A CLP-Based, Diagnosticity-Driven System for Concept Combinations

Georgios Tagalakis
University College Dublin
Department of Computer Science
Dublin 4, Ireland
georgios.tagalakis@ucd.ie

Daniela Ferrari
University of Siena
Department of Philosophy
and Social Sciences
via Roma, Siena 53100, Italy
daniela.ferrari@ucd.ie

Mark T. Keane
University College Dublin
Department of Computer Science
Dublin 4, Ireland
mark.keane@ucd.ie

Abstract

Diagnosticity operates as an important selection criterion for several computational models of concept combination. Unfortunately, it has not been clear how the diagnosticity of property and relational predicates of the concepts combined can be formalized and quantified. Using an information retrieval method we compute, in a uniform manner, diagnosticity values of concepts predicates. We go on to present a reasoning system that attempts to create meaningful interpretations of novel noun-noun combinations. The system is based solely on diagnostic predicates values and a set of constraint satisfaction rules. We show the effectiveness and plausibility of our methods and discuss their potential.

1 Introduction

Noun-noun combinations play a central role in language and cognition. People produce two main types of interpretation to noun-noun composite concepts: *relational* interpretations that hinge on a relation being found to connect the two concepts (e.g., *finger cup*: "a cup to clean fingers in") and *property* interpretations that hinge on transferring a property from one concept to the other (e.g., *finger cup*: "a thin cup") (see [Levi, 1978]). In this process, the *diagnostic* predicates of the concepts combined have been found to play a key role. Furthermore, several computational models use the *diagnosticity* constraint as an important selection criterion (e.g., [Costello and Keane, 2000]). However, the diagnosticity mechanisms have so far only been formalized for property predicates, partly because it has not been clear how one can operationally define diagnosticity for relational predicates. Property predicates typically represent perceptual or physical aspects of objects. Relational predicates usually put the concept in connection with other concepts or refer to functions.

In this study, we use a new method for characterizing the diagnosticity of property and relational predicates, which is based on van Rijsbergen's [1979] *tf-idf* information retrieval schema. We applied this method to compute the diagnosticity values of the predicates of a set of concepts. The data are input to a reasoning system that attempts to provide meaningful interpretation(s) for a given noun-noun combination based on its interpretation type, the diagnosticity values of the predicates of its constituent concepts and a small set of rules.

2 Formalizing Diagnosticity

The idea underlying the formalization of diagnosticity is that a feature predicate $\mathcal{F}$ is important for a concept C within a set of concepts, if it occurs often in descriptions of concept C, and rarely in descriptions of the rest of the concepts in the set [Ferarri and Keane, 2005]. Given a set of descriptions of concepts, we can thus compute diagnosticity values for each feature relative to a concept $D_{C,\mathcal{F}}$ by adapting the information retrieval *tf-idf* formula [van Rijsbergen, 1979] as:

$$D_{C,\mathcal{F}} = ff_{C,\mathcal{F}} \times \frac{log_2 N_{CS}}{n_{\mathcal{F}}} \tag{1}$$

where $ff_{C,\mathcal{F}}$ is the frequency of the feature in a certain concept, and $log_2 N_{CS}/n_{\mathcal{F}}$ is a non-linear measure derived from the ratio between the number of concepts N_{CS} in the set and the number of concepts $n_{\mathcal{F}}$ described by the feature at issue.

Once the diagnosticity values for concept predicates are derived, we can class every concept as property or relational depending on whether its total diagnosticity is given mainly by relations or properties. When we make a combination from two single concepts each of them bears diagnosticity of property features (DPF) and diagnosticity of relational features (DRF). These values can in turn be summed as to characterize the combination with a total property diagnosticity (TPD) and a total relational diagnosticity (TRD).

A list of predicates was formed by asking subjects to describe a set of concepts and any associated knowledge. Formula 1 was used to compute diagnosticity values for their diagnostic predicates. Predicates with one standard deviation or more above the mean score were defined as highly diagnostic (HD). A combination was classified as relational, when $TRD > TPD$ or property, when $TPD > TRD$.

