Towards Operational Abduction from a Cognitive Perspective

P.D. Bruza†, R.J. Cole‡, D. Song†, Z. Abdul Bari‡

†Distributed Systems Technology Center
The University of Queensland
St Lucia, Qld, 4072, Australia
{bruza,dsong}@dstc.edu.au

‡School of Information Technology and Electrical Engineering
University of Queensland
St Lucia, Qld, 4072, Australia
{rcole,zeen}@itee.uq.edu.au

Introduction

In the mid-nineteen eighties, Don Swanson, an information scientist, made a chance discovery by connecting two disparate on-line medical literatures, one dealing with Raynaud's disease, the other with dietary fish oil (Swanson, 1986a; Swanson, 1987). Patients with Raynaud's disease suffer from intermittent blood flow in the extremities - fingers, toes, and ears. At the time, there was neither a general treatment, nor a cure. Swanson formulated the explanatory hypothesis that fish oil may be a beneficial treatment, which was later verified by clinical trials. Swanson's discovery highlights a more widely occurring phenomenon. In order to deal with the information explosion, disciplines and expertise are becoming increasingly specialized and insular with little awareness of kindred, or potentially allied, specializations. As a consequence, disparate bodies of knowledge form, and with them “undiscovered public knowledge” (Swanson, 1986b). Automated, or semi-automated knowledge discovery systems can help counter this growing lack of awareness by suggesting potentially relevant connections between these islands of knowledge.

Swanson's discovery is an example of an abductive scientific discovery. Gabbay and Woods (2005a) have convincingly argued that abduction has its roots in cognitive economy. Put crudely, it is cheaper to “guess”, than to pursue a deductive agenda in relation to a problem at hand. It is interesting to briefly consider Gabbay and Woods' conjecture within the framework of Gärdenfors' three level model of cognition (Gärdenfors, 2000). How information is represented varies greatly across the different levels. The subconceptual level is the lowest level where information is carried by a connectionist representation. Within the uppermost level information is represented symbolically. It is the intermediate, conceptual level (or “conceptual space”), which is of particular relevance to this account. Here properties and concepts have a geometric representation in a dimensional space. For example, the property of “redness” is represented as a convex region in a tri-dimensional space determined by the dimensions hue, chromaticity and brightness. The point left dangling for the moment is that representation at the conceptual level is rich in associations, both explicit and implicit. We subscribe to the view that associations and analogies generated within conceptual space play an important role in hypothesis generation. Gärdenfors (2000, p48) alludes to this point when he states, “most of scientific theorizing takes place within the conceptual level”. His conjecture is aligned with Gabbay and Woods' insights regarding the cognitive economic basis of abduction: Within the conceptual space, inference takes on a decidedly associational character because associations are often based on context-sensitive similarity (e.g., semantic or analogical similarity), and notions of similarity are naturally expressed within a dimensional space. Inference at the symbolic level, however, is a linear, deductive process. It may well be that because associations are formed *below* the symbolic level of cognition, significant cognitive economy results. This is not only interesting from a cognitive point of view, but also opens the door to providing both a principled and computationally tractable abductive logical system.

In light of the introductory remarks above, it is our conviction that it would be misguided to adopt a traditional, symbolic perspective of an abductive logical system by assuming a propositional knowledge representation and proof-theoretic approaches for driving it. Gabbay and Woods (2003) argue that this perspective is conceptually incomplete - it ignores what is going on ``down below”. In terms of Gärdenfors’ model, ``down below” can be interpreted as the conceptual and subconceptual levels of cognition. Even if one does not accept Gabbay and Woods’ objection, another can be mounted from an operational stance. Textual information cannot automatically be rendered into a propositional representation. In addition, deductive approaches have well documented and daunting complexity results. Granted, the complexity challenges can be to a degree circumvented by the use of heuristics, but the dearth of large-scale symbolic logical systems reasoning over text suggests significant operational challenges not likely to be surmounted soon. For these reasons, we feel strongly that from both the conceptual and operational perspectives, a purely symbolic approach does not pave the way towards abductive logical systems on the scale needed to replicate Swanson-like discoveries. It is our conviction that in order to construct such systems, a cognitively motivated knowledge representation is required. More specifically, we advocate semantic spaces as a computational approximation of Gärdenfors' conceptual space. Hypotheses generated from semantic spaces do not have a proof-theoretic basis, but rather they are computations of associations by various means within the space. In ensuing sections we will show how conceptual space can be approximated computationally by means of semantic space. In addition, its feasibility for abduction will be demonstrated by replicating Swanson’s discovery by operational means.

In summary, the goal of this account is to introduce semantic spaces to the model-based reasoning and abduction community and to illustrate their potential for principled, operational abduction in “the large”. In order to construct an operational abductive system, two major questions should be addressed:

- How should the knowledge be represented?
- How to generate and justify hypotheses?

Both of these questions set the tone for following sections.

Approximating cognitive knowledge representation by semantic space

In order to illustrate how the gap between cognitive knowledge representation and actual computational representations can be bridged, the Hyperspace Analogue to Language (HAL) model is used (Lund and Burgess, 1996). HAL produces representations of words in a high dimensional space that seem to correlate with the equivalent human representations. Burgess, Livesay and Lund (1998) note “...simulations using HAL accounted for a variety of semantic and associative word priming effects that can be found in the literature...and shed light on the nature of the word relations found in human word-association norm data”.

HAL takes a corpus of text as input and learns a representation of words by accumulating weighted associations of co-occurring words in the context of fixed length window. More specifically, given a vocabulary of n words drawn from the corpus in question, HAL computes an n x n matrix by moving a window of length l over the corpus by one word increments, ignoring punctuation, sentence and paragraph boundaries. All words within the window are considered as co-occurring with strength 1. When the counts of the sliding window are aggregated, the strength of association between words becomes proportional to the distance between the words, because words that are closer together co-occur in more windows. Each row i in the matrix represents the accumulated weights of association of words that occur before i within context windows. Conversely, column i represents the accumulated weights of association of words that appear after *i* within context windows. By way of illustration, Table 1 depicts a HAL matrix constructed from the text “Beneficial effect of fish oil on blood viscosity” [1], with $n = 8$ and $l = 5$.

