Quantifying Simplicity of Generalized Trees

Mahsa Kiani, Virendrakumar C. Bhavsar, and Harold Boley
Mahsa.Kiani@unb.ca
Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada

Introduction

Simplicity, as the opposite of complexity [1], is a fundamental concept of cognitive science, which indicates the degree of plainness. The explanation of simplicity can be problem-specific [2]. The term simplicity has been associated with falsifiability, causality, communicability, as well as a low number of parameters. It is challenging to define simplicity precisely due to its ambiguity and variety of meanings. To quantify simplicity, one should determine what to measure [3].

Simplicity can be categorized into two distinct types: syntactic and semantic. While syntactic simplicity is defined based on the number of concepts and principles [4], semantic simplicity is defined based on the meaning [5].

Simplicity of data structures has been defined using different approaches.

- Fast execution time of operation of a data structure is an indication of its simplicity [6].
- Size (breadth and depth) and edge weights. for a weighted tree are considered as the measure of simplicity [7].

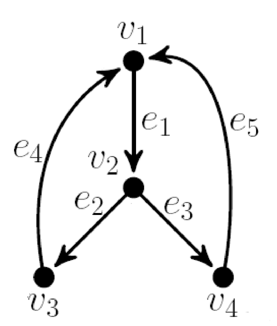
Contribution

We represent the metadata of each record stored in the database using a hierarchical graph, which is called a attributed generalized tree (AGT). The existing unlabeled/ vertex-labeled generalized tree structures have been extended to consider edge labels as well as edge weights. Then, we propose a function which considers the AGT characteristics to quantify its simplicity.

Representation of Data

Metadata of each record is represented as an Attributed Generalized Tree.

- Edge weights express users' assessment regarding the relative importance of the attributes represented by edge labels.
- Labels are unique and appear in lexicographic left-to-right order.
- Weights are in the real interval [0, 1] and for each generalized tree its edge weights are normalized.



Vertex	v_1	v_2	v_3	v_4
$l(v_i)$	A	B	C	D

Edge	e_1	e_2	e_3	e_4	e_5
$l(e_i)$	la	lb	lc	ld	le
$w(e_i)$	1.0	0.3	0.7	1.0	1.0

Simplicity Function

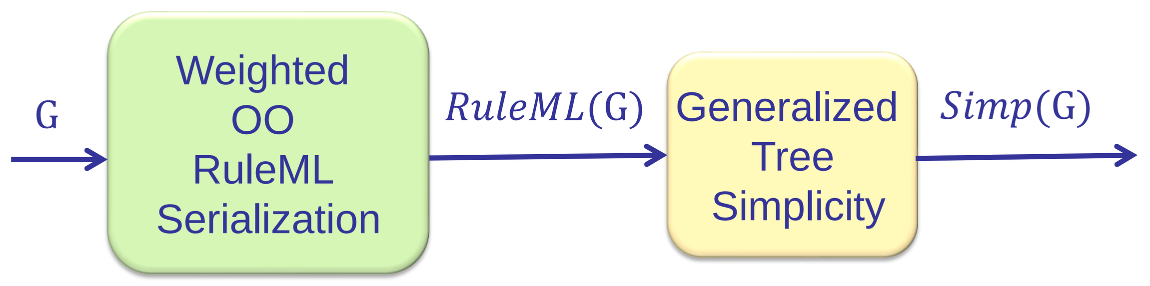
- Branch factor (B): based on the number of outgoing edges from each vertex
- Position factor (P): based on the depth of the source and destination vertices
- Weight factor (W): based on the weight of each edge

Analysis

- Size: With gradual increase in the number of edges, depth, and breadth in AGTs, the simplicity value decreases.
- Depth: With increasing depth of source and destination vertices of forward edges, back edges, and cross edges, the simplicity value decreases. Also, as the depth of the branching increases, the simplicity value decreases.
- Length: Increasing the length of forward as well as back edges results in the gradual decrease of the simplicity.
- Forward Edges versus Back Edges: The simplicity function could distinguish forward and back edges. When a forward edge and a back edge have the same characteristics except for their directions, the effect of a forward edge on simplicity is greater than the effect of a back edge.

Architecture

- G: An Attributed Generalized Tree
- $RuleML\ (G)$: RuleML representation of G
- $Simp(G)$: Simplicity value in interval [0, 1]



