Analogy as Functional Recategorization: A Perspective from HowNet

Tony Veale

Department of Computer Science,
University College Dublin, Belfield, Dublin 6, Ireland.
Tony.veale@UCD.ie
http://www.cs.ucd.ie/staff/tveale/home/

Abstract. One generally accepted hallmark of creative thinking is an ability to look past conventional labels to recategorize a concept or object based on its behaviour and functional potential. So while taxonomies are useful in any domain of reasoning, they typically represent the conventional label set that creative thinking attempts to look beyond. If a linguistic taxonomy like WordNet [1] is to be useful in driving linguistic creativity, it must support some basis for recategorization, to allow an agent to reorganize its category structures in a way that unlocks the functional potential of objects, or that recognizes similarity between literally dissimilar ideas. In this paper we consider how recategorization can be used to generate analogies using the HowNet [2] ontology, a lexical resource like WordNet that in addition to being bilingual (Chinese/English) also provides explicit semantic definitions for each of the terms that it defines.

1 Introduction

Analogy is a knowledge-hungry process that exploits a conceptual system's ability to perform controlled generalization in one domain and re-specialization into another. The result is a taxonomic leap within an ontology that transfers semantic content from one term onto another. While all taxonomies allow vertical movement, a system must fully understand the effects of generalization on a given concept before any analogy or metaphor can be considered either deliberate or meaningful. Research has shown that the most satisfying analogies are those that operate at the causal level of representation, since causality allows an analogy to offer an deep explanation of a poorly understood phenomenon (e.g., [3], [4]). Thus, *the atom as miniature solar-system* is satisfying because the source and target concepts are causally structured around the notion of rotation. So to support analogy, a taxonomy must provide a basis of abstracting over the causal behaviour of concepts.

When comparing agents or artefacts, this causality can be abstracted out by considering the functional or behavioural commonality between target and source: a surgeon can be meaningfully described as a repairman since both occupations have the function of restoring an object to an earlier and better state; a footballer can be

meaningfully described as a gladiator or a warrior since each exhibits competitive behaviour; and a scalpel can be compared to a sabre, a sword or a cleaver since each has a cutting behaviour; and so on.

Theories of metaphor and analogy are typically based either on structure-mapping [3,4] or on abstraction e.g., [5,6,7,8,9,10]). While the former is most associated with analogy, the latter has been a near-constant in the computational treatment of metaphor. Structure-mapping assumes that the causal behaviour of a concept is expressed in an explicit, graph-theoretic form so that unifying sub-graph isomorphisms can be found between different concepts. In contrast, abstraction theories assume that analogous concepts, even when far removed in ontological terms, will nonetheless share a common hypernym that captures their causal similarity. Thus, we should expect an analogous pairing like *surgeon* and *butcher* to have different immediate hypernyms but to ultimately share an abstraction like *cutting-agent* (see [8,9]).

However, the idea that a standard ontology will actually provide a hypernym like *cutting-agent* seems convenient almost to the point of incredulity. The problem is, of course, that as much as we want our ontologies to anticipate future analogies and metaphors with these pro-active categorizations, most ontologies simply do not possess terms as prescient as these. This is the question we address in this paper: if we assume that our ontologies lack these structures, can we nonetheless enable them to be added via automatic means? We argue that we can, by looking not to purely conceptual representations but to integrated natural-language representations derived from multiple sources and multiple languages.

2 Abstraction Theories of Analogy

That analogy and metaphor operate across multiple levels of conceptual abstraction has been well known since classical times. Aristotle first provided a compelling taxonomic account of both in his Poetics (see [5], for a translation), and computationalists have been fascinated by this perspective ever since. While the core idea has survived relatively unchanged, one must discriminate theories that apparently presume a static type-hierarchy to be sufficient for all abstraction purposes (e.g., [6]), from theories that posit the need for a dynamic type hierarchy (e.g., [7, 8]). One must also differentiate theories that have actually been implemented (e.g., [6,8,9]) from those that are either notional or that seem to court computational intractability (e.g., [5,6]). Perhaps most meaningfully, one must differentiate theories and implementations that assume hand-crafted, purpose-built ontologies (e.g., [6]) from those that exploit an existing large-scale resource like WordNet (e.g., [8,9]). In the former, one has the flexibility to support as many functional abstractions like *cutting-agent* as are believed necessary, but at the cost of appearing to anticipate future analogies by hand-crafting them into the system.