In the experiments reported later, the row and column in the HAL matrix corresponding to a given word *i* are added together to produce a single vector representation for that word. In the context of Table 1, the term “fish” would be represented by (ben: 3, eff: 4, of: 5, fish: 0, oil: 5, bld: 4, visc: 3). The row and column vectors are added together, thus combining pre and post co-occurrence counts and resulting in a symmetric matrix. The column vectors are then normalized to unit length. A matrix of such representations is termed a *HAL space*.

In practice, different variations of semantic space are possible. For example, stop words such as “the”, “on”, “of” etc. may be ignored. Also, HAL is but one scheme for computing term co-occurrence weights. Other weighting schemes include log-likelihood (Dunning, 1994) and odds-ratio (Lowe, 2001).

Table 2 shows part of the normalized HAL vector for the word “Raynaud” computed by applying the HAL method to a collection of 111,603 titles of core journal documents drawn from the MEDLINE collection (the dimensions are ordered by decreasing strength of association). This example demonstrates how a word is represented as a weighted vector whose components correspond to other words. The weights represent the strengths of association between “Raynaud” and other words it co-occurred with. The Raynaud vector is then an aggregated representation of the contexts in which the word “Raynaud” appears within the collection.

	ben	eff	of	fish	oil	on	bld	visc
ben								
eff	5							
of	4	5						
fish	3	4	5					
oil	2	3	4	5				
on	1	2	3	4	5			
bld		1	2	3	4	5		
visc			1	2	3	4	5	

Table 1: Example HAL matrix

[1] The example text is derived from the title of a MEDLINE electronic medical document: “*Beneficial effect of fish oil on blood viscosity in peripheral vascular disease*”.

Raynaud	
Dimension	Value
nifedipine	0.44
sleroderma	0.36
ketanserin	0.22
synthetase	0.22
sclerosis	0.22
prostaglandin	0.22
dazoxobin	0.22
E1	0.15
calcium	0.15
vasolidation	0.15
platelet	0.15
…..	….
blood	0.07
viscosity	0.07
vascular	0.07
...	...
fish	0.0

Table 2: Raynaud representation via HAL

The quality of HAL vectors is influenced by the window size: the longer the window, the higher the chance of representing spurious associations between terms. A window size of eight or ten has been used in various studies (Lund and Burgess, 1996; Burgess, Livesay and Lund, 1998; Bruza and Song, 2002).

More formally, a semantic space *S* used in this article is an *n* x *n* matrix, where *n* is the size of the term vocabulary. *S*[*i*,*j*] denotes the strength of co-occurrence of the terms *i* and *j*. The vector representation of a word *j* is the *j*'th column of *S,* and is denoted: s_j. The length of the vector, s_j, is given by:

$$|s_j| = \sqrt{\sum_{i=1}^{n} S[i,j]^2}$$

A vector s_j is normalized to unit length by dividing each of its components by the length of the vector:

$$normalize(s_j) = \frac{s_j}{|s_j|}.$$

HAL is an exemplar of a growing ensemble of computational models emerging from cognitive science, which are generically referred to as *semantic space* (Lund and Burgess, 1996; Landauer and Dumais, 1997; Patel, Bullinaria and Levy 1997; Burgess, Livesay and Lund, 1998; Landauer, Foltz and Laham, 1998; Levy and Bullinaria, 1999; Lowe, 2000; Lowe, 2001; Sahlgren, 2002). The term "semantic" derives from the intuition that words seen in the context of a given word contribute to its meaning. Colloquially expressed, the meaning of a word is indicated by the "company it keeps" (Kintsch, 2001). Even though there is ongoing debate about specific details of the respective models, they all feature a remarkable level of compatibility with a variety of human information processing tasks such as word association. Semantic spaces provide a geometric, rather than propositional, representation of knowledge. They can be considered to be approximations, albeit rather primitive, of the conceptual space proposed by Gärdenfors (2000).

Within a conceptual space, knowledge has a dimensional structure. As described earlier, colour can be represented in terms of three dimensions: hue, chromaticity, and brightness. Gärdenfors argues that a property is represented as a convex region in a dimensional space. In terms of the example, the property "red" is a convex region within the tri-dimensional space made up of hue, chromaticity and brightness. The property "blue" would occupy a different region of this space. A domain is a set of integral dimensions in the sense that a value in one dimension(s) determines or affects the value in another dimension(s). For example, the three dimensions defining the colour space are integral since the brightness of a colour will affect both its saturation (chromaticity) and hue. Gärdenfors extends the notion of properties to concepts, which are based on domains. The concept "apple" may have domains taste, shape, colour etc. Context is modelled as a weighting function on the domains, for example, when eating an apple, the taste domain will be prominent, but when playing with it, the shape domain will be heavily weighted (i.e., it's roundness). Observe the distinction between representations at the symbolic and conceptual levels. At the symbolic level "apple" can be represented as the atomic proposition *apple(x)*, whereas within conceptual space (the conceptual level of cognition), it has a representation involving multiple inter-related dimensions and domains. Colloquially speaking, the token "apple" (symbolic level) is the tip of an iceberg with a representation rich in association within the conceptual level. Gärdenfors points out that the symbolic and conceptual representations of information are not in conflict with each other, but are to be seen as "different perspectives on how information is described".

Barwise and Seligman (1997) also propose a geometric foundation to their account of inferential information content by use of real-valued state spaces. In their state space, the colour "red" would be represented as a point in a tri-dimensional real-valued space. For example, brightness can be modelled as a real-value between white (0) and black (1). Integral dimensions are modelled by so called observation functions defining how the value(s) in dimension(s) determine the value in another dimension. Note this is a similar proposal, albeit less expressive, to that of Gärdenfors as the representations correspond to points rather than regions in the space.