3 Diagnosticity-Driven Interpretation Generation

To test the potential of diagnosticity as a driving force in concept combination, we developed a reasoning system (in Prolog) using constraint programming techniques. The techniques were developed after analysis of the data of a previous

Algorithm 1 Diagnosticity-Driven Combination Rules for Relational Noun-Noun Composite Concepts

get all MN and HN's property and relational diagnostic features (DPF_{MN}, DRF_{MN}, DPF_{HN}, DRF_{HN})
if 1 or more $DPF_{MN} \geq HD$ **then**
  if all $DRF_{MN} < HD$ **then**
    HN that [DPF_{MN}], s.t. $DPF_{MN} \geq HD$
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} = DRF_{max_{MN}}$
    MN [DRF_{HN}] (HN), s.t. $DRF_{HN} = DRF_{max_{HN}}$
  else
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} \geq HD$
    MN [DRF_{HN}] (HN), s.t. $DRF_{HN} = DRF_{max_{HN}}$
    HN that [DPF_{MN}], s.t. $DPF_{MN} \geq HD$
  end if
else
  if all $DRF_{MN} < HD$ **then**
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} = DRF_{max_{MN}}$
    MN [DRF_{HN}] (HN), s.t. $DRF_{HN} = DRF_{max_{HN}}$
  else
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} = DRF \geq HD$
    MN [DRF_{HN}] (HN), s.t. $DRF_{HN} = DRF_{max_{HN}}$
  end if
end if

empirical study, where 30 people were asked to provide interpretations to 60 novel noun-noun combinations with predicates of varied diagnosticity scores and distributions.

3.1 Architecture and Implementation

The system has an interface that asks the user to enter the modifier noun (MN) and the head noun (HN) of a composite concept. This input is parsed and checked against its knowledge base. The knowledge base is in clausal form and consists of a declarative definition of concepts with their predicates and their corresponding diagnostic values as computed above. If both nouns are known to the system, it proceeds to apply the reasoning procedures. The core procedure for the interpretation of relational combinations is sketched in Algorithm 1 and that of the property ones in Algorithm 2. Action constraints are represented as binary functions and enforce valid transitions between states.

3.2 Evaluation

Evaluation of the interpretations produced was conducted with 30 noun-noun concepts, half of which were property and half relational, with constituent concepts ($N = 50$) of varying diagnosticity scores. All reasoning sub-mechanisms were tested. The results were compared with existing people's data. Sixty two interpretations were produced. 81% ($N = 50$) of the interpretations were deemed to be sensible by two judges. Half of the meanings generated were among those people had produced. For the property combinations in particular, it was found that the interpretations generated were those people produced more often.

4 Discussion

Currently, the application of the new method for computing diagnosticity to "diagnostically mark" concept predicates of concept glosses of large, standardized knowledge bases (e.g., WordNet) is being examined. The benefits of exploiting such a rich information for concept combination has been demonstrated in [Tagalakis and Keane, 2004]. Diagnosticity marking can be an important enhancement to the knowledge representation of ontology systems in general.

As regards the reasoning model, the percentage of correct interpretations it produces based on diagnostic values is very encouraging; especially when one considers that the system is self-contained and does not take into account other key aspects of concept understanding (e.g., plausibility, informativeness, previous knowledge). Future work will provide additional model validation and programming methodology details.

Algorithm 2 Diagnosticity-Driven Combination Rules for Property Noun-Noun Composite Concepts

get all MN and HN's property and relational diagnostic features (DPF_{MN}, DRF_{MN}, DPF_{HN}, DRF_{HN})
if 1 or more $DPF_{MN} \geq HD$ **then**
  if all $DRF_{MN} < HD$ **then**
    HN that [DPF_{MN}], s.t. $DPF_{MN} \geq HD$
  else
    HN that [DPF_{MN}], s.t. $DPF_{MN} \geq HD$
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} \geq HD$
  end if
else
  if all $DRF_{MN} < HD$ **then**
    HN that [DPF_{MN}], s.t. $DPF_{MN} = DPF_{max_{MN}}$
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} = DRF_{max_{MN}}$
  else
    HN [DRF_{MN}] (MN), s.t. $DRF_{MN} \geq HD$
    HN that [DPF_{MN}], s.t. $DPF_{MN} = DPF_{max_{MN}}$
  end if
end if

References

[Costello and Keane, 2000] F. J. Costello and M. T. Keane. Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, 24:299–349, 2000.

[Ferarri and Keane, 2005] D. Ferarri and M. T. Keane. Diagnosing conceptual combinations: A methodology for operationalizing diagnosticity in conceptualization. Unpublished manuscript, 2005.

[Levi, 1978] J. Levi. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York, 1978.

[Tagalakis and Keane, 2004] G. Tagalakis and M. T. Keane. Automating the interpretation of novel, noun-noun compounds using WordNet. In P. Gervás and K. M. Gupta, editors, *Proceedings of the European Conference on Case-Based Reasoning 2004 Workshops*, pages 225–236, Madrid, 2004. Complutense University of Madrid.

[van Rijsbergen, 1979] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 1979.