LNAI (Lecture Notes in Artificial Intelligence), 2464, pp. 20-27, 2002.

RADAR : Finding analogies using attributes of structure

Brian P. Crean[1] and Diarmuid O'Donoghue[2]

[1]Galway-Mayo Institute of Technology, Castlebar Campus
Westport Road, Castlebar, Co.Mayo, Ireland also [2]
Brian.crean@gmit.ie
[2]Department of Computer Science, NUI Maynooth,
Co. Kildare, Ireland.
Diarmuid.odonoghue@may.ie
http://www.cs.may.ie/~dod/

Abstract. RADAR is a model of analogy retrieval that employs the principle of systematicity as its primary retrieval cue. RADAR was created to address the current bias toward semantics in analogical retrieval models, to the detriment of structural factors. RADAR *recalls* 100% of structurally identical domains. We describe a technique based on "derived attributes" that captures structural descriptions of the domain's *representation* rather than domain contents. We detail their use, recall and performance within RADAR through empirical evidence. We contrast RADAR with existing models of analogy retrieval. We also demonstrate that RADAR can retrieve both semantically related and semantically unrelated domains, even without a complete target description, which plagues current models.

1. Introduction

An analogy is a structure-based comparison between a problem domain (the target) and some known domain (the source). Existing similarities form the basis for transferring additional information from the source to the target domain. *"So, IF the nucleus is like the sun and the planets orbit the sun, THEN the electrons must also revolve about the nucleus"*. Yet much research has focused on interpreting predetermined analogies rather than discovering new candidate analogies.

Analogy retrieval is an essential yet complex process, where numerous similarity constraints (semantic, structural and pragmatic) can be considered in judging the usefulness of a source/target pairing [1]. The retrieval process is further complicated when one considers that cross-domain analogies need not share superficial semantic features between the participating domains. Structural similarity has received a great deal of acceptance in analogical mapping, yet has had startlingly little impact on analogy retrieval models. We believe that it is not unreasonable to include structural considerations within the retrieval process itself. Systematicity [2] would allow a deliberate broadening of knowledge to reach retrieval and subsequent stages, and thereby form new analogies

Frequently, domains lacking obvious semantic resemblance can be used to form useful analogies [3], for example *"Nuclear war is like a game of tic-tac-toe*" (adopted from the motion picture "War Games"). Taking a recognised domain like tic-tac-toe that frequently ends in a draw (no winner) can lead to recognising the futility of nuclear war i.e. no winner. In order to interpret a possible candidate analogy we first must discover it. Yet in current models, retrieval is governed by the semantic content (meaning) of entities common to both analogs. Such systems can only retrieve domains that are represented similarly - and the retrieval of structurally similar domains is purely accidental. It is merely a happy coincidence that any semantically similar domains ever have the same structure as each other. A crucial consequence of these inflexible retrieval models, is their inability to explain the role of analogy in creativity, inspiration and insight as identified by Boden [4].

Initially, we want to retrieve a *structurally* identical source domains, given some target domain problems. However, analogies are typically formed between a smaller target domain and a larger source - since the source must supply additional information to the target domain. Therefore, not only are we looking for *isomorphic* source domains, but we also wish to identify *homomorphic* sources.

These factors will undoubtedly increase the possibility of retrieving inappropriate source domains. Current retrieval models reject many inappropriate sources by their inability to form a large mapping. However, our technique could use the analogical validation phase to reject inappropriate comparisons. A model of structural retrieval would also have to operate in a computationally tractable manner. In this paper we introduce the RADAR (Retrieving Analogies utilising Derived AttRibutes) model and its method of encoding structural traits of domains through derived attributes. We detail its operation, and how it improves over current analogy retrieval models.

2. Existing Analogy Retrieval Models

Before we discuss models of analogy retrieval we briefly describe our objectives from an information retrieval perspective. As we shall see, existing models have very poor *recall,* because they either ignore or severely dis-favour semantically distant domains. Thus, they are impotent from the perspective of generating creative analogies. Furthermore, the *precision* of some models is very poor because semantic similarity is independent of the structural similarity required to form an analogy. We wish to explain how structurally similar domains, even ones that are semantically un-related to the target, might be retrieved. Such comparisons underlie many breakthroughs in science - "*the heart is a pump, the brain is a telephone networks, light is a wave vs. light is a particle"*. Crucially, no current model of analogical retrieval can explain how a target domain can cause the retrieval of a semantically unrelated source. Consider the following target sentence (taken from [5]), where Felix is a cat and Mort is a mouse.