A HAL representation is an approximation of a Barwise and Seligman state space whereby the dimensions are words, and there are no observation functions. A word, or combination of words, like a noun compound are represented as a point, or vector, in the space. This point represents the ``state" in the context of the associated text collection from which the semantics space is learnt. If the collection changes, the state of the word may also change. In other words, HAL, and semantic spaces in general, allow for flexible "meanings" of words to be represented. These meanings can be studied longitudinally as they evolve over time (McArthur and Bruza, 2003). HAL, however, does not make provision for observation functions, so integral dimensions cannot be modeled. Despite HAL being a somewhat primitive approximation[2] of conceptual space, it nevertheless has an encouraging track record of cognitive validity (Burgess, Livesay, and Lund, 1998).

From an operational perspective, semantic spaces have been constructed from very large collections of text, for example, a corpus of Usenet news comprising 160 million words (Lund and Burgess, 1996), so they have a demonstrated track record of knowledge representation in the large.

In short, semantic spaces are a promising, pragmatic means for large-scale knowledge representation. Moreover, due to their cognitive credentials, semantic spaces would seem to be apt foundation for underpinning computational variants of human reasoning, like abduction.

Abduction from semantic space

Swanson's Raynaud/fish oil discovery was apparently a coincidence (Weeber et al., 2001). By reading two distinct and unconnected literatures for different reasons he was able to make the connection between Raynaud's phenomenon and dietary fish oil. The literatures were bridged via intermediate terms as depicted in figure 1 (adapted from Weeber et al. 2001). The basic architecture of the discovery is summarized by A-B-C, where C denotes the phenomenon, e.g., Raynaud, and A represents the potential cure, or treatment, e.g., fish oil. The discovery between C and A is made by means of intermediate B-terms, e.g., "platelet aggregation", "vascular reactivity" and "blood viscosity". It is important to keep in mind that the connection between C and A is indirect. Statistically it is a weak signal.

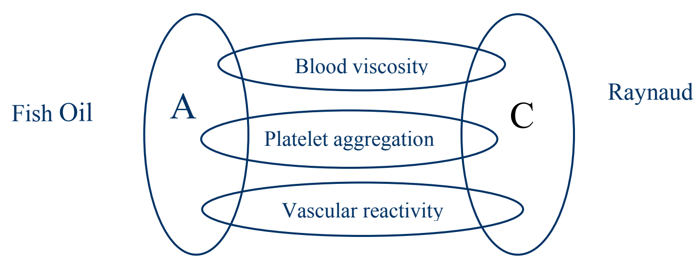

Figure 1: Swanson's Raynuad/fish oil discovery

Two modes of discovery underpin the A-B-C scheme: open and closed. The open mode of discovery involves the generation of a hypothesis, for example, that "fish oil is a treatment for Raynaud's phenomenon". The open discovery mode involves two steps. Firstly, there is the problem of identifying

[2] Gärdenfors (2000) argues that information representation undergoes a large dimensional reduction between the subconceptual and conceptual levels of cognition. HAL spaces can have a large dimensionality so it could be argued that its representations would correspond to the subconceptual level of cognition. This distinction is not relevant to the operational aspects of abduction to be presented in the next section.

salient B-terms in relation to the abduction trigger C (Raynaud). Secondly, once salient B-terms have been identified, these are then used to make connections to potential A-terms (e.g., fish oil). The closed mode of discovery involves the justification of the hypothesis. This account of operational abduction will focus on the open mode of discovery.

The open mode of discovery can be implemented in semantic space as follows. A large corpus of documents is identified which would contain multiple literature "islands" across a variety of specializations. A semantic space is constructed from the corpus, for example, by using HAL. The abduction trigger is represented in a semantic space by the C vector (Raynaud). The abduction problem is reduced to finding a strong association to the A vector (fish oil). The intuition underlying the open mode of discovery is if more of the B-terms are shared by the respective A and C vectors, then the connection between A and C is stronger. The dot product of the A and C vectors (in both higher and lower dimensionality) is a means of computationally realizing this intuition provided the B-terms are prominently weighted in the respective representations:

$$s_i \cdot s_j = \sum_{k=1}^{n} S[k,i].S[k,j]$$

Alternatively, the cosine between A and C can be calculated in order to bridge the connection between them (Gordon and Dumais, 1998). When A and C are both normalized to unit length, cosine equates to dot product.

$$\text{cosine}(i,j) = \frac{s_i \cdot s_j}{|s_i||s_j|}$$

Under the assumption that the B-terms are prominently weighted in the A and C representations,
another computational means to establish the A-C connection, is to consider A and C as points, and to measure the distance between them. The closer the points, the stronger the connection. The Minkowski family of metrics includes the Euclidean distance metric (*r=2*) and is given by:

$$\text{minkowski(i, j)} = \sqrt[r]{\sum_{k=1}^{n} \left| S[k,i] - S[k,j] \right|^r}$$

Euclidean distance (r=2) was employed by Burgess and Lund in their experiments evaluating HAL (Lund and Burgess, 1996; Burgess, Livesay and Lund, 1998).

Information flow has shown some promise for computing suggestions relevant to the Raynaud/fish oil discovery (Bruza, Song and McArthur, 2004). It is essentially a heuristic asymmetric form of dot product. Information flows from a source to a target vector and is motivated from information flow in real-valued state spaces, of which HAL is an exemplar (Barwise and Seligman, 1997, Song and Bruza, 2003). Information flow thresholds the co-occurrence weight of the source vector, the intuition being that prominently weighted dimensions in the source vector contribute the most to its essential "meaning" and are thus reliable for underpinning the flow computation. Information flow measures how well these prominent dimensions map into the target vector:

$$flow(i,j) = \frac{\sum_{x|S[x,i]>\partial \wedge S[x,j]>0} S[x,i]}{\sum_{x|S[x,i]>\partial} S[x,i]}$$

A common setting for the threshold ∂ is the mean weight of non-zero components in source vector s_i. A high degree of information flow is achieved when many of the dimensions above the threshold in the source vector s_i are also present in the vector s_j.

