European Conference on Artificial Intelligence (ECAI) Workshop on Creative Systems, Lyon, France, 21-26 July 2002.

Searching for Serendipitous Analogies

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

Abstract. Analogical reasoning is an acknowledged process behind many episodes of creativity. Typically, the creator chances upon information unrelated to the given problem – and solves the problem by analogy with this accidental source of inspiration. Current models of analogical retrieval do not explain how semantically unrelated source domains are retrieved. We present the RADAR algorithm that maps domains into a separate structure space, where domains with similar topological attributes are co-located. Each axis in structure space records the occurrence frequency of that feature in each domain. Nearest neighbour retrieval in structure space identifies structurally similar domains - from a diversity of semantic backgrounds. Structure based retrieval opens the possibility for creating an analogy model with far greater creativity potential than human reasoning.

1 INTRODUCTION

Analogy plays an acknowledged role in many creative episodes, and much work has been carried out into modelling the analogy process. While work has been carried out on discovering analogies, none appears to have been carried out into discovering *creative* analogies. In this paper we describe a computational model capable of acting as the driving force behind the discovery of creative analogies.

In this paper we assume that we have some given problem domains. This domain is described by a collection of predicate calculus assertions, involving objects, first-order rations and high-order causal relations. We also assume that some background memory exists with a large number of other domains, in which we base our search for creativity. The task is to select all domains with the potential for forming an analogy with the given target. Additionally, we wish to support the semantically distant comparisons that are common to so many creative analogies.

First, we set the background for this project. There is a vast difference between *interpreting* a given creative analogy and *discovering* a new creative analogy. Interpreting a supplied analogy does not require significant "exploration", as a good analogy must be relatively easy to interpret - even if its implications are quite profound and far reaching. However, when discovering a creative analogy, we rely on the serendipity explanation - that it was a chance happening upon the source domain that enabled the analogy.

Semantic based retrieval uses the problem (or target) domain's description to identify previous domain description. Typically, these candidate sources share some content in common with the target problem. This direct semantic similarity may be augmented with a semantic lexicon, like WordNet, enabling retrieval via synonymous descriptors. However, this approach doesn't explain why serendipity plays such an important part in many creative episodes. Nor does it explain how an algorithm might go about deliberately searching for serendipitous analogical comparisons.

Consider the problem of generating the famous *solar-system:atom* analogy. The two domains originate in different disciplines (nuclear physics and astro-physics). Thus, the vocabulary used to describe each is very different, but must still be retrieved by a creativity model. For example, the target relation between the nucleus and the electron is the "electromagnetic attraction", and Rutherford would most likely have thought of it as such. (The distinction between the four "fundamental forces" being a core distinction in physics). The corresponding relationship between source's planet and electron is "gravity". It is only after we have identified the analogy that we can generalise these relations and identify the "attracts" relation in each domain.

This example again highlights the difference between generating and interpreting an analogy. Generating an analogy must counteract the representational "vocabulary" problems. For example, WordNet does not currently contain specific relations for gravitational-attraction and electromagnetism, which would be essential in modelling a physicist's expert knowledge. Thus, seeing gravity and electromagnetism as instances of attraction is arguably more a result of this analogy than a driving force behind its discovery.

In contrast, conveying a newly discovered analogy can be done in such a way as to highlight the newly discovered similarities. When conveying this analogy the term "attracts" can be used when describing each domain. This makes the processes of both remembering and interpreting the analogy much simpler.

While some creative analogies result in significant new discoveries in science or other disciplines, many less profound but still creative analogies are invented every-day. Indeed, for tutoring and instruction it is often vital that the source domain is semantically unrelated to the given problem, allowing teaching to be based in a domain that is familiar to the student. Coaches, instructors and tutors frequently invent and use analogies between semantically different domains (ie between-domains analogies). Examples include "golf-putting is like standing on train tracks" [1], "driving a race-car is like ice skating", and "programming a computer is like writing a film script". These between-domains comparisons are often the stock-in-trade of instructors, who regularly use them so that student's will come to see the 'something strange' as being 'something familiar'. The semantic overlap between these analogies is minimal, and retrieval based on semantic similarity seems inadequate in explaining these regularly occurring creative analogies.

