

Kurzfassung

Empirische Analysen natürlicher Sprache zeigen, dass die Stärke der Abhängigkeiten zwischen Textfragmenten mit zunehmender Distanz einem Potenzgesetz (*power-law*) folgt. Motiviert durch diese Erkenntnis, widmet sich die vorliegende Arbeit der Frage, welche Modellklassen fähig sind, dieses fundamentale Verhalten zu modellieren.

Hierfür wird ein formales Framework entwickelt, das präzise mathematische Definitionen für Modelle, Modellklassen sowie die Potenzgesetz-Eigenschaft bereitstellt. Dieses Vorgehen gewährleistet die Konsistenz der Ergebnisse mit der aktuellen Forschung.

Im Kontext dieses Frameworks wird zunächst das etablierte Resultat reproduziert, dass **Hidden-Markov-Modelle** nicht in der Lage sind, ein solches Verhalten aufzuweisen. Der Beitrag dieser Arbeit besteht hierbei in der rigorosen Formalisierung, die zudem eine höhere Allgemeingültigkeit der Aussage ermöglicht. Darüber hinaus wird ein Modell vorgestellt, das die Potenzgesetz-Eigenschaft erfüllt, was in dieser formalen Ausarbeitung ein neues Ergebnis darstellt.

Keywords

natural language, mutual information, power-law, hidden Markov models