T_1 : "*Felix bit Mort causing Mort to flee from Felix"*

We wish to find domains that are *structurally* similar to T_1, so that we can later identify any cross-domain analogies. So, for the given source we might wish to identify a candidate source domain such as $S_{2,}$ where Mary is a woman and Frank is a man**.**

S_2 : *"Mary seduced Frank causing Frank to kiss Mary"*

Of course, most cross-domain analogies will be invalid but we must at least have the potential to retrieve them. Identifying the useful analogies from these cross-domain comparisons is the task for the later *mapping* and *validation* phases.

2.1 Analogy Retrieval Models

With domains described by content vectors, MAC/FAC [6] creates a skeletal description by counting the number of times distinct predicates and objects occur in a representation. Estimating the structural overlap between source and target is presented by a dot product computation on the respective content vectors. The highest dot product and all those within a 10 % range are deemed structurally similar. Crucially, MAC/FAC cannot retrieve semantically distant domains due to weak predicate and object similarity in the content vectors. Its frailty is also evident when presented with a partial source and target description. Consider S_3 (additional information to S_2)

S_3 : *"Frank and Susanne are married. Mary seduced Frank causing Frank to kiss Mary, so Susanne divorced Frank"*.

The respective T_1/S_3 content vector is weakened due to again the non-identicality of the domains on which MAC/FAC is dependant, signifying it is vulnerable when dealing with partial descriptions.

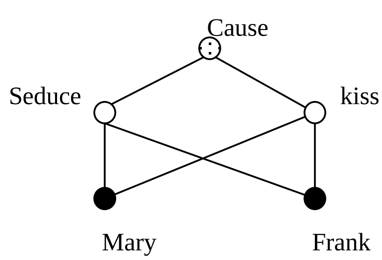
ARCS [7] uses tokenised semantic similarity between representations as pre-selection criteria in the creation of a parallel constraint satisfaction network. Structural correspondence is determined on the basis of isomorphic similarity between source(s) and target predicate arguments. ARCS uses a standard parallel connectionist relaxation algorithm to allow the network to settle. After the network settles sources are deemed suitable based on their activation strength. ARCS perseverance with semantic similarity as a filtering process means it casts semantically disparate domains aside from the outset, i.e. Felix the cat is not similar to Frank the man. In relation to the original T_1 and $S_{2,}$ ARCS immediately rejects any possible retrieval due to weak semantic correspondence, the only recognised similarity is the higher-order predicate, *cause*. This indicates that ARCS is unable to retrieve semantically distant domains.

HRR's [5] combine semantic and structural similarity with an inclusive vector representation. Semantic identity is encoded into each entity in a representation. The structural traits of the representation are brought to the surface as role filler assignments (e.g. agent, patient) and are encoded along with each entity or predicate. HRR's structural retrieval is diminished by its tendency to favour more semantically similar source(s). Consider T_1 and S_2, the absence of semantic similarity between objects and predicates in both source and target, severely limits the possibility of semantically distant domains being retrieved (also borne out by Plates original results [5]). The resultant "convolution" product (used by HRR's to estimate similarity) is considerably weak. HRR's suffer from a similar affliction as MAC/FAC in its inability to tolerate the loss of significant information. If presented with T_1 and S_3, the absence of identical structures between representations will weaken the measures of similarity even further. Though HRR's attempt to combine semantics and structure in retrieval, problems with non-identical structure and semantics are very problematic to HRR's.

Though CBR (Case-Based Reasoning) is similar to analogy retrieval in many respects, we exclude CBR from discussion in this paper as CBR is primarily concerned with intra-domain retrieval and is not concerned with retrieving semantically distinct domains.

3. Derived attributes

One method to effect structural-based retrieval is to re-describe the domain description via micro-features. We use *derived attributes* to determine the structural characteristics of a domain representation. Derived attributes describe features of the representation itself, rather than qualities related to the real world. These structural attributes are not tailored to suit individual domains but are used to describe any conceivable domain. Each derived attribute describes a particular construct of the domain through a calculated value. A number of simple derived attribute types (Fig 1) can be used to capture the structural characteristics of the domain's representation. Derived attributes such as the number of predicates and objects in a domain can convey simple structural detail. The identification of cyclic paths (loops) in the representation is also beneficial i.e. in fig 1, the path containing the entities *cause, seduce, frank* and *kiss* is an instance of a loop as it cycles back to the starting point. These derived attributes can distinguish between identically structured domains and similarity-structured domains - an important distinction as already highlighted in the operation of current models. For illustrative purposes the predicate representation of, S_2, *cause (seduce, kiss), seduce(mary, frank), kiss (frank, mary)* can be described by the derived attributes in fig 1.