Basic metaphors that frame much of our understanding typically involve abstract source domains - the future is a container, an argument is a war *etc*. Identifying the source for each

[1] Department of Computer Science, National University of Ireland, Maynooth, Co. Kildare, Ireland.
[2] Department of Information Technology, Galway-Mayo Institute of Technology, Castlebar, Co. Mayo, Ireland.

target requires using structure to find the framing source - relying on semantics for retrieval is too constrained.

Semantic retrieval ignores the key role that *structure* plays in analogy. Structure should be the basis for retrieval, not semantics. In this paper we describe an algorithm that efficiently performs structure-based retrieval on analogy domain descriptions.

2 ANALOGY AND CREATIVITY

In this section we briefly describe analogy and the analogical reasoning process. We assess its role in some well-known examples of creative scientific reasoning. Then we examine the notion of creativity, and why we believe the RADAR (Retrieving Analogies with Derived AttRibutes) model can be considered an engine for creative scientific discovery.

Analogy is a form of reasoning that identifies, and extends, the structural similarity between two domains of information. Kekule acknowledged the key role of analogy behind his creative "invention" of organic chemistry, and later his creative "invention" of aromatic chemistry. Indeed, Hoffman [2] describes many famous scientific discoveries either driven by, or widely understood as analogical comparisons.

Every analogy juxtaposes two key collections of information - the source and the target. In scientific creativity the target (problem) is known *a priori*, though unlike the analogies described in much of the computational modelling literature, it typically contains much irrelevant information.

Combined work across the branches of cognitive science in recent years has greatly improved our understanding of this core cognitive process. In this paper we present a computational model that efficiently identifies creative candidate sources when presented with a target problem.

Top-down and Bottom-up

Many of the major discoveries in science, especially the physical sciences, are driven in the bottom-up manner. Newly discovered facts contradict the predictions of an existing theory, thereby calling for a new all-encompassing theory. Observing the behaviour of smoke particles led to the analogy that gasses are like billiard balls. Kekule's well-documented carbon ring analogy was driven by $C6H6$ (and other molecules) whose behaviour contradicted the existing carbon chain theory.

Analogy also plays a role in "top-down" scientific creativity, by enabling theoretical advances. This form of creativity generally suggests novel inferences that might require experimentation to verify. Einstein's discovery of relativity had to wait until detailed observations of a solar eclipse were made, before his theory was validated (laboratory experiments being all-but impossible in astrophysics). Of course these approaches often proceed in tandem - involving iterative combinations of either (or both) bottom-up and top-down flavours. Kekule's carbon ring involved two distinct discoveries - the ring structure and also the carbon double-bond [3].

Creative Limitations

While a source & target comparison may generate a novel interpretation, each source domain implicitly creates its own conceptual limitations. Each analogy has its own focus and its own blind spots. Analogical comparisons create not only novel interpretation and inferences, but also they implicitly take attention away from competing theories. The very presence of a valid mapping further inhibits our search for a new comparison, so every breakthrough also represents a new block to subsequent creativity.

Because of the nature of analogy, you cannot separate the mapping from its implications. Likewise, it is impossible to separate the inspiration behind creativity from the limitations of that creative interpretation - however irrelevant those limitations may appear at the time. For example, the limitations of Newtonian mechanics did not become clear until Einstein's breakthrough.

It has been shown that people do not spontaneously recognise analogies, even when all information is available to them. In tests on human subjects Gick and Holyoak [4] showed that only when people are explicitly prompted to identify an analogy does solution rates jump from about 10% to around 90%. For this reason we believe that computational models have greater creative potential than people. Computers can be forced to search out creative sources, forced to look for alternate mappings and examine their inferences. In this paper we attempt to create the driving engine for a creativity machine that is capable of generating analogies that people could use to re-interpret any given problem domain.

