# Hierarchical Inference Applied to Cyc

Dario GARCIA-GASULLA [a] Ulises CORTÉS [a]
e-mail: {dariog,ia}@lsi.upc.edu
[a] *KEMLg, Universitat Politècnica de Catalunya*

**Abstract.** Hierarchical graphs are a frequent solution for capturing symbolic data due the importance of hierarchies for defining knowledge. In these graphs, relations among elements may contain large portions of the element's semantics. However, knowledge discovery based on analyzing the patterns of hierarchical relations is rarely used. We outline four inference based algorithms exploiting semantic properties of hierarchically represented knowledge for producing new links, and test one of them on a generalization of Cyc's KB. Finally, we argue why such algorithms can be useful for unsupervised learning and supervised analysis of a KB.

## 1. Introduction

Hierarchical graphs (HG) include generalization and specialization relations usually found in taxonomies and ontologies. These graphs are frequently used for Knowledge Representation (KR) (*e.g.*, ontologies, social networks, semantic web data, *etc.*), but there are still few knowledge discovery techniques for this type of structure[3]. Hierarchical properties, such as inheritance and abstraction, are powerful tools for KR and knowledge discovery. These properties however are difficult to capture through the components of a graph which has motivated the use of logics for that purpose (FOL, Inductive Logic Programming, *etc.*). Some research fields have developed learning tools for HG[3], but none of these approaches use hierarchical semantics as basis for inferential learning. Some consider inheritance, but focus on spatiotemporal relations through Bayesian inference [1]. Some apply logic inference instead as its basis for learning [4]. In that context, our hypothesis is that features provided by strong generalization and inheritance properties capture enough of the data semantics as to empower inference to achieve learning.

In this paper we outline algorithms based on inference to be applied on a HG. Given a graph's topology and assuming properties of hierarchical knowledge the algorithms use the principles of inference to predict edges. These algorithms are an unsupervised learning methodology for HG. The results can also be used in a supervised manner to detect flaws in KBs (*e.g.*, inconsistencies and unbalances). We test one of these algorithms on a generalized version of Cyc KB, and overview the results.

## 2. Background

Conceptual Clustering obtains a classification from a set of data descriptions[6], which may include a hierarchy. These methods work on weakly structured data and use weak interpretations of hierarchy (*e.g.*, no multiple inheritance). [2] presents a methodology for hierarchy discovery. Given a graph, it produces a set of hierarchies ranked by likeliness, turning any graph into a HG. Assuming the importance of hierarchy for defining the data, it predicts links according to it. Its missing links prediction is based on its proposed hierarchy, meaning it can only be applied to graphs hierarchized by their method. DeSTIN[1] is a Deep Learning System (DLS) which builds a hierarchy of nodes in which each layer represents a different spatiotemporal scale. Each node includes a set of patterns and of belief states. DLS are most suitable for uniformly and statically distributed data (*e.g.*, sensory information), but may not be appropriate for knowledge without a stable hierarchical pattern (*e.g.* ontologies). Logic-based systems represent hierarchical properties through logic predicates, like Cyc[5] which includes components for reasoning and learning. Cyc's KB is an ontology represented hierarchically through collections in a graph like structure. Cyc's predicates are strongly typed and logical deduction is its main source of reasoning. Opencog[4] KB is based on a hypergraph generalization which defines several types of vertices and edges. It uses a wide variety of methods for learning, among which there is probabilistic logic inference, clustering and pattern mining. Other approaches like Link Prediction and Probabilistic Graphical Models are relevant contributions to the field [3] but no not consider hierarchies for learning.

## 3. Motivation and algorithms

We argue that the topology of hierarchical edges among symbolic entities holds a large portion of its meaning and that it can be used to infer knowledge about those entities. We implement algorithms for extending a KB's only using its distribution of hierarchical relations based on an unlabeled, directed graph in which there is a single type of edge and node. For any two nodes linked in the graph, we will refer to the node source of the relation as the *child* and to the node destination as the *parent*. The relevancy of this model is the achievement of inferential reasoning without using higher level tools such as logics. While logics allow a more accurate and rich approach to inference, they partly detach the KR and reasoning from the graph. Our proposal works only with knowledge found in the graph. In the current context with large hierarchical KBs available this work can contribute to the construction, enrichment and evaluation of these KBs.

The algorithms we define apply the principles of inference to hierarchies. Deductively, one may say that a child node inherits the edges from its parent node. This specialization process can be implemented through a top-down transitivity closure for edges (*i.e.*, the edges of each parent are directly inherited by each children), and is frequently found in most hierarchical datasets. Inductively, one may say that the relations of a parent are a generalization of the relations of all its children. This generalization process can be implemented through a bottom-up weighted transitivity for edges. To evaluate the existence of the edge $A \rightarrow B$ we consider how many of A's children are also children of B and use that data as weighted evidence. Finally, abductively one can find parenthood relations which explain the existence of other relations. Given $B \rightarrow A$ and $C \rightarrow A$, an explanation for the latter could be a newly proposed relation $C \rightarrow B$. We define two of those algorithms, the first proposes the edge $C \rightarrow B$ given that most of A children are also of B (and therefore by induction also C). The second one proposes the edge $C \rightarrow B$ given C shares a pattern of edges which is shared by B but not by most children of A. More details on the algorithms can be found in [3].

| | | | |
|---|---|---|---|
| car | land transportation device | object underneath which a human can move past | |
| wheeled vehicle | single purpose device | self-powered vehicle | man-made pollution source |

**Table 1.** Sample of induced parents for *automobile model*

## 4. Testing on Cyc

We test our algorithms on OpenCyc, the open version of Cyc's KB, which contains 120K *Class* concepts and 537K relations among them. To transform it into a HG we kept the two most common relations, *rdf:type* and *rdfs:subClassOf* since both their meanings are hierarchical. The rest of relations were omitted but we still work with most of the KB (these two relations represent over 80% of all edges). We unlabel them and consider the resultant relation as a unique transitive (through deduction) relation. After applying these changes the resultant HG is composed by 116K nodes and 354K edges.