Derived Attributes

Number of Loops : 3
Loop size : 4
Connectivity : 3
Number of predicates : 3
Number of atoms : 5

Fig. 1. Derived attribute description of S_2

The derived attributes depicted in fig 1 are just a subset of the structural composition of domains, other derived attribute types such as *number of roots*, *longest path* etc. can also be incorporated. These structural attributes are independent of any semantic primitive descriptions housed in these domains and hence cannot be influenced by semantic considerations.

4. RADAR

We now describe RADAR (Retrieving Analogies using Derived AttRibutes), which bases analogy retrieval in the 'search space' of structural attributes. The premise of RADAR's retrieval algorithm is that domains with the same derived attributes must also have identical structure. Each domain in long-term memory is examined for its structural description, which generates various derived attributes types and their values. These values are stored in derived attributes stores and linked to the source in long-term memory.

When presented with a target analog, the target is also rendered into its own structural attributes, and the target's derived attributes activate the corresponding stores. All concepts in memory that have the same attribute value for a particular type will be identified through spreading activation. This indicates they are structurally identical for each structural feature. Activation values discriminate the strongest source(s) from dissimilarly structured analogs through a threshold level that rejects any concept below a certain similarity measure. The stronger the activation value, the more structurally identical a candidate source is to a target.

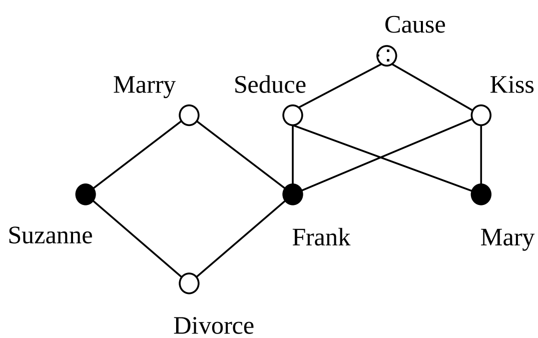
4.1 Identical Structure Retrieval

Presenting S_2 (source) and T_1 (target) from above, we demonstrate how RADAR retrieves identical structures. Retrieval is driven by the derived attributes of T_1, so any source that shares a corresponding derived attribute value will receive activation. RADAR successfully retrieves S_2 with and overall similarity metric of 100%. This 100 % signifies that both source and target share all derived attribute values and hence are structurally identical. This was as expected because the source representation is isomorphic with the target. Current models rely heavily on semantic similarity to guide retrieval but this demonstrates retrieval where semantic similarity is negilable, consequently are constrained by the targets semantic influence and unable to retrieve this valid analogy.

It also demonstrates RADAR's ability to retrieve creative analogies such as the *"Nuclear war is like a game of tic-tac-toe",* which again current models cannot retrieve. Identical results were obtained when presented other structurally identical domains that *do* share some semantic similarity, again Plate's "Spot bit Jane causing Jane to flee from Spot", Gick and Holyoak's [8] surgeon/general analogy, etc. These domains contain identical structural and RADAR, as expected, retrieves the identical sources based on identical structural attribute values.

4.2 Partial Structural Retrieval

This demonstration examines RADAR's ability to retrieve useful sources from memory, where the structural overlap between source and target is incomplete. (missing predicates and objects). This is vital as analogies are frequently used for learning, requiring that the source have more structure than the target - and consequently a different structure to it. Each source is broken down into its derived attribute pairings (fig 2) and stored in structure memory. RADAR again creates a basis for retrieval by analyzing the target, T_1 for its derived attribute values. Again, RADAR successfully retrieved the appropriate source, S_3, though with a smaller similarity scoring. As we can see from fig 2 the structure of the analogues do not match exactly, and this is reflected in the derived attribute values (again fig 2). But a subsection of the representations do share identical structure - the loops structure. Both share three loops of size four. The loop structures identify a structural similarity and brings it to the surface.



Derived Attributes

Number of Loops : 4
Loop size : 4
Connectivity : 4
Number of predicates : 5
Number of atoms : 8

Figure 2: Structural representation of the S_3's Love Triangle Story

We readily accept the argument that if more information were missing then retrieval accuracy would decay. This is of course a fact of retrieval, poor representations lead to miss-guided retrieval (if any). Remove the predicate *Divorce (Susanne, Frank)* from the representation, and the resultant derived attribute values would reflect this change. But we would argue that there does come a point in reminding, in the human cognitive process and cognitive modeling, when significant information is missing, retrieval will tend to be poor or not take place at all. This is perfectly analogous to the existing situation where semantically based retrieval performs poorly with missing information - but cannot retrieve outside its own domain. Likewise if more predicates or objects were added to the target, i.e. the proposition *move-in-with (Frank, Mary)* then the structure changes, but again there is a common substructure (loop structure). This experiment confirmed that RADAR operates successfully when presented with partial domains, on the provision that there exists some coherent structure between the representations. RADAR is the only model that considers partial source/target parings in retrieval.