2.1 Scientific Creativity

Boden [5] identifies 2 classes of creativity; *P-creative* ideas are new to the individual agent generating the novel idea whereas *H-creative* ideas are historically new to all reasoning agents. Without a knowledge base containing all known ideas in history, any search for H-creative knowledge must be based in some psychological knowledge base - that is, a search for p-creativity. In RADAR the hunt for creativity is rooted in a memory of domain descriptions, where each domain is a collection of predicate calculus assertions. While RADAR searches for p-creative inferences, its architecture is centred on identifying inferences that are more likely to be h-creative. This issue will be developed in later sections.

Ritchie [6] identifies the essential qualities of creativity as *directed, novel* and *useful.* We describe the RADAR model as *directed,* as its creativity is based on some given target problem for which we seek a new interpretation. All RADAR's activities are driven by the target problem. RADAR's output is *novel* as it locates a previously unidentified source comparison, that generates p-creative inferences that are new to the target domain. Finally, its output is potentially *useful* as the identified source makes novel predictions that are cognitively plausible about the target problem.

Colton and Steel [7] identify justification of findings as a key requirement for computational creativity. In RADAR, justification comes in two forms, both relating to the phases of analogy known as mapping and validation. An analogy is only identified when a sufficient proportion of a domain's elements participate in the inter-domain mapping. Secondly, the inferences mandated by such a sufficiently credible comparison are subject to validation. At the lowest level validation can ensure that inferences do not directly contradict known facts - any analogy mandating such inferences can be immediately rejected. However, the focus in this paper is on improving the recall of analogical retrieval models. Our concern is that we identify all plausible sources of creativity, identifying domains with the potential of forming an analogy with the supplied target.

Boden [8] identifies three levels of creativity: *improbable, exploratory* and *transformational*. The first creates novel combinations of familiar ideas. At the more advanced end of this spectrum, transformation causes "the shock of amazement" at a fundamental restructuring of ideas. Interestingly, she places analogy into the improbable category, not the transformational. To place analogy in this category is to imply that *no* analogy is exploratory or transformational.

Keklue's 1855 vision of two carbon atoms "dancing in the street" was revolutionary. Not only did this dancing analogy resolve the conflict caused by many carbon compounds, but it simultaneously invented organic chemistry. Faraday discovered benzene in 1826, but its behaviour contradicted expectations. Kekule's "carbon-chain is like a snake biting its tale" analogy (*ie* carbon ring) resolved this conflict in 1865, and simultaneously invented "aromatic chemistry".

It is difficult to see these analogies as anything other than transformational. The first analogy introduced the revolutionary carbon-carbon bond (previously atoms only connected with non-identical atoms). The second introduced alternating carbon-carbon double bonds - a key structural element to DNA's double helix structure. Einstein said that much of his theory of relativity can be derived by considering an analogy between distance and the time taken to travel between logs floating up and down stream.

3 ANALOGY MODELS

An analogy is a comparison between two domains of information (called the *source* and *target*) that supports learning. Gentner [9] highlighted that the key to interpreting analogies lies in aligning the *structure* of these two domains - and in extending the observed similarities to generate new information. Analogy researchers subdivide the process into several distinct phases, and where individual differences remain, Keane *et al* [10] represents a typical model composed of: *representation, retrieval, mapping, adaptation* and *induction*. This decomposition has facilitated computational modelling of individual phases, as well as the entire process.

We argue that it is the *retrieval* phase that provides the key to discovering creative analogies. Retrieval identifies the source domain, which in turn acts as the driving force behind the creative inspiration. Mapping and subsequent phases act as filters on the domains that were initially identified by retrieval. The remainder of this paper is devoted to the analogical retrieval phase, and in creating a model that is capable of identifying "creative" source domains for any given target problem.