This current work follows the latter course. We intend to automatically construct a new taxonomy of analogically-useful abstractions like *cutting-agent*, by analysing the semantic content of the definitions assigned to each word-sense in HowNet. Past

work (e.g., [8]) has attempted this automatic construction of *analogically-friendly* taxonomies from WordNet, resulting in an approach that is as much information-extraction from free text as it is semantic inference. This is because WordNet's glosses, unlike the semantic definitions of HowNet, are free-form sentences designed for human, rather than machine, consumption. For instance, Figure 1 below illustrates how features can be lifted from WordNet glosses to create new intermediate taxonyms, or dynamic types, from which analogies can be generated:

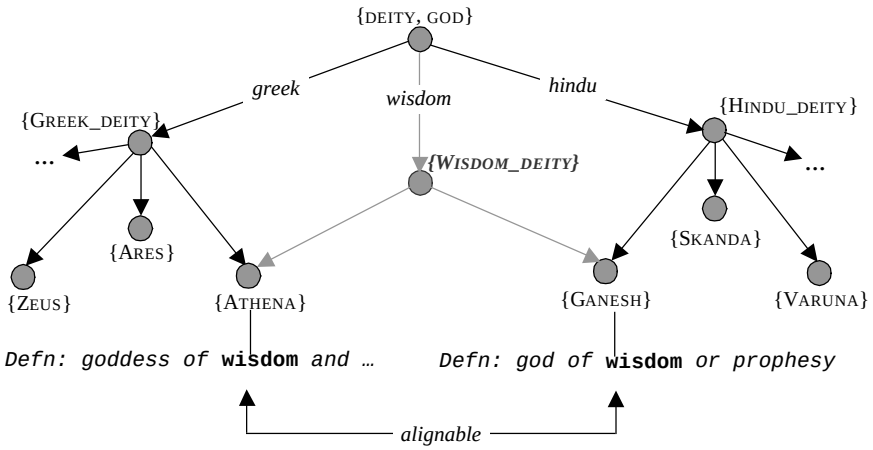

Fig. 1. Analysis of the WordNet gloss for *{Athena}* suggests that the word-form "wisdom" has analogical potential, since it is alignable with another use in *{Ganesh}*. This leads to the construction of the dynamic sense *{Wisdom_deity}* which can be used to make analogical leaps between these concepts.

The explicitly-structured semantic forms that one finds in HowNet definitions will clearly make this lifting of features more logical and less heuristic. In general, this makes HowNet an ideal knowledge-source for a computational model of metaphor and analogy (e.g., see [10] for a topical perspective).

3 Comparing WordNet and HowNet

Generalization can be considered "controlled" if, when moving to a higher level of abstraction in a taxonomy, a conceptual system is able to precisely quantify that meaning which is lost. In this sense at least, most large-scale taxonomies do not provide a significant degree of control. Perhaps nowhere is this observation more keenly felt than in weak lexical ontologies like Princeton WordNet (PWN). In PWN [1], generalization of a concept/synset does not generally yield a functional or behavioural abstraction of the original concept. WordNet's taxonomy is designed not to capture common causality, function and behaviour, but to show how existing lexemes relate to each other. For example, the common abstraction that unites

{surgeon, sawbones} and *{tree_surgeon}* is not a concept that captures a shared sense of repair, improvement or care, but *{person, human}*. To be fair, much the same must be said of other taxonomies, such as that of HowNet [1,11], a Chinese/English semantic dictionary, and even Cyc [12]. However, as we shall demonstrate, HowNet contains the necessary basis for such abstractions in its relational semantic definitions.

PWN and HowNet have each been designed according a different theory of semantic organization. PWN is *differential* is nature: rather than attempting to express the meaning of a word explicitly, PWN instead differentiates words with different meanings by placing them in different synsets, and further differentiates synsets from one another by assigning them to different positions in its ontology. In contrast, HowNet is *constructive* in nature, exploiting sememes from a less discriminating taxonomy than PWN's to compose a semantic representation of meaning for each word sense.

Nonetheless, HowNet compensates strongly with its constructive semantics. For example, HowNet assigns the concept *surgeon*|医生 the following definition:

{*human*|人:*HostOf*={*Occupation*|*职位*},*domain*={*medical*|*医*},{*doctor*|*医治*:*agent*={~}}}

which can be glossed thus: "a surgeon is a human with an occupation in the medical domain who acts as the agent of a doctoring activity." The *{~}* serves as a self-reference here, to mark the location of the concept being defined in the given semantic structure. The oblique reference offered by the tilde construct serves to make the definition more generic (thereby facilitating analogy), so that many different concepts can conceivably employ the same definition. Thus, HowNet uses the above definition not only for surgeon, but for medical workers in general, from orderlies to nurses to internists and neurologists.