Abduction and dimensional reduction

Gärdenfors (2000, p220) contends that the information received by the senses and processed by the lower, subconceptual level of cognition is high dimensional, rich and unstructured. This statement is strongly related to research into consciousness summarized by Gabbay and Woods (2005a), and covered in more detail by Austin (1998). Information-theoretic research suggests that the senses process between 10^7 and 10^{11} bits per second. For any given second, only about 16 to 20 bits of information enter into consciousness (Austin, 1998 p278). It is clear that a lot is going on "down below". In the light of this, Gärdenfors reflects, "What is needed is some way of transforming the input into a mode that can be handled on the conceptual or symbolic level. This basically involves finding a more *economic* form of representation: going from the subconceptual to the conceptual level usually involves a *reduction of the number of dimensions that are represented*" (Italics in the original). Gärdenfors' reflection is clearly in tune with Gabbay and Woods' (2003) conjecture mentioned earlier - practical reasoning, like abduction, has its roots in cognitive economy.

The connection between dimensional reduction and cognition also appears to have an operational reflection in semantic space. For example, the reduction of semantic spaces into a lower dimensionality via singular value decomposition (SVD) has facilitated the replication of a number of cognitive effects, most notably in the semantic space model produced by Latent Semantic Analysis (LSA) (Landauer and Dumais, 1997; Landauer, Foltx and Laham, 1998). In this connection, Burgess, Livesay and Lund (1998, p219) also mention that some of the cognitive effects replicated using a 140,000 dimension HAL semantic space, could also have been achieved within a dimensionality of one or two hundred.

The relation between SVD and operationally abducing Swanson-like discoveries is the following. In text, some terms are not explicitly mentioned, but remain implicit to the discourse. A vector representation of a term t in a lower dimensional space includes weighted associations corresponding to the strength of "guesses" regarding how strongly terms implicitly co-occur with t. For example, the value in the "fish" component of the Raynaud vector in table 2 is zero as "fish" never explicitly co-occurs with the term "Raynaud" because the two literatures were disjoint. However, when the corresponding HAL space of 28,779 dimensions is reduced to a dimensionality of 50 by SVD, the corresponding Raynaud vector in the reduced space shows a weight of 0.00054 in the fish component. Admittedly this weight is small in an absolute sense, but nevertheless promotes the component "fish" to within the top 1143 out of 28,779 (4%) components within the Raynaud vector. This effect, can be viewed as an operational realization of "hasty", or pre-inductive generalization (Gabbay and Woods, 2003). Pre-indictive generalisation is mechanism for producing suggestions, and is therefore relevant to abduction.

SVD is a powerful technique from linear algebra (Golub and Van Loan, 1996). The SVD theorem states that any nxm matrix S with rank r can be decomposed into three matrices: $S = UDV^T$ where U and V are unitary nxr and mxr matrices respectively. D is an rxr diagonal matrix whose values are monotonically increasing singular values of S. The columns of U and V are the eigenvectors of SS^T and S^TS respectively.

Dimensional reduction is performed by taking only the first k eigenvectors ($k < m$) and singular values to approximate S by $S_k=U_kD_kV_k^T$, where U_k and V_k are nxk and mxk matrices composed of the first k columns of U and V respectively. The Eckart-Young theorem states that S_k is the closest rank-k approximation to S in the sense of the matrix 2-norm. Stated formally,

$$S_k = \min_{rank(B)=k} \|S - B\|_2$$

An intuition which can be ascribed to the Eckhart-Young theorem is that SVD tries to capture as much of the variation in the data in S within the given number of dimensions (k). Bear in mind that data often exhibits regularities, for example, clustering. One of the assumptions behind semantic space models is that words with similar "meanings" will tend to cluster. In this connection, consider figure 2. The left hand cluster contains 30 points in 3D and the right hand cluster 20 points. Dimensional reduction from 3D to 2D involves projecting the points down onto the 2D plane defined by the 2 eigenvectors, which define the axes of the plane. Observe how the first eigenvector drives an axis between the two clusters, with a bias towards the left cluster as it has more points. The second eigenvector drives an axis along the two clusters. It is as if the first eigenvector functions like a weighted average of the two clusters thus giving a global representation encompassing the two clusters. The second tries to capture the remaining variation by positioning itself optimally so the separation between points in 3D is also manifested when projected down on this second axis (eigenvector). The Eckhart-Young theorem essentially says that the low dimensional subspace is conditioned so as to preserve as much as possible, on average, the length of vectors that are projected into it.

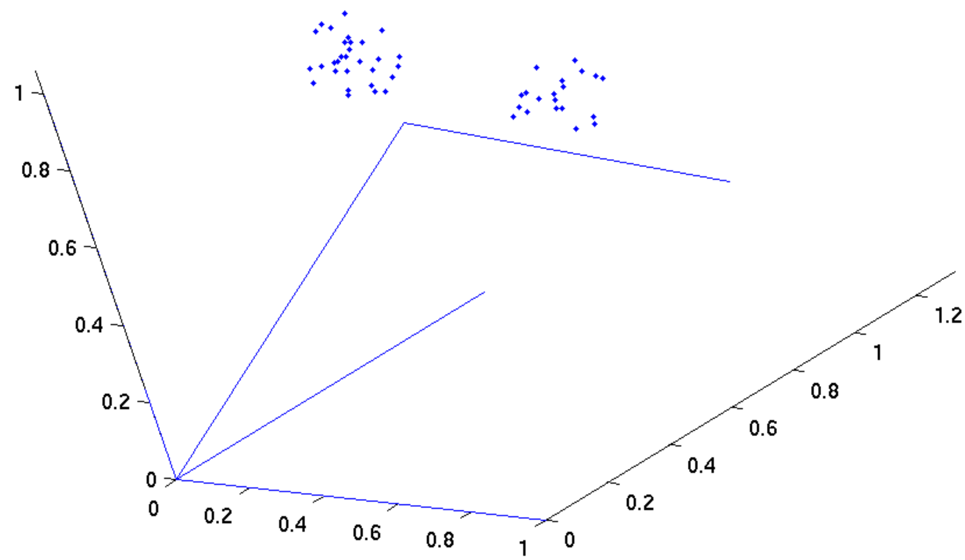

Figure 2: Eigenvectors in 2D generated from 3D data points

Generalizing the above example to n dimensions leads to the intuition that the lower dimensional approximation to S yields to "guesses" of (implicit) term co-occurrence. If a group of terms commonly co-occur together then they are likely to be combined together in an eigenvector. This has the effect that if one of these terms occurs in a document then the other terms will also be included. This effect causes S_k to have non-zero entries where S had zero values, the positive weight in S_k reflecting the strength of the pre-inductive generalization.