Information Retrieval

Before proceeding to examine existing analogy retrieval models and their limitations, we first examine metrics used to describe information retrieval algorithms. Information retrieval algorithms are frequently described in terms of two key qualities: *recall* and *precision*. *Precision* is defined as the ration between the number of relevant domains retrieved, divided by the total number of domains retrieved. *Recall* is defined as the number of relevant domains retrieved divided by the total number of relevant domains in the long-term memory.

Ideally we would like both precision and recall of our analogy retrieval model to be 100%. But, from a creative perspective it is recall that is most important. We want to ensure that RADAR supports maximum creativity, even if this entails that many inferences are not creative. The post-retrieval phases identify the mapping and inferences, and these must assume responsibility for rejecting all the fruitless comparisons – thereby improving precision.

We now look at three existing analogy retrieval algorithms, and how each retrieves candidate sources when presented with some target problem. It should be noted that the objective of the following models is to identify candidate sources as opposed to *creative* candidate sources. Thus, the candidate sources RADAR seeks are more general than the other models.

MAC/FAC [11] uses normalised content vectors to represent a domain's contents. Retrieval in MAC/FAC is based on the dot product of two content vectors, and so is incapable of semantically distant retrieval. This two-stage algorithm firstly identifies potentially useful domains, before computing structural similarity. Thus the recall of MAC/FAC is poor.

ARCS [12] uses WordNet to allow synonyms to influence retrieval, and thus would have higher recall than MAC/FAC. This greatly increases the range of candidate sources that are considered as possible sources. However, distant domains are frequently described using a different vocabulary and thus ARCS would still not retrieve some semantically distant domains. Indeed, analogies are often said to invent similarity between domains, rather than relying on pre-existing similarity. Thus, while recall in ARCS improves on MAC/FAC it is not designed to address the problem of semantically distant retrieval.

Plate's [13] Holographic Reduced Representation (HRR's) use common attributes of domain objects as a basis for retrieval. Domains are identified when their objects share common attributes. Interestingly, HRR's also includes structure in the retrieval process, by introducing role-filler bindings into the representation. However, recall on semantically distant candidate sources with "missing" causal and other relations (*ie* creative analogies) will still result in poor recall for HRR's.

4 THE RADAR MODEL

We now examine the structure of the RADAR model for structure-based retrieval. We see each phase of analogy, subsequent to retrieval, as acting as a filter on the information passed to it. Successfully identifying an inter-domain mapping, causes the inference and validation process to be invoked. Thus, the creativity of an analogy model depends directly on the creativity of the initial retrieval phase. Any model that fails to identify potentially useful candidate source cannot be considered a useful model of analogical creativity. A useful retrieval tool capable of identifying creative analogies must support the following:

i. Retrieve semantically distant domains
ii. Favour structurally similar domains
iii. Allow the (possible) inclusion of semantic, pragmatic and other factors in retrieval
iv. Operate on a large memory containing a many domain descriptions from a variety of disciplines.

We adapt the five-phase model of analogy [10], by expanding the retrieval phase. To favour creative analogies, we must use structure (not semantics) as the basis for retrieval – the means of performing structure-based retrieval are explained in the next section. While this approach will inevitably identify many fruitless sources (thereby affecting precision) it will enable 100% recall.

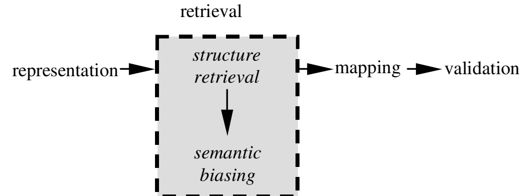

Figure 1: A two-part model of Analogical Retrieval

The architecture of an analogy model encompassing structure-based retrieval is displayed in Figure 1. This details the expanded model of the retrieval phase. Semantic biasing is not currently implemented, but could easily be added as an ordering mechanism for the selected domains.

From a pool of structurally similar candidate sources, we may wish to begin with semantically similar domains – though this has not been implemented in the RADAR model. This structure ensures we identify semantically disconnected sources, but can still include a semantic biasing if required. Finally, we point out that this is not intended to directly model human behaviour, but rather it serves as a means to implement a computational model with greater potential for generating creative analogies.