4 Extracting Functional Structure

Our scheme for converting HowNet's constructive definitions into a more differential form hinges on the use of the tilde as a self-reference in relational structures. For instance, consider the semantic definition that HowNet gives to *repairman*|*修理工*:

{*human*|人:*HostOf*={*Occupation*|*职位*}, {*repair*|*修理*:*agent*={~}}}

Noting the position of *{~}*, we can infer that a repairman is the agent of a repairing activity, or in differential terms, a *repair-agent*. Now, since HowNet defines *repair*|*修理* as a specialization of the reinstatement activity *resume*|恢复, we can further establish *repair-agent* as a specialization of *resume-agent*.

	resume-agent	
repair-agent	***doctor-agent***	***amend-agent***
repairman|修理工	surgeon|医生	reviser|修订者
watchmaker|钟表匠	herbalist|药农	

Fig. 2. Portion of a three-level functional hierarchy derived from HowNet

This double layer of abstraction establishes a new taxonomy that organizes word-concepts according to their analogical potential, rather than their formal ontological properties. For instance, as shown in Figure 2, *resume-agent* encompasses not only *repair-agent*, but *doctor-agent*, since HowNet also defines the predicate *doctor*|*医治* as a specialization of the predicate *resume*|恢复.

In general, given a semantic fragment *F:role={~}* in a HowNet definition, we create the new abstractions *F-role* and *F'-role*, where *F'* is the immediate hypernym of *F*. The role in question might be *agent*, *instrument*, *location*, *patient*, or any other role that HowNet supports. By way of example, Figure 3 illustrates a partial hierarchy derived from the HowNet semantics of various form-altering tools:

	AlterForm-instrument		
cut-instrument	***stab-instrument***	***split-instrument***	***dig-instrument***
knife|刀	sword|宝剑	grater|擦菜板	scissors|剪
razor|剃刀	lance|长矛	glasscutter|玻璃刀	chainsaw|油锯

Fig. 3. a hierarchy of form-altering instruments derived from instances of AlterForm|变形状

5 Evaluating Analogical Competence

We evaluate the analogical potential of the newly derived functional taxonomy using four criteria: *topology* – the branching structure of the new taxonomy dictates its ability to generate analogies; *coverage* – the percentage of unique HowNet definitions that can be functionally re-indexed in the new taxonomy; *recall* – the percentage of unique definitions for which at least one analogy can be found using the new taxonomy; and *parsimony*– the percentage of abstractions in the new taxonomy that can be used to generate analogies.

5.1 Topological Characteristics of the New Functional Taxonomy

The new functional taxonomy contains 1579 mid-level abstractions and 838 upper-level abstractions. In total, the taxonomy contains only 2219 unique

abstractions, revealing that in 8% of cases, the upper-level abstraction of one concept serves as the upper-level abstraction of another.

Analogies will be generated only if two or more unique concept definitions are co-indexed under the same mid-level or upper-level abstraction in the new functional taxonomy. For example, knight|骑士 and gladiator.|打斗者 are both co-indexed directly under the mid-level abstraction *fight-agent*. Likewise, gladiator|打斗者 is indexed under *HaveContest-agent* via *fight-agent*, while footballer|足球运动员 is indexed under *HaveContest-agent* via *compete-agent*. The upper-level of abstraction, represented here by *HaveContest-agent*, is necessary to facilitate analogy between semantically distant concepts.

Nonetheless, we note that a certain degree of *metaphoric licence* has already been exercised by HowNet's designers in assigning semantic structures, so that even semantically distant concepts can still share the same mid-level abstraction. Creative analogies like "Death is an assassin" can, as shown in Figure 4, be understood via a single generalization.

	MakeBad-agent	
	kill-agent	***attack-agent***
	assassin|刺客	intruder|侵略者
	Death|死神	man-eater|食人鲨

Fig 4. semantic diversity among concepts with the same mid-level abstraction

Furthermore, because HowNet contains 95,407 unique lexical concepts (excluding synonyms) but only 23,507 unique semantic definitions, these definitions must be under-specified to the extent that many are shared by non-identical concepts (e.g., *cart*|*板车* and *bicycle*|*单车*, are simply defined as manual vehicles).

5.2 Analogical Coverage

Since this new taxonomy is derived from the use of *{~}* in HowNet definitions, both the coverage and recall of analogy generation crucially depend on the widespread use of this reflexive construct. However, of the 23,507 unique definitions in HowNet, just 6430 employ this form of self-reference. The coverage of the new taxonomy is thus 27% of HowNet definitions.

5.3 Analogical Recall

A majority of the abstractions in the new taxonomy, 59%, serve to co-index two or more HowNet definitions. Overall, analogies are generated for 6184 unique HowNet definitions, though these individual definitions may have many different lexical realizations. The recall rate thus is 26% of HowNet's 23,507 unique definitions, or

96% of the 6430 HowNet definitions that make use of {~}. The most productive abstraction is *control_agent*, which serves to co-index 210 unique definitions.