SVD is used as the basis of Latent Semantic Analysis which has a promising track record of cognitive validity across a number of information processing tasks. Landauer, Foltz and Laham (1998) allude to SVD's abductive capability as follows, "This text segment [Y] is best described as having so mush of Abstract Concept 1 and so much of Abstract Concept 2, and this word [X] has so much of Concept 1 and so much of Concept 2, and combining those two pieces of information, *my best guess* is that word *X* actually appeared 0.6 times in context *Y*" (Emphasis ours). An "abstract concept" is related to the spectral decomposition of SVD:

$$S_k = \sum_{l=1}^{k} u_l \sigma_l v_l^T \qquad\qquad S_k[i,j] = \sum_{l=1}^{k} U[i,l]D[l,l]V[j,l]$$

where the column vectors u of U and v of V are interpreted as describing abstract concepts, and σ is a singular value along the diagonal in D. The notion of "abstract concept" is a reference to eigenvector and has an intriguing ring to it. Ding (2005) refers to the reduced space as "intrinsic semantic space", presuming the projection is optimal[3]. Are, then, the axes (eigenvectors) of the reduced space, the fundamental axes of "meaning" embedded in matrix S? The literature tends toward the agnostic in relation to this fascinating question, notwithstanding LSA's track record. We feel, however, that there is quite some mileage left in this question when it is placed in an abductive perspective. Principal component analysis

[3] "Optimal" can best be conceived of in relation to a specific task. For example, using semantic space to replicate human word association norms.

(PCA) is kindred technique to SVD. Gärdenfors (2000, p242) observes, “Thus the first principal component (read eigenvector) is the spatial direction in the data set that has the highest variation and is thus the maximally “explanatory” dimension”. Gärdenfors admits that equating variation to explanatory power is debatable, but nevertheless the question beckons as to whether eigenvectors in reduced space support the inference of Swanson-like explanatory hypotheses. This question will be taken up again in the context of the experimental results presented in the next section.

SVD can be used for operational abduction in the same way in the reduced space as is done in the non-reduced space. The association mechanisms presented above can be used in the reduced space to compute potential A-C connections.

In summary, this section has detailed a number of mechanisms for computing associations from semantic spaces. The next section reports on how they perform in the context of operational abduction by assessing their ability to replicate the Swanson Raynaud/fish oil discovery.

Abduction at work

The experiments reported in this section summarize results and insights gleaned from our attempts to operationally abduce the Swanson Raynaud/fish oil discovery from semantic spaces (Bruza, Song and McArthur, 2004; Cole et al., 2005). In the construction of a semantic space from a text corpus, there are a variety of operational decisions to be made. For example: which stop word list to use, whether or not to stem terms, which term weighting best fosters discovery, which mechanism to use to discover potential A-C connections, and whether or not to employ dimensional reduction. Many of these decisions have to do with reducing, or conditioning the space to emphasize those dimensions likely to bear on the discovery. Fundamentally, this has much to do with the problem of relevance, which will be addressed shortly. The goal of the experiments is to evaluate the effect of these decisions on the open discovery problem.

A corpus of 111,603 MEDLINE core clinical journal articles from the period 1980-1985. (This period is before Swanson’s discovery). Only the titles of the articles were used as Swanson's original discovery was made solely on the basis of document titles.

Two stop words lists were employed: A standard collection of stop words commonly used in information retrieval experiments and a list of tailoured stop words for literature based discovery from Swanson’s Arrowsmith system (Swanson and Smalheiser, 1997). After applying the normal stop word list there were 34,716 distinct words in the collection, while after applying the Arrowsmith stop words the number of distinct words fell to 28,779. Three weighting schemes were used: HAL, log-likelihood, odds-ratio. Three discovery mechanisms were employed: Cosine, Euclidean distance, Information flow. The semantic space was optionally reduced in dimension using singular value decomposition. Stemming was found to have little or no positive impact on the results and so all results reported are from experiments performed without stemming.

Discovery of B-terms

The guiding intuition behind the discovery of B-terms follows Gordan and Lindsay’s conjecture that the best B-terms are those that are semantically and statistically close to the C-term (Gordon and Lindsay, 1996). This view was supported by Gordon and Dumais, where a semantic neighbourhood[4] was computed around the Raynaud vector in a lower dimensional space computed via SVD (Gordon and Dumais, 1998). We have found in our experiments that priming the semantic space with log-likelihood weighting and then ranking all the dimensions within the C vector on decreasing weight has encouraging success in promoting salient B-terms (“platelet”, “vascular”, “blood”, “viscosity”), irrespective of whether the stop words are tailoured, or not (Cole et al., 2005) . The semantic space was not reduced. This result is welcome as it

[4] The cosine of the C vector and all other vectors are computed and ranked on decreasing order of cosine. The vectors of terms semantically related to C are typically highly ranked.

removes the computational cost of computing both the semantic neigbourhood and performing dimensional reduction. In addition, log-likelihood weighting may reduce the dependency on tailoured stop words, which are time consuming to manually compile and maintain.

Discovery of the A-C connection

Once a set of B−terms has been discovered, they can be used to prime an A − C discovery in the following way. The C vector (corresponding to “Raynaud”) is normalized to unit length. Thereafter, those dimensions corresponding to the B−terms are given a maximal weight of one. This corresponds to the situation where the scientist is giving positive relevance feedback with respect to those terms (s)he assesses as salient to addressing the phenomenon represented by C. In the results reported below, the known B−terms “blood”, “viscosity”, “vascular”, and “platelet” were boosted manually. Cosine and Euclidean distance are measured to all terms in the vocabulary, and ranked. Terms that co-occur in a title with Raynaud are discarded from the ranking as only indirect connections are of interest. The ranks of the terms “fish” and “oil” are then inspected. Both the actual rank, and the percentage distance from the top of ranking are reported. Comparisons with the information flow metric are made on the basis of previously published results.