Introducing a semantic similarity *threshold* could help to improve the precision but reduce recall - effectively achieving similar recall to existing models. Hence, we do not intend to introduce a semantic similarity threshold to RADAR.

Clearly such a profligate retrieval strategy places great strain on mapping and other phases. Many structurally similar domains must be rejected either by virtue of a small inter-domain or by identifying the invalid inferences generated. RADAR forms part of a larger project that identifies many invalid inferences - however, we shall not discuss the validation phase further herein.

4.1 Attributes of Structure

We introduce the concept of "attributes of structure". These are simple numeric attributes representing structural qualities of the *representation* of each given domain. Rather than representing attributes of the content, we represent attributes of the *representation* itself. This abstracts away from the problem data, and allows us reason about the structure independently from the content. For example, we might count the number of predicates found in each domain and thereby favour similarly proportioned candidate sources. Other structural attributes then represent different qualities derived from the representation.

All candidate sources, together with the target domain, are mapped into this N-dimensional structure space. Domains that are structurally similar are mapped into similar regions of this n-dimensional space. We can then use retrieval algorithms to identify candidate sources that are located in the vicinity of the target domain within structure space. Of course the actual distance varies depending upon the set of structural attributes employed, but similarly structured domains are generally located in the same regions of structure space.

An important factor from a computational creativity perspective, is that structure based retrieval is *oblivious* to the target domain semantics. Thus, retrieving a semantically dis-connected domain is equally as likely as finding one related to the presented target. We have therefore overcome the overbearing semantic restriction repelling semantically distant sources. In practice, typical retrieval episodes randomly intersperse semantically related and disconnected domains - as dictated by each sources structural similarity to the given problem. Of course, pure random selection of candidate sources also does this, but our algorithm ensures that the candidate sources are *structurally* capable of forming analogy with the given target. Thus we can compute the candidate inferences with the standard "pattern completion" algorithm for analogical inference.

Let the structure index of the target domains be: *ta1* for target attribute1, *ta2* for target attribute 2. Similarly let the structure index of each candidate source be: *s1a1* for structural attribute 1 of source domain 1, *s2a2* for structural attribute 2 of source domain 2 etc. The distance between the target and any source is calculated by the following equation:

$$d = \sqrt{(ta1 - s_x a1)^2 + (ta2 - s_x a2)^2 + \ldots} \qquad (1)$$

Diagrammatically, we can think of each domain occupying a location in structure space. Domains close to the target share similar structural features with it. Distant domains then have little structural similarity with the target – regardless of the semantic content of that domain description.

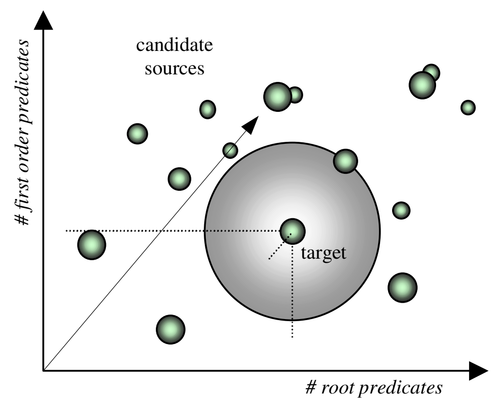

Figure 2: Nearest Neighbours in Structure Space

Structure space represents an abstract region in which we can reason about domains of information. In structure space we are free from the semantic contents of each description, and can examine the similarities and differences between domain descriptions.