Due to space limitation we will present only the results obtained by the inductive algorithm. Further analysis on the algorithms precision can be found in [3]. For the inductive algorithm we considered a 99% uncertainty threshold (*i.e.*, 99% children of *n* had to share an edge for it to be induced by *n*). We added further restrictions regarding the specificity level of the candidate nodes to avoid too abstract inductions and too similar inferences[3]. The algorithm produced 2.181 relations from induction and 2.152 from applying deduction to the newly induced edges. As an overview of the results let us focus on the inductions produced for the element *automobile model*. Originally, the node had 7 transitive parents (*e.g.*, type of object, facet collection, road vehicle model, etc.) and 631 transitive children (*e.g.*, Volvo 140). 17 additional parents were discovered by induction (see a small sample in Table 1) and 85 new relations were produced with deduction for its children. One of the discovered parent was *car*, which was captured in OpenCyc's KB through an *ObjectProperty* by a generalizing predicate. Although this relation was deleted as only Class-to-Class relations were kept, the algorithm was capable of re-discovering it from the patterns of *automobile model* children.

In the original KB, 6 of the 17 induced parents were related with *car* and therefore with *automobile model*, but those relations were deleted in the adaptation process. The algorithm rebuilt those relations from patterns of *automobile model* children. This implies that the children of an element contain information regarding the element's properties, information which can be used to produce new data. The remaining 10 elements were not directly related with *car* in OpenCyc's KB, meaning the algorithm produced information which was not directly found on the original KB. The relations added by applying deduction to the induced edges are less relevant since they are a direct effect of the induced relations. A brief example are elements *AcuraCL* and *ChevyS10Blazer*, children of *automobile model* which by deduction are now children of the 17 new parents of *automobile model*, including those of Table 1. Finally, in Table 2 we present a sample of other inductions which we consider to be semantically interesting.

## 5. Conclusions and discussion

In this paper we propose the use of algorithms inspired on inferential reasoning for learning in HG. For that purpose it is necessary to assume that hierarchical relations represent both a top-down specialization and a bottom-up generalization of concepts. We applied

| | |
|---|---|
| coral reef - natural thing | the smell of chemically acidic substances - rust causing stuff |
| pure breed - tame animal | unalloyed metal - stuff composed entirely of one element |
| beheading - hostile actions | mass cell division -event in which some quantity changes |
| group of sports events - conflict | typical artillery target type - high payoff target type |
| medical school graduate - skilled worker | logical feature of binary predicates - truth function |
| new years resolution - agreement topic | product with warranty - repairable product |

**Table 2.** Sample of inductions. Left element of cell is induced to be a children of the right element of cell

these algorithms on a directed, unlabeled graph, and tested it on a generalization of a rich KB to study the type of relations we can infer. The results seem to indicate that such methodologies can be useful to produce additional information about concepts without the use of exogeneous inferential knowledge such as logics. At the same time, the universality of inference allows the algorithms to produce unbounded and complex learning which is hard to obtain through statistical methods.

To the best of our knowledge no other contribution in the current state of the art combines hierarchical features with inference in a similar manner. These algorithms could be applied to any KB which can be reduced to a HG, the relations of which satisfy the properties of hierarchically represented knowledge. The results proved the approach to be highly dependent on the KB's topology and on the existence of irrelevant relations and sparsity on the graph. The method is an unsupervised learning approach for semantic learning which produces new relations. Additionally, by supervising the results the algorithms can be used to improve the contents and topology of a KB by finding inconsistencies and unbalances in it. Both functionalities are currently appealing as there are many large symbolic datasets available which could benefit from it.

## 6. Acknowledgments

Thanks to the ACIA fellowship for assistance to CCIA'2013.

## References

[1] I. Arel, D. Rose, and R. Coop. Destin: A scalable deep learning architecture with application to high-dimensional robust pattern recognition. In *Proc. AAAI Workshop on Biologically Inspired Cognitive Architectures*, pages 1150–1157, 2009.

[2] A. Clauset, C. Moore, and M. E. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.

[3] D. Garcia-Gasulla and U. Cortés. *Hierarchical Link Prediction and Reasoning*. PhD thesis, Proposal, 2013.

[4] B. Goertzel. *OpenCog Prime: Design for a Thinking Machine*. http://opencog. org, 2008.

[5] D. Lenat, R. Guha, K. Pittman, M. Shepherd, and D. Pratt. Cyc: toward programs with common sense. *Communications of the ACM*, 33(8):30–49, 1990.

[6] R. Michalski. Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts. *Journal of Policy Analysis and Information Systems*, 4(3):219–244, 1980.