5. Performance and Future Work

RADAR's overall performance was examined with a long-term memory containing frequently cited domains, chosen randomly from the analogy literature. In all seventy domains were stored. Domains were of varying complexity with an average of 8 predicates (ranging from a minimum of 1 to a maximum of 25 predicates) and 14 entities (also ranging from a minimum of 3 to a minimum of 39 entities) per domain. In the investigation, the largest loop construct considered was six. Selection is based on the highest scoring source domain, or group of sources joint highest retrieval score. Retrieval was classified as successful if the target caused the corresponding source identified in the literature to be retrieved.

RADAR retrieved the common matches on each occasion and significantly out performed other models in its ability to retrieve appropriate candidate source domains. RADAR retrieved an average of 4 sources (6%) when presented with a target. In each case the correct source was amongst the joint highest active sources. Significantly, RADAR can retrieve similar structured source when presented with identical and partial representations even when they share no semantic overlap between objects or predicates, where other models are deficient.

	MAC/FAC	ARCS	HRR	RADAR
Identical Structural Retrieval	Yes	Yes*	Yes +	Yes
Partial Structural Retrieval	No	Yes *	No	Yes
Identical Semantic distant Domain retrieval	No	No	No	Yes
Partial Semantic distant Domain retrieval	No	No	No	Yes

Table 1. Comparison of RADAR against common retrieval models

* on the provision that pre-selection will have semantic information
\+ retrieval accuracy will vary considerably from target

Derived attributes can be manipulated in order to re-describe the structural description of a representation and increase performance of the retrieval process. Similar to *weighted features* [9], where feature descriptors are weight based on their importance or usefulness, certain derived attributes can be marked as more relevant than others. Alternatively a nearest neighbour algorithm can be used to locate similar sources [O'Donoghue, Crean, in press]. Another technique is to simply increase the number of derived attributes used to describe a domain.

Domain retrieval using derived attributes is only as efficient as the derived attributes that supplement the raw domain information. Taking just five attribute types each with just 10 values, and making the best case assumption that our data is distributed evenly along each value, then each location in derived attribute space would represent just 10 domains, for a base of 1,000,000 domains. This indicates the potential retrieval power of derived attributes. Of course, efficiency is increased with additional attribute types and values describing new structural qualities with particular utility to analogy retrieval.

6. Conclusion

We lay no claim that this is how structural retrieval is performed in human cognition. The focus of this work was on creating a computational model that is capable of structural domain retrieval in a computationally tractable manner - which overcomes the semantic restriction suffered by other models. We have demonstrated the ability of derived attributes in describing the structural make-up of domains. We then demonstrated their ability to retrieve semantically related and un-related domains, whether presented with a partial or complete target domain. We detailed these findings through the recall and precision performance of RADAR, using commonly cited domains from the analogy literature. RADAR successfully overcomes the limitations suffered by current retrieval models.

References

1. Eskeridge, T.C. “A Hybrid model of continuous analogical reasoning”, In Branden (ed.), Advances in Connectionist and Neural Computation Theory, Norwood, NJ: Ablex, 1994
2. Gentner D “Structure-mapping: A theoretical framework for analogy”, Cognitive Science, volume 7, pp 155 –170, 1983.
3. Kolodner, J. L. “Educational Implications of Analogy a view from Case-based Reasoning”, American Psychologist, pp 57 - 66 Volume 52, 1997.
4. Boden, M. A. “The Creative Mind”, Abacus, 1994.
5. Plate Tony A., “Distributed Representations and Nested Compositional Structure”, Graduate Department of Computer Science University of Toronto, 1994 Ph.D. Thesis.
6. Gentner D, Forbus D., “ MAC/FAC : A model of Similarity-based Retrieval”, pp 504 - 509, Proceedings of the 13th Conference Cognitive Science Society, 1991.
7. Thagard P., Holyoak K. J., Nelson G, Gochfield D., “Analog Retrieval by Constraint Satisfaction”, pp 259 - 310, Artificial Intelligence Volume 46, 1990.
8. Gick M. L. and Holyoak K. J., “Analogical Problem Solving”, Cognitive Psychology, volume 12, pp 306 – 355, 1980
9. Blum A. L., Langley. P. “Selection of relevant features and examples in machine learning”. Artificial Intelligence volume 97, pp 245--271, 1997.