4.2 List of Structural Attributes

Now we list the structural attributes actually used in the RADAR model. The derived attributes of topography used by RADAR are:

1) Number of *first-order* predicates - specifying associations between objects. Useful high-order relations include `cause`, `result-in`, `prohibits`, `xor`, `but-not` and `neither`. Because analogies frequently introduce causal predicates (high-order relations) into the target domain, we should expect our analogy retrieval system to favour the selection of candidate sources with a greater number of causal relations than the presented target.
2) Number of *high-order* predicates – specifying associations between first-order predicates. Generally, source domains contain more causal relations than the impoverished target.
3) Number of *root* predicates – predicates that are not arguments to other predicates. Root predicates play a crucial role in incremental mapping models like IAM and SME. Including root-predicates in structure space will facilitate identification of suitable sources, particularly isomorphic sources. While a useful source will generally have more causal structure and therefore fewer root predicates, it was felt that including roots would be useful in many retrieval problems.
4) Number of *objects* – representing domain entities. In almost all examples of analogies found in the literature, the number of objects in the source and target are identical.
5) Number of *unique first-order predicates* - this differs from the number of predicates, as we do not count duplicate predicates. This identifies domains that rely on repeated use of a small number of predicates, from domains that use a few instances of many different predicates.

6) Number of *unique high-order predicates* - some domains rely on repeated usage of a single causal relation whereas others rely on a combination of different relations (eg cause + inhibit).
7) *Maximum agent object usage* - this counts the frequency that each object is used in either the agent or patient role of a relation. If an object appears in the agent role of two relations, then this value will be two.
8) *Maximum patient usage* – this counts the frequency of objecxts used in the paritent role of the domains realtions.

These last two structural attributes play a key role in distinguishing between structurally similar, though not identical, domain descriptions. Consider the domain descriptions in Figure 3, both involving three objects and three relations. On the left we have the "love triangle" domain and on the right we have the "un-requited love" domain. Although both are semantically similar in terms of their use of the same objects and relations, the structure of these domains is central to their semantics.

Structure based retrieval must be able to distinguish between these structures, and when presented with one structure should favor the isomorphic domain over the homomorphic domain. When presented with a domain analogous to the love-triangle, we wish to retrieve the love-triangle source. More importantly, we do *not* wish retrieve the "un-requited love" domain on the right of figure 3.

In the love-triangle domain, each object is used once as an agent and once as a patient. But in the un-requited love domain one object never occupies the patient role of the loves relation. These structural attributes help distinguish between the two domains, based purely on the structure of the representation.

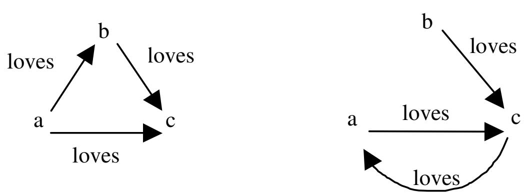

Figure 3: Structurally distinct domains

Of course, this is not an exhaustive list of all possible derived attributes, but these attributes do serve to distinguish between a variety of domain structures. Additional attributes might include identifying "loops" with the representation, the ratio of predicates to objects etc. Another distinction might be to quantify the use of commutative (`adjacent-to`) and non-commutative (`taller-than`) relations within a domain description. First order relations (taking objects as arguments) are central to a domain description. Such relations can also be classified using their temporal signature, whether they are `actions` or `states`. Quantifying the number of each type of first-order relation may also help improve precision. Experiments are ongoing to identify the most appropriate set of these structural attributes.

RADAR Retrieval Algorithm
1. For each candidate source
2. For each derived attribute type
3. Populate Structure space by computing the value for that structural attribute
4. Retrieve similar sources
5. For each derived attribute type
6. Compute the targets value for that structural attribute
7. Compute the distance to each candidate source
8. Sort candidate sources by distance, and select top K.

5 RESULTS

We now examine some results generated by RADAR, and compare them to the algorithms discussed earlier. Throughout this section we assume there is a known target domain, for which we seek a source domain that supports some creative interpretation. We examine the retrieval performance under a number of different circumstances to highlight its abilities. Firstly, we examine retrieval of isomorphic sources from a background memory of semantically similar domains.

Before beginning however, we again point out that while RADAR has much better recall than competing algorithms, correspondingly its precision is quite poor. We see this as an inevitable consequence of our search for creative source domains. Improving the precision of RADAR could make use of the semantic biasing of other retrieval algorithms, but would also require a better model of analogical validation.