5.4 Parsimony of Abstraction

Overall, 1,315 of the 2219 nodes in the new taxonomy prove useful in co-indexing two or more unique definitions, while 904 nodes serve to index just a single definition. The parsimony of the new taxonomy is thus 59%, which reveals a reasonable, if not ideal, level of representational uniformity across HowNet's semantic definitions.

6 Conclusions and Future Work

While just 27% of HowNet's definitions are sufficiently structured to support analogy, we are encouraged that almost all of this generative potential can be achieved with a new functional taxonomy that is straightforward and efficient to construct. Furthermore, though 27% may seem slim, these analogically-friendly {~} structures are concentrated in the areas of the HowNet taxonomy that can most benefit from analogical re-description. As revealed in Table 1 below, some areas of HowNet are more amenable to analogical reasoning than others.

Table 1. Analogical coverage/recall for different areas of HowNet

	Humans	**Artefacts**	**Animals**
Coverage	.65	.68	.42
Recall	.54	.58	.16
Parsimony	.50	.54	.22

Clearly, however, we have just scratched the surface of what can usefully be derived from the lexico-semantic content of HowNet. For instance, our investigations with HowNet suggest that the full semantic richness of Chinese orthography may play a considerable role in supporting creative reasoning at a linguistic level, if only because it opens a window onto a different cultural perspective on words *and* concepts.

Most Chinese entries in HowNet are multi-character terms whose composite orthography affords a kind of semantic transparency that other writing systems (like that of English) do not possess. Thus, 手术刀, meaning "scalpel", is a composite not just of characters but of ideas, for 手术 means "surgery" and 刀 means "knife". Likewise, 哲学家, which translates as "philosopher", is a composition of 哲学 ("philosophy") and 家 ("specialist" or "scientist"). In turn, philosophy|哲学 is organized by HowNet as a specialization of knowledge|知识, as is logic|辩学, mathematics|数学, lexicography|词典学 and even midwifery|产科学. By

decomposing compound terms in this way and generalizing the extracted modifiers, yet another three-level taxonomy can be constructed. For instance, from these examples the partial taxonomy of Fig. 4 can be derived.

	Knowledge-instrument	
Mathematics-human	***philosophy-human***	***midwifery-human***
mathematician|数学家	philosopher|哲学家	midwife|产科

Fig. 4. Portion of an alternate three-level hierarchy derived from compound Chinese terms

The analogical potential of this ontologization becomes clear when one notices that it supports the classical analogy of philosopher as midwife.

References

1. Miller, G. A.: WordNet: A Lexical Database for English. Communications of the ACM, Vol. 38 No. 11 (1995)
2. Dong, Z.: Knowledge Description: What, How and Who? *The Proceedings of the International Symposium on Electronic Dictionaries*, Tokyo, Japan (1988)
3. Falkenhainer, B.; Forbus, K.; and Gentner, D.: Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, 41, pages 1-63 (1989)
4. Veale, T., Keane, M. T.: The Competence of Sub-Optimal Structure Mapping on 'Hard' Analogies. The proceedings of IJCAI'97, the Int. Joint Conference on Artificial Intelligence, Nagoya, Japan. Morgan Kaufman, San Mateo California (1997)
5. Hutton, J.: Aristotle's Poetics. Norton, New York (1982)
6. Fass, D: An Account of Coherence, Semantic Relations, Metonymy, and Lexical Ambiguity Resolution. In: Small, S. I, Cottrell, G. W., Tanenhaus, M.K. (eds.): Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology and Artificial Intelligence. Morgan Kaufman, San Mateo California (1988)
7. Way, E. C.: Knowledge Representation and Metaphor. Studies in Cognitive systems, Kluwer Academic Publishers (1991)
8. Veale. T.: Dynamic Type Creation in Metaphor Interpretation and Analogical Reasoning: A Case-Study with WordNet. In the proceedings of ICCS2003, the 2003 International Conference on Conceptual Structures, Dresden, Germany (2003)
9. Veale, T.: WordNet sits the S.A.T.: A Knowledge-Based Approach to Lexical Analogy. *The proceedings of ECAI'2004, the 16th European Conf. on Artificial Intelligence*. John Wiley: London (2004)
10. Veale, T.: Analogy Generation in HowNet. In the proceedings of IJCAI'05, the 19th International Joint Conference on Artificial Intelligence. Morgan Kaufmann: CA.
11. Wong, S.H.S.: Fighting Arbitrariness in WordNet-like Lexical Databases – A Natural Language Motivated Remedy. *The proceedings of GWC 2004, the 2nd Global WordNet conference*. Edited by Sojka, Pala, Smrz, Fellbaum, Vossen (2004)
12. Lenat, D., Guha, R.V.: Building Large Knowledge-Based Systems. Addison Wesley (1990)