Semantic Space	fish	oil
Cosine, Normal, HAL	368 (1.06%)	982 (2.83%)
Cosine, Normal, LL	4493 (12.94%)	374 (1.08%)
Cosine, Normal, Odds	197 (0.57%)	445 (1.28%)
Cosine, ARR., HAL	515 (1.79%)	915 (3.18%)
Cosine, ARR., LL	700 (2.43%)	270 (0.94%)
Cosine, ARR., Odds	346 (1.20%)	384 (1.33%)
Eucl., Normal, HAL	89 (0.26%)	35 (0.10%)
Eucl., Normal, LL	13254 (38.18%)	363 (1.05%)
Eucl., Normal, Odds	**12 (0.03%)**	**5 (0.01%)**
Eucl., ARR, HAL	23 (0.08%)	14 (0.05%)
Eucl., ARR., LL	9016 (31.33%)	100 (0.35%)
Eucl., ARR., Odds	**5 (0.02%)**	**4 (0.01%)**
Flow, ARR, HAL	**17 (.06%)**	**27 (0.09%)**

Table 3: A-C similarities in higher space after boosting B−term weights in the Raynaud vector.

Discovery based on Euclidean distance in a semantic space primed with odds ratio terms weights performed outstandingly well by promoting both “fish” and “oil” to a very high degree. It does not depend on tailoured stop words, which is a significant benefit. It should be noted that all of the successful replications of the Raynaud/fish oil discovery of which the authors are aware, have relied on external props, like specialized stop words, external knowledge sources, or significant manual intervention. Information flow also showed encouraging performance, but its performance depends on tailoured stop words. Overall, the results suggest that once relevant B-terms have been indentified, it is certainly possible to bridge the A-C connection by operational means.

The effect of dimensional reduction

Despite SVD's inherent abductive character, its performance in replicating the Swanson discovery is disappointing (Gordon and Dumais, 1998). It has had more notable success in discovering B-terms. Table 4 depicts the Raynaud vector of Table 2 in a reduced space of 50 dimensions computed by SVD. Note how a couple of the salient B-terms have been significantly promoted, in particular "platelet" and "blood" which occur at positions 1 and 6 respectively. These terms however have been promoted at the expense of "viscosity" which is a relatively rare term within the corpus. SVD, because it tries to capture as much variation in the data as quickly as possible, tends to neglect infrequent terms in lower dimensions. Dimensional reduction has also introduced spurious terms such as "cancer" and "chemotherapy" into the Raynaud vector due to the co-occurence of these terms with the terms that Raynaud co-occurs with. These can be characterised as pre-inductive generalizations that went wrong. Note also that dimensional reduction has produced a small non-zero entry for fish. While small, this value is within the top 4% of the values in the Raynaud vector.

Table 4: Raynaud vector in the low dimensional space (k=50)

platelet	0.17
pulmonary	0.07
renal	0.07
cancer	0.07
lung	0.07
blood	0.07
ventricular	0.07
chemotherapy	0.07
….	
fish	0.00075

We found that when the Arrowsmith stop words were used dimensional reduction did improve the strength of the A—C connection. However the improvement was much less that achieved by boosting salient B-terms. When the normal stop words were used, dimensional reduction reduced the strength of the A—C connection. This is because the reduced space becomes dominated by terms such as "patient" and "trial" that are not relevant to the A—C connection. All in all, this suggests that SVD is sensitive to the frequency bias of more commonly occurring terms. While dimensional reduction using SVD clearly has abductive capability, more research is required to properly harness it's capability. SVD's intriguing potential is succinctly expressed by Landauer, Foltz and Laham (1998), "The relationships inferred by LSA are also not logically defined, and they are not assumed to be consciously rationalizable as these could be. Instead, they are relations of similarity – or of context sensitive similarity – but they nevertheless have mutual entailments of the same general nature and also give rise to fuzzy indirect inferences that may be weak or strong and logically right or wrong".

The Architecture of an Abductive Knowledge Discovery system

There have been quite a number of attempts at replicating the Swanson's Raynaud/fish oil discovery and a subsequent discovery linking migraines with magnesium (Gordon and Lindsay, 1996; Swanson and Smalheiser, 1997; Gordon and Dumais, 1998; Weeber et al., 2001; Bruza, Song and McArthur, 2004; Srinivasan, 2004, Cole et al., 2005). Most have not positioned their work from an abductive perspective. In addition, there is a great variation in the technical details of the operational systems involved. In this section, the architecture of abduction is motivated from a cognitive perspective. This architecture not only serves as a high level blue print for operational abduction, but also provides a framework for general discussion regarding previous work.

Gabbay and Woods (2005b) contend that the (human) abductive reasoner solves "The Cut Down" problem. This can be imagined as funnel taking a space of possibilities and refining them through successive filters until the hypothesis, (or hypotheses), to be discharged, comes dripping out as a premiss (or premises). More specifically, *Gen* is a sublogic of hypothesis generation, resulting in a space of hypotheses ***S***. The members of ***S*** are each a "possible hypothesis for possible conjecture". Next, the engagement sublogic, *Engage,* engages those elements of ***S*** *relevant* to the abduction problem at hand. The result of *Engage* is a proper subset of ***S***, namely ***R,*** the set of relevant hypotheses for possible conjecture. In turn, the plausibility filter contracts ***R*** to a set of possibilities for actual conjecture, represented by ***P***. Finally, the discharge sublogic *Dis* transforms the plausible hypotheses into a premise (or premises) by subjecting it (or them) to a test filter of which Woods and Gabbay identify two varieties: "The filter of independent confirmation" and "The filter of theoretical fruitfulness". The distinction between hypothesis and premise is important. Swanson may have actually conjectured several hypotheses in relation to Raynaud's, but "fish oil" was the only premiss he discharged by writing an article. In summary, the triple (***S***, ***R***, ***P***) represents a filtration structure on the initial space of possibilities, in which succeeding spaces are cut downs of their predecessors.