5.1 Isomorphic Same-Domain Example

Consider the following target domain (from [13])*"Spot bit Jane causing Jane to flee from Spot"*. We now wish to identify *creative* candidate sources for this target problem.

```
cause(bite, flee)
bite (spot, jane)
flee-from(jane, stop)
```

The topographic structure is clear from a diagrammatic perspective.

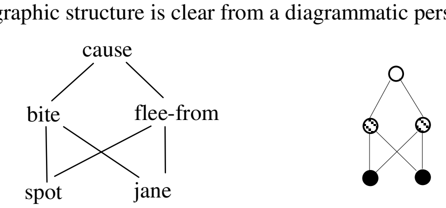

Figure 4: Domain Topology

We wish to identify domains that contain two objects (Spot and Jane), two distinct first order relations (bite and flee) and one causal relation (amongst the attributes listed in the previous section). RADAR will place each of the following sentences at the same location in structure space, even though their retrieval generated by the other algorithms will vary.

i) "Spot bit Mary causing Mary to flee from Spot"
ii) "Rex bit John causing John to flee from Rex".
iii) "Mort bit Felix causing Felix to flee from Mort".
iv) "John hit Rex causing Rex to flee from John".

Because each of these domains is structurally identical to the given target, the distance in structure space between these domains and the target is 0 in each case. Therefore, this indicates structural identity with the target and we are guaranteed to retrieve them.

5.2 Isomorphic Semantically-Distant Retrieval

Now consider the problem of retrieving the following domains. Note that while the semantics of each of these domains is very different to the source, they are all structurally identical to it. Because of the lack of semantic overlap the other algorithms will not retrieve these potential sources of creativity.

v) "The comet skimmed off the atmosphere causing the comet to leave the atmosphere"
vi) "The city encroached on the countryside causing the countryside to recede from the city"
vii) The tanks approached on the infantry causing the infantry to counter-attack the tanks".

Each of these domains can be considered a source of creative inspiration for the earlier target domain. Therefore, we would like a creative analogy retrieval algorithm to be capable identify these. RADAR identifies that the structural attributes of these domains are identical and thus are co-located in structure space.

5.3 Homomorphic Same-Domain Retrieval

Now, consider the following domains that are also semantically similar to the target, and that the other algorithms will retrieve but RADAR will not.

viii) "Spot and Jane fled from the bull".
ix) "Spot caused Jane to flee from Spot"
x) "Spot saw Jane causing Spot to flee from Jane"

Although each of these sentences contains information from the same semantic domain as the earlier source, they can not form an analogy with that source. This is caused by the lack of structural similarity with that target. Keane et al [10] refer to the IAM constraint to ensure that at least half the entities in the source domain participate in a mapping for the comparison to even be considered an analogy.

5.4 Homomorphic Retrieval

Consider the problem of generating (not just interpreting) the frequently referenced *"solar-system is like an atom"* analogy. This is a particularly challenging retrieval problem firstly because the source and targets are semantically quite distant. Secondly, the target is missing a causal relation and hence has a different (though somewhat similar) structure.

Useful source domains generally contain more causal structure than the driving target, and so more causal (high-order) relations than the target. Thus we expect that the performance of RADAR might be improved by reducing the influence of this axis in structure space retrieval. This is easily achieved by scaling that axis by a constant factor ($\beta < 1$) before performing retrieval, while setting $\beta = 0$ will eliminate the influence of this attribute on retrieval. However, more testing must be performed before this can be properly verified.

$$d = \sqrt{(ta1 - s_x a1)^2 + \beta(ta2 - s_x a2)^2 + ...} \qquad (2)$$

The target 'atom' domain would cause the retrieval of domains with a similar number of objects, first-order relations etc. This will include the universe domain, though in this case the distance in structure space will be greater than 0. The usual description of the universe contains the object 'planets' though this should be nine separate objects - one per planet. Structure based retrieval would thereby favour retrieving Fluorine with its 9 electrons. Because we use K-Nearest Neighbours retrieval, we can retrieve all domains with similar (though not identical) structural qualities. Thus any of the smaller elements would suffice. The following representation of the atom domain containing just one electron object, would favour retrieving the domain description containing the single object 'planets'.

