov Chain Model (MM), Extended Markov Chain (EMC), Hidden Markov Model (HMM), Conditional Random Fields (CRF), Decision Tree (DT), Maximum Entropy Markov Model (MEMM) and Support Vector Machine (SVM) are the popular statistical approaches used by many researchers across the globe. This paper focuses on mathematical analysis of different statistical approaches used in machine transliteration of named entity which would be beneficial for many upcoming researchers to know the mathematics used behind the curtains.
Analyzing Probability Vectors for Named Entity Statistical Machine Transliteration

References

- Kevin Knight, 2009. Bayesian Inference with Tears, a tutorial workbook for natural language researchers
- Kevin Knight, 2009. Training Finite-State Transducer Cascades with Carmel
Analyzing Probability Vectors for Named Entity Statistical Machine Transliteration

- Charles Sutton and Andrew McCallum, An Introduction to conditional random fields for relational learning, University of Massachusetts, USA

Index Terms

Computer Science
Artificial Intelligence

Keywords
Conditional Random Fields
Decision Trees
Hidden Markov Model
Markov Chain

Statistical Machine Transliteration