It is apparent from the literature, and from the experiments documented in the previous section, that it is not difficult to come up with operational variants of *Gen* producing ***S***. For example, *cosine*, Euclidean distance, and *information flow* are all examples of mechanisms for operationally realizing the sublogic *Gen*. The real challenge is gaining sufficient operational command of the relevance filter. Relevance is a subtle, multidimensional and dynamic concept. By way of illustration, the "Nifedepine" dimension in the Raynaud HAL vector of table 2 expresses a relevant, strong association, as Nifedepine was used as a treatment for Raynaud's phenomenon in clinical trials, without success. Bear in mind, however, this association is *not* relevant for finding a treatment for Raynaud's. Therefore, this association should not figure highly, if at all, in operational mechanisms trying to make the discovery. Such subtlety is very hard to determine operationally. For this reason, many of the systems cited above resort to specially tailored stopword lists, and/or external sources of knowledge like the MeSH (Medical Subject headings) or Unified Medical Language Systems (UMLS) in order to grapple with relevance. Substantial manual intervention is often required, for example, the compilation of tailoured stopword lists. There is a significant cost in the deployment of such measures, so the question arises whether the reliance on them can be lessened, or even removed. Preliminary indications from statistical term-weighting in semantic spaces point to an encouraging possibility for gaining operational command of relevance without relying on external props or large amounts of manual intervention. More substantive experiments are needed to bear this out. It is likely, however, that statistical term weighting will only turn out to be a partial solution to operational relevance.

In operational systems, *Gen* and *Engage* are confounded. The generation sublogic *Gen*, propped up with *ad hoc* mechanisms to cope with the relevance problem produces rankings of terms or phrases (the hypotheses) with the implicit assumption that the ranking function is also a ranking in the *plausibility* of the hypotheses. The scientist would peruse such rankings and identify those suggestions for actual conjecture. Therefore, the plausibility filter ***P*** is ultimately performed by the scientist using the system. However, the system can provide support for this, for example, by providing explanatory facilities of why the system determined a suggestion A to be highly ranked in relation to the abduction trigger C. Such operational support has been termed the "closed mode of discovery". Similarly, the discharge sublogic *Dis* is the province of the human scientist. Again the system can provide a supporting role, for example, by retrieving a focused sub-collection of documents to aid the determination of "theoretical fruitfulness".

It would be a mistake to conceive of operational abduction as a linear filtration process. We envisage it to be an interactive process incorporating feedback to condition vector representations. By way of illustration, the scientist can first browse suggestions for B-terms from the system in relation to the abduction trigger C, use his or her background knowledge to enhance these suggestions, or change their weights. Consider "blood viscosity". Curiously, this has a weak statistical association with Raynauds' in the underlying corpus. The scientist could boost the respective weights reflecting background knowledge of its importance. In an operational setting this amounts to the scientist interactively applying a context function to the weights in the C vector. The conditioned C vector can then be used to compute A-C suggestions.

Summary and future work

In the introduction we asserted the need for abductive knowledge discovery systems to enhance our awareness in light of the information explosion. For example, by providing creative suggestions spanning disparate islands of knowledge. We have argued that a principled and pragmatic approach to operational abduction can be taken from a cognitive perspective. Two questions were raised: How should the knowledge be represented and how to generate and justify the hypotheses? In relation to the first question, semantic spaces have been presented as an approximation, albeit primitive, of the conceptual level of cognition, which harbours geometric (dimensional) representations of information. In addition, semantic spaces offer a computationally attractive knowledge representation, which circumvents the operational challenges of symbolic knowledge representation. In relation to the second question, it has been speculated, for reasons of cognitive economy, that abduction is rooted in conceptual level of cognition, where inference has a natural associational, analogical character. Various mechanisms were detailed which compute hypotheses by association within semantic space. Operational abduction was illustrated by implementing Swanson's A-B-C discovery scheme and showing promising results in semi-automatically replicating his Raynaud/ fish oil discovery from a semantic space derived from a corpus of biomedical literature.

Future work will be directed at going beyond the A-B-C discovery scheme. Dunbar (1999) concludes from cognitive studies that scientists frequently resort to analogies when there is not a straightforward answer to their current problem. Analogical reasoning plays an important role in hypothesis formation. The question is how to generate analogies from semantic space. In this connection, the work of Eliasmith and Thagard (2001) provides an interesting avenue for further exploration. They present a computational model for replicating colloquial analogies by vector convolutions in a dimensional space. The holographic reduced representations (HRRs) used for this purpose are somewhat akin to those found in semantic spaces.

A criticism that can be leveled at the "flat", dimensional representations of a semantic space is the lack of structural relationships. This is a fair criticism of semantic space, but not of the conceptual level of cognition, which does cater for these. Our goal is to first ascertain how far semantic spaces can be exploited without recourse to structural relationships.

A major challenge facing abductive systems is gaining operational command of relevance: It is easy to generate hypotheses; far harder is it to generate relevant hypotheses (while at the same time excluding those that aren't). From a cognitive perspective it seems that relevance and dimensional reduction are inextricably bound as information passes from the subconceptual to conceptual level of cognition. Experiments, thus far, with dimensional reduction of semantic space have not been promising in producing the operational equivalent, however, more research in this area is certainly warranted.

Finally, it is important to stress that the view of abduction presented in this account does not rest on traditional conception of logic. Gabbay and Woods (2003, p63) speculate that a logic of "down below" could be "a logic of semantic processing without rules". We feel that abduction from semantic spaces falls very much within the ambit of such speculation and actually reinforces it. In turn, this view of a ``logic'' is also aligned with C.S. Peirce's view of abduction: ``No reason whatsoever can be given for it [abduction], as far as I can discover; and it needs no reason, since it merely offers suggestions'' (Peirce Edition Project, 1998, p217).

Acknowledgments

The work reported in this paper has been funded in part by the Co-operative Research Centre for Enterprise Distributed Systems Technology (DSTC) through the Australian Federal Government's CRC Programme (Department of Industry, Science & Resources). Thanks to Marc Weeber for the Arrowsmith stop word list.

References

Austin, J., 1998, *Zen and the Brain: Towards an Understanding of Meditation and Consciousness*. MIT Press.

