

UNIVERSITÄT LEIPZIG

Abstract

Fakultät für Mathematik und Informatik

Institut für Informatik

Probabilistic models of natural language semantics

by Ingmar Schuster

This thesis tackles the problem of modeling the semantics of natural language. Neural Network models are reviewed and a new Bayesian approach is developed and evaluated. As the performance of standard Monte Carlo algorithms proved to be unsatisfactory for the developed models, the main focus lies on a new adaptive algorithm from the Sequential Monte Carlo (SMC) family. The Gradient Importance Sampling (GRIS) algorithm developed in the thesis is shown to give very good performance as compared to many adaptive Markov Chain Monte Carlo (MCMC) algorithms on a range of complex target distributions. Another advantage as compared to MCMC is that GRIS provides a straight forward estimate of model evidence. Finally, Sample Inflation is introduced as a means to reduce variance and speed up mode finding in Importance Sampling and SMC algorithms. Sample Inflation provides provably consistent estimates and is empirically found to improve convergence of integral estimates.

UNIVERSITÄT LEIPZIG

Zusammenfassung

Fakultät für Mathematik und Informatik
Institut für Informatik

Probabilistic models of natural language semantics

von Ingmar Schuster

Diese Dissertation befasst sich mit der Modellierung der Semantik natürlicher Sprache. Eine Übersicht von Neuronalen Netzwerkmodellen wird gegeben und ein eigener Bayesscher Ansatz wird entwickelt und evaluiert. Da die Leistungsfähigkeit von Standardalgorithmen aus der Monte-Carlo-Familie auf dem entwickelten Model unbefriedigend ist, liegt der Hauptfokus der Arbeit auf neuen adaptiven Algorithmen im Rahmen von Sequential Monte Carlo (SMC). Es wird gezeigt, dass der in der Dissertation entwickelte Gradient Importance Sampling (GRIS) Algorithmus sehr leistungsfähig ist im Vergleich zu vielen Algorithmen des adaptiven Markov Chain Monte Carlo (MCMC), wobei komplexe und hochdimensionale Integrationsprobleme herangezogen werden. Ein weiterer Vorteil im Vergleich mit MCMC ist, dass GRIS einen Schätzer der Modellevidenz liefert. Schließlich wird Sample Inflation eingeführt als Ansatz zur Reduktion von Varianz und schnellerem auffinden von Modi in einer Verteilung, wenn Importance Sampling oder SMC verwendet werden. Sample Inflation ist beweisbar konsistent und es wird empirisch gezeigt, dass seine Anwendung die Konvergenz von Integralschätzern verbessert.