Architecture of the System

Weighted OORuleML Module

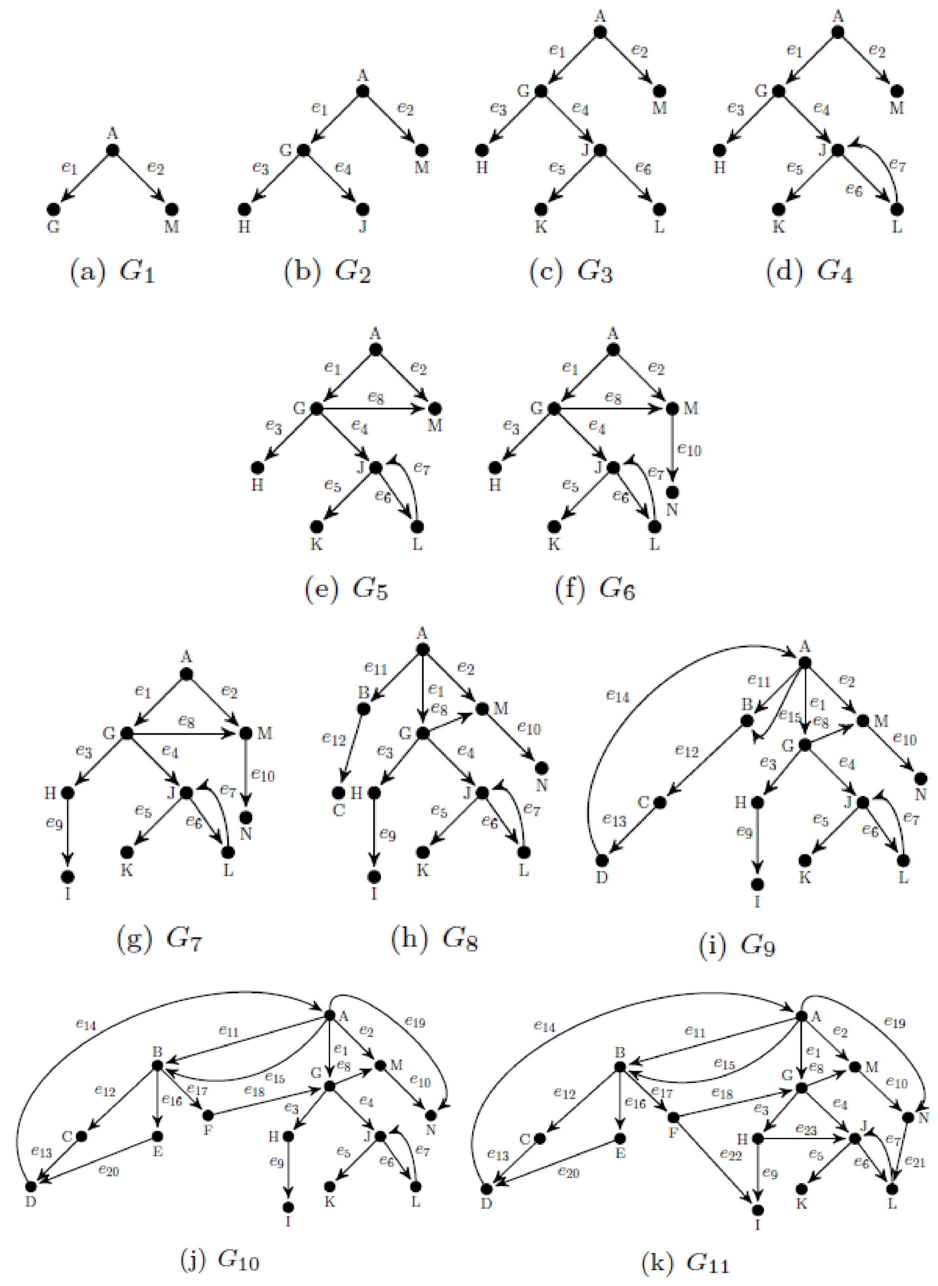
Each attributed generalized tree is uniformly represented and interchanged using a weighted extension of Object Oriented RuleML [8]. This approach preserves all structural information of each generalized tree including its hierarchical structure, vertex labels, edge labels, and edge weights.

Simplicity Module

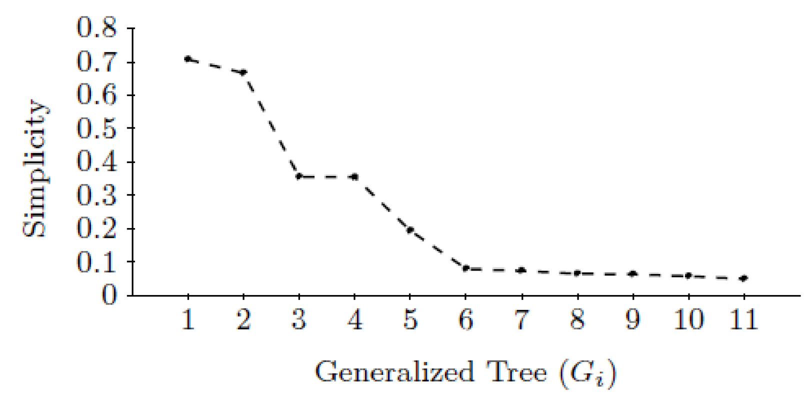
The simplicity measure is defined recursively to map an arbitrary single AGT G to a value in interval [0, 1]. The simplicity algorithm uses the cycle-detection strategy of the depth-first search in order to handle cycles in traversing G. The AGT simplicity algorithm has linear time complexity.

Computational Experiments

- Generalized tree dataset: $\mathbf{G}=\{G_1, G_2, \ldots, G_{11}\}$
- The simplicity value decreases with increasing the complexity of AGT structure.



AGTs in the Dataset



Simplicity Values

Conclusion

Simplicity is a fundamental concept in cognitive science. We have quantified the simplicity of arbitrary Attributed Generalized Trees (AGT). Simplicity function maps an AGT to a value in the interval [0, 1]. The function can be used in AGT similarity matching for domains with open world assumption.

Publication:
M. Kiani, V.C. Bhavsar, and H. Boley, "Semantic Computing of Simplicity in Attributed Generalized Trees", Proc. IEEE 15th Int'l Conf. on Cognitive Informatics & Cognitive Computing, 2016.

References:

[1] N. Rescher, "Complexity: a philosophical overview", 1998.
[2] H.A. Keuzenkamp, M. McAleer, and A. Zellner, "The enigma of simplicity", 2002.
[3] J.C. Lindahl, "Anthropological Approaches to the Philosophy of Translation", 1999.
[4] A. Kulka, "Methods of Theoretical Psychology". 2016.
[5] E.O. Aboh, and N. Smith, "Complex processes in new languages", 2009.
[6] T.H. Cormen, C. Stein, "Introduction to algorithms", 2009.
[7] L. Yang, B.K. Sarker, V.C. Bhavsar, H. Boley, "A weighted tree simplicity algorithm for similarity matching of partial product descriptions", 2005.
[8] H. Boley, B. Grosof, M. Kifer, M. Sintek, S. Tabet, and G. Wagner, "Object-oriented ruleml", 2004.