

発表概要

Packrat Parser のメモ化領域の自動調整手法

森永 孝仁^{1,a)} 前田 敦司² 山口 喜教²

2013年1月15日発表

現在一般的に使われている構文解析手法では、解析時間を線形時間に保つため、バックトラックを行わない。しかし、バックトラックを行わない解析手法では、先読みできる記号数が限られ、解析できる文法に制限ができてしまう。バックトラックを行い、parsing expression grammar で表現できる広範囲な構文規則を解析する手法として packrat parsing という手法がある。この手法では、解析対象文字列のすべての位置において1度解析した結果を保存するメモ化という手法を用いて、同じ解析を繰り返し行わないことで、単純なバックトラックでは最悪の場合に指数時間が必要だった解析時間を線形時間に抑えている。本発表では、この手法の問題点である、解析対象文字列のサイズに比例してメモ化領域が肥大化する点を改善することを目標としている。本発表での提案手法として、効果的にメモからエントリを削除するための手法と、メモ領域のサイズと解析効率のトレードオフを軽減するための手法の2点がある。まず、前者として直近にメモを行った付近で頻繁にバックトラックが発生することを前提とし、時間的局所性を用いて最近最も用いていないエントリをメモから削除する手法や、バックトラックを行う際に入力対象の細かい部分から広い部分に解析が進むことを考慮し、最も古いエントリをメモから削除する手法を提案する。また、後者の手法として、メモのヒット率が悪くなってきた場合に、動的にメモ領域の大きさを変化させ解析効率を維持する手法を提案する。以上の3つの手法を組み合わせ、解析効率をなるべく維持した状態でメモ領域を削減を行う。

Automatic Optimization of Memoization Area
in Packrat ParserTAKAHITO MORINAGA^{1,a)} ATSUSHI MAEDA² YOSHINORI YAMAGUCHI²

Presented: January 15, 2013

Many of widely used parsing techniques avoid backtracking to keep time complexity linear. However, these techniques can lookup only limited number of symbols, thus placing some restriction on grammars. Packrat parsing is a parsing algorithm which use backtracking to handle wide range of grammars defined in parsing expression grammar. With this algorithm, a technique called memoization is used to store results of parsing performed in a position to eliminate duplicated analysis in backtrack to keep parsing time linear. In this presentation, we propose methods to eliminate entries from memo effectively, and a method to set trade-off between memo size and parsing efficiency. For entry elimination, we propose two methods; the first one assumes backtrack occurs mainly near the last position memoized and eliminate least-recently used entry, while the second one considers parsing order and eliminates oldest entry from the memo. For trade-off control method, we propose an algorithm to modify memo size dynamically when hit-rate degrades to keep parsing efficiency. Combining these methods, our system effectively reduce memo size while keeping parsing as efficient as possible.

¹ 筑波大学情報学群情報科学類
College of Information Science, School of Informatics,
Tsukuba University, Tsukuba, Ibaraki 305-8537, Japan

² 筑波大学コンピュータサイエンス専攻
Department of Computer Science, Tsukuba University,
Tsukuba, Ibaraki 305-8537, Japan

a) morinaga@ialab.cs.tsukuba.ac.jp