Barwise, J. and Seligman, J., 1997, *Information Flow: The Logic of Distributed Systems*. Volume 44 of Cambridge Tracts in Theoretical Computer Science, Cambridge University Press.

Bruza, P.D., and Song, D., 2002, Inferring query models by computing information flow. in: *Proceedings of the 11th International Conference of Information and Knowledge Management (CIKM 2002)*, ACM Press, pp. 260–269.

Bruza, P.D., Song, D., and McArthur R., 2004, Abduction in semantic space: Towards a logic of discovery, *Logic Journal of IGPL*, 12:97–110.

Burgess, C., Livesay, K., and Lund K., 1998, Explorations in context space: words, sentences, discourse. *Discourse Processes*, 25(2&3):211–257.

Cole, Bruza, Song, Abdul Bari, 2005, The mechanics of literature-based discovery. Submitted to the *28th International Association for Computer Machinery Special Interest Group on Information Retrieval (ACM SIGIR)*.

Ding, C., 2005, A probabilistic model for dimensionality reduction in information retrieval and filtering. *Journal of the American Society for Information Science and Technology*, 56.

Dunbar, K., 1999, How scientists build models invivo science as a window on the scientific mind. in *Model-Based Reasoning in Scientific Discovery*, L. Magnani, ed., Kluwer Academic/Plenum Publishers, New York, pp. 85–99.

Dunning, T., 1994, Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

Eliasmith, C., and Thagard, P., 2001, Integrating structure and meaning: a distributed model of analogical mapping, *Cognitive Science*, 25:245–286.

Gabbay, D., and Woods, J., 2003, Agenda Relevance: A Study in formal Pragmatics, Volume 1 of *A Practical Logic of cognitive Systems*. Elsevier.

Gabbay, D., and Woods, J., 2005a, The Reach of Abduction: Insight and Trial, Volume 2 of *A Practical Logic of cognitive Systems*. Elsevier.

Gabbay, D., and Woods, J., 2005b, Filtration structures and the cut down problem in abduction. In: *Mistakes of Reason: Essays in Honour of John Woods*, K. Peacock and A. Irvine, eds. University of Toronto Press.

Gärdenfors, P., 2000, *Conceptual Spaces: the Geometry of Thought*. MIT Press, London.

Golub, G. and Van Loan, C., 1996, Matrix Computations. John Hopkins University Press.

Gordon, M. D., and Dumais, S., 1998. Using latent semantic indexing for literature based discovery. *Journal of the American Society for Information Science*, 48:674–685.

Gordon, M. D. and Lindsay, R. L., 1996, Towards discovery support systems: A replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between raynaud's and fish oil. *Journal of the American Society for Information Science*, 47:116–128.

Kintsch. W., 2001, Predication. *Cognitive Science*, 25:173–202.

Landauer, T.K., and Dumais, S.T., 1997, A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge, *Psychological Review*, 104: 211–240.

Landauer, T.K., Foltz, P.W., and Laham, D., 1998, An introduction to latent semantic analysis, Discourse Processes, 25(2&3):259-284.

Levy, J.P., and Bullinaria, J.A., 1999 Learning lexical properties from word usage patterns: Which context words should be used? in: *Connectionist Models of Learning, development and Evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop*, R.F. French and J.P. Sounge, eds., Springer, London, pp. 273–282.

Lowe, W., 2000, What is the dimensionality of human semantic space? in: *Proceedings of the 6th Neural Computation and Psychology workshop*, Springer Verlag, pp. 303–311.

Lowe, W., 2001, Towards a theory of semantic space, in: *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, J. D. Moore and K. Stenning, eds., Lawrence Erlbaum Associates, pp. 576–581.

Lund, K., and Burgess, C., 1996, Producing high-dimensional semantic spaces from lexical co-occurrence, Behavior Research Methods, Instruments & Computers, 28(2):203–208.

McArthur, R., and Bruza, P.D., 2003, Discovery of tacit knowledge and topical ebbs and flows within the utterances of online community, in: *Chance Discovery*, Y. Ohsawa and P. McBurney eds., Springer Verlag, pp. 115–132.

Patel, M., Bullinaria, J.A., and Levy, J.P., 1997, Extracting semantic representations from large text corpora, in: *Connectionist Models of Learning, development and Evolution: Proceedings of the Fourth Neural Computation and Psychology Workshop*, R.F. French and J.P. Sounge, eds., Springer, London, pp. 199–212.

Peirce Edition Project, ed., 1998, The nature of meaning, in: Essential Peirce: Selected Philosophical Writings. Vol 2 (1893-1913), Indiana Univ. Press, pp. 208–225.

Sahlgren, M., 2002, Towards a flexible model of word meaning, Paper presented at the *AAAI Spring Symposium 2002*, Stanford University, Palo Alto, California, USA.

Song, D., and Bruza, P.D. 2003, Towards context sensitive information inference. *Journal of the American Society for Information Science and Technolog*y, 54(4):321–334.

Srinivasan, P., 2004, Text mining: Generating hypotheses from Medline. *Journal of the American Society for Information Science and Technolog*y, 55(5):396–413.

Swanson, D.R., 1986, Fish oil, Raynaud's syndrome, and undiscovered public knowledge, *Perspectives in Biology and Medicine*, 30(1):7-18.

Swanson, D. R., 1986, Undiscovered public knowledge. *Library Quarterly*, 56:103–118.

Swanson, D. R., 1987, Two medical literatures that are logically but not bibliographically connected. *Journal of the American Society for Information Science*, 38:228–233.

Swanson, D.R. and Smalheiser, N.R., 1997, An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence*, 91(2):183–203.

Swanson, D. R. and Smalheiser, N. R., 1999, Implicit text linkages between Medline records: Using Arrowsmith as an aid to scientific discovery. *Library Trends*, 48:48–59.

Weeber, M., Klein, H., Jong van den Berg, L., and Vos, R., 2001, Using concepts in literature-based discovery: Simulating swanson's raynaud/fish-oil and migraine/magnesium discoveries, *Journal of the American Society for Information Science and Technology*, 52(7):548–557.