

発表概要

型付き字句規則によるテキストデータ向け文法推論

本多 峻^{1,a)} 倉光 君郎¹

2016年10月27日発表

近年、Web 上においてテキストデータが増加しており、それらを利用するためのテキスト処理技術の重要性が高まっている。しかし、テキストデータは柔軟な表現を持ち、様々なフォーマットが存在するために、構文解析プログラムの開発コストが問題となる。本発表では、テキストデータから構文規則を導出するための推論アルゴリズムを示す。我々の新しいアイデアは、アノテーション付きの字句規則による解析結果から、構文規則を決定的に推論することにある。これにより、構文解析プログラムの自動生成を実現し、その開発コストを削減する。また、推論された構文規則と手書きにより作成した構文規則を比較することで、推論アルゴリズムの評価を行う。

Grammar Inference for Text Data by Typed Lexical Rule

SHUN HONDA^{1,a)} KIMIO KURAMITSU¹

Presented: October 27, 2016

In recent years, the increase of text data raise for importance of text processing technique. Text data are represented in flexible and various formats. Therefore, it is involve significant cost to develop a parser. In this presentation, we present the inference algorithm to generate syntactic rule from text data. Our new idea is to deterministically infer the syntax rules from the analytical results using the annotated lexical rules. Thus, we provide automatic parser generating for text data that can reduce development costs. We evaluate the inference algorithm by comparing inferred syntactic rule to handwriting one.

¹ 横浜国立大学工学府
Faculty of Engineering, Yokohama National University,
Yokohama, Kanagawa 240-8501, Japan

a) honda-shun-xr@ynu.jp