Ph.D. Thesis Assessment for Jiří Havelka

Dr. Keith B. Hall
Johns Hopkins University
Center for Language and Speech Processing
Baltimore, MD USA

Overview

Havelka's thesis presents an account of graph-theoretic properties of totally-ordered directed trees, commonly referred to as dependency trees. While these structures have been studied thoroughly in the graph-theory literature, Havelka derives novel characteristics of dependency trees with respect to the individual edges within them. This approach is most attractive when the trees are realizations of linguistic dependency relationships.

While much of the thesis focuses on the reformulation of graph-theoretic characterizations of the trees and efficient algorithms to automatically detect such phenomena, the final chapter presents an empirical analysis of 19 languages in terms of these edge-based dependency tree attributes. An all too brief analysis of this data follows which offers some interpretation of the data.

A substantial subset of the work presented in the thesis has been published in condensed form in very respectable computational linguistics conferences. The work has been published in both the Association of Computational Linguistics (ACL) and the North American chapter of the ACL (NAACL) in 2007 as well as in the Prague Bulletin of Mathematics. The work has been well accepted by the subset of the computational linguistics community interested in dependency analysis, both manual and automatic.

Novelty

The primary point of novelty in this work is the restatement of some well known properties of dependency trees in terms of the attributes of the edges contained within the trees. This does, of course, introduce novel properties of edges that are presented formally.

The algorithms presented throughout the thesis, although sometimes obvious implementations of the edge properties are, are presented in extensive detail. Algorithms are presented throughout the thesis in extensive detail. These sometimes include obvious implementations of the defined properties. Some claims are made about potential complexity results of algorithms which are not supported (e.g., at the end of section 4.3). While this does not detract from the novelty of the associated algorithms, supporting such claims would have provided an extra level of completeness to the work.

Spread through the entirety of the thesis is a novel notion of dependency analysis based on upper and lower tree analysis (Gaps, non-planar sets, etc.). This mechanism allows for the low complexity results of a number of the presented algorithms. While the idea is based on previous work, Havelka projects the concept throughout the thesis in the analysis of all edge characteristics.

Technical Correctness

Proofs are found throughout the thesis, some of these obvious, but others are quite insightful. While the proofs are flawless, the interpretation of them is often limited. Many of the attributes provide additional methods for determining non-projectivity and non-planarity; however, this is not always clear until later chapters.

In general, the thesis is probably overly formal for the types of results presented. Many of the theorems could be thought of as lemmas for a few very important theorems. One of those theorems is that edge-based attributes are sufficient for characterising projectivity, planaryity, well-nestedness, etc. Another would be the fact that upper-type properties influence the design of efficient algorithms for detection. Nonetheless, the work is correct and clear.

Presentation/Exposition

All aspects of the thesis are presented in very clear language. The organization of the work is logical; subsequent chapters are built upon the ideas presented in earlier chapters. As mentioned above, the work may be overly formal, but this is certainly an acceptable attribute.

Schematic examples accompany many of the conceptual definitions, providing a graphical interpretation of the ideas presented. This proves helpful because much of the work is based on objects that live in a geometric space and have a natural geometric interpretation.

Impact

Other work, most notably that of Nivre and Kuhlmann in the past few years, has explored many of the tree-based concepts presented in this thesis. However, the unified approach presented by Havelka, making use of edge-based analysis, is cleaner and offers a consistent story regarding dependency tree analysis.

Presently, there is an increasing interest in automatic analysis of dependency structures. While the individual findings in this thesis may not be significant, the work as a whole is certainly significant and should have a moderate impact on the field.

Summary

Havelka's work in this thesis provides a unified presentation of a novel interpretation of dependency tree properties. It presents simple algorithms for detecting these properties based on a theory of edges-based characteristics. Many of these characteristics are novel, in particular those that divide the type of non-projective (and non-planar) edges into upper and lower sets.

This work fulfills the expectations of a Ph.D. thesis. The merit of this work has been assessed by experts in the field by way of the conference publications related to the findings presented within.