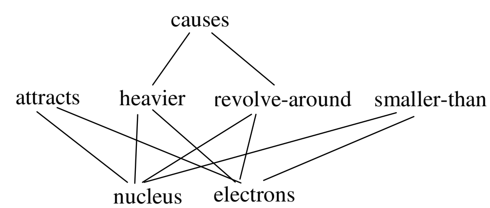

Figure 5: Domain Topology

Thus, RADAR is capable of identifying semantically distant but structurally *similar* source domains. This we feel is a crucial quality for generating creative analogies. It is this ability that separates RADAR from existing retrieval algorithms.

6 CONCLUSION

Retrieval models hold the key to generating a computational mode of truly creative analogical reasoning. Current models are hampered in their search for creative analogies by basing analogical retrieval in the semantics of the given problem domain. This approach as two main problems, firstly it considers only a small percentage of the possible candidate sources that are structurally similarity to the problem. Secondly, those domains that are considered are generally those that a human problem solver will already have considered. Thus, the potential of these models for identifying a creative comparison is severely constrained. Algorithms will remain blind to sources of creativity until structure, and not semantics, becomes the driving force behind analogical retrieval.

Structure based retrieval is a useful and flexible means of identifying candidate sources. Crucially, it overcomes the semantic constraint suffered by alternative retrieval models, greatly increasing the possibility of generating truly creative analogies. Domains are mapped into structure space where each axis identifies a topological feature of a domain description. A variety of structural attributes are derived from each domains representation, thereby creating a structure space that accurately reflects the structure of each domain's description. Analogical retrieval then takes place in structure space, by identifying domains located in the same region of structure space. Therefore, analogical retrieval is based on the structural similarity required by an analogy. This technique is oblivious to the semantic content these domains, resulting in a more creative model of analogical retrieval. Structure based retrieval opens the possibility for creating an analogy model with far greater creative potential than a human reasoning agent.

REFEREENCES

[1] Leadbetter D., Smith D., Price Nick, Huggan J. "David Leadbetter's Faults and Fixes: How to Correct the 80 Most Common Problems in Golf", 1996.

[2] Hoffman, R. "Monster Analogies", AI-Magazine, 3, 11-35, 1995.

[3] O'Donoghue, D. Computational Modelling of Creative Cognition Conference, DCU, Dublin, Ireland, 1997.

[4] Gick and Holyoak,

[5] Boden, M. A. "The Creative Mind", Abacus, 1992.

[6] Ritchie G. "Assessing Creativity", Proceedings of AISB Symposium on AI and Creativity in Arts and Science, York, March. 2001.

[7] Colton and Steel

[8] Boden, M. A. "Creativity and Artificial Intelligence", Artificial Intelligence, 103, 247-356, 1998.

[9] Gentner, D. "Structure-Mapping: A Theoretical Framework for Analogy", Cognitive Science, 7, 155-170, 1983.

[10] Keane, M. T. Ledgeway, T., Duff, S. "Constraints on Analogical Mapping: A comparison of Three Models", Cognitive Science, 18, 387-438, 1994.

[11] Forbus, K. Gentner, D. Law K. "MAC/FAC: A model of Similarity-based Retrieval", Cognitive Science, 19, 2, 1995.

[12] Thagard, P. Holyoak K. J. Nelson, G. Gochfeld, D. "Analogue Retrieval by Constraint Satisfaction", Artificial Intelligence, 46, 259-10, 1990.

[13] Plate T. "Structured operations with Distributed Vector Representations", in "Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational and Neural Sciences", New Bulgarian University, Sofia, Bulgaria, July 1998.