Discovering Information Flow Using a High Dimensional Conceptual Space

Dawei Song and Peter Bruza

Distributed Systems Technology Centre
Level 7, General Purpose South
The University of Queensland, QLD 4072 Australia

{dsong, bruza}@dstc.edu.au

ABSTRACT

This paper presents an informational inference mechanism realized via the use of a high dimensional conceptual space. More specifically, we claim to have operationalized important aspects of Gärdenfors's recent three-level cognitive model. The connectionist level is primed with the Hyperspace Analogue to Language (HAL) algorithm which produces vector representations for use at the conceptual level. We show how inference at the symbolic level can be implemented by employing Barwise and Seligman's theory of information flow. This article also features heuristics for enhancing HAL-based representations via the use of quality properties, determining concept inclusion and computing concept composition. The worth of these heuristics in underpinning informational inference are demonstrated via a series of experiments. These experiments, though small in scale, show that informational inference proposed in this article has a very different character to the semantic associations produced by the Minkowski distance metric and concept similarity computed via the cosine coefficient. In short, informational inference generally uncovers concepts that are carried, or, in some cases, implied by another concept, (or combination of concepts).

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval]:** Information Search and Retrieval– *Retrieval models*; H1.1 [**Models and Principles**]: Systems and Information Theory]– Information theory; I.2.0 [**Artificial Intelligence**]: General– Philosophical foundations; I.2.6 [**Artificial Intelligence**]: Learning– *Concept learning.*

General Terms

Theory, Algorithms, Experimentation.

Keywords

Conceptual space, Information flow, Informational inference.

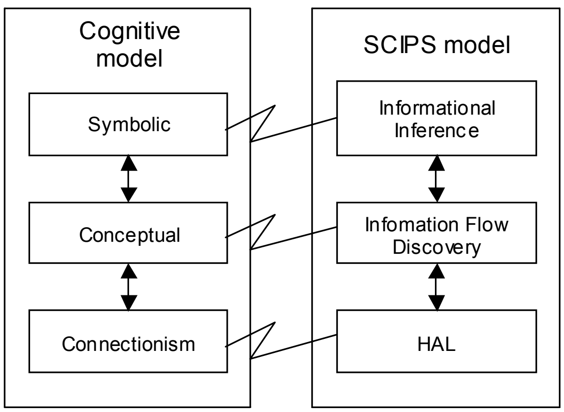
1. INTRODUCTION

Consider the text fragment "*Welcome to Penguin Books, U.K*". A human can quickly make the judgment that this text probably refers to Penguin, the publisher. The text "*Antarctic Penguins*", on the other hand, would lead to the judgment that the text is referring to with penguins of the animal variety. Human have the ability to make hasty, though reliable judgments about what terse text fragments are about (or are not about). Many of us do this daily while scanning the subject descriptions of emails, or the title captions in the result set from a search engine. In situations involving large amounts of incoming electronic information (e.g., defence intelligence), judgments about content (whether by automatic or manual means) are sometimes performed based simply on a title description or brief caption because it is too time consuming, or too computationally expensive to peruse whole documents.

This article is about how to automatically infer that one piece of information *carries* information about another. This goes beyond the traditional term co-occurrence relationships and we refer it to as *Informational Inference*. For example, given a company named NEC, we may conclude that NEC also carries the information "a computer company", "an electronics corporation", etc. This is also referred to as "*information flow*" by Barwise and Seligman (1997): *x carries/bears/conveys the information that y*. The discovery of information flow will be able to enhance our cognitive power and thus become more aware in our ever more complex information environment. We draw upon research from both cognitive science and logic.

Currently, symbolic and connectionist approaches dominate in cognitive science. The former views cognition as symbolic manipulation, while the latter models associations using artificial neural networks. However, neither of them provides appropriate modelling tools for the mechanisms of *concept learning*, which are fundamental for many cognitive phenomena, for example, the aforesaid information flow between two concepts. Gärdenfors (2000) proposes a three-level cognitive model, which embodies the symbolic, conceptual and connectionist perspectives. He introduces conceptual spaces as a bridge between the symbolic and connectionist approaches. A conceptual space is built upon geometric structures representing concepts and properties.

Informational inference has been proposed in terms of rules prescribing properties of aboutness (Bruza, Song and Wong, 2000). It therefore suffers from the disadvantages inherent to the symbolic approach, for example, it sustains no creative inductions, no genuinely new knowledge and no conceptual discoveries (Gärdenfors 2000). Gärdenfors' conceptual level allows informational inference to be defined in terms of vector representations at the conceptual level. In order to operationalize Gärdenfors' cognitive model, we have proposed *Semiotic Cognitive Information Processing Systems (SCIPS)* - next generation information retrieval devices (Bruza and Song 2001). In this article, we will use vector representations, which are obtained from the connectionist level via the Hyperspace Analogue to Language (HAL) approach (Lund and Burgess, 1996; Burgess, Livesay and Lund, 1998). The conceptual level will then feature a theory of information flow, which underpins the informational inference at the symbolic level (Barwise and Seligman, 1997). In this way, the architecture of a SCIPS is shown as below:



2. CONCEPTUAL SPACE

Within the conceptual level of Gärdenfors' cognitive model, information is represented geometrically. For example, the property colour can be represented as a ternary vector of three dimensions: Hue, chromaticity, and brightness. Hue is manifested directly from the wavelength of the light, so a hue of 445 nano-metres corresponds to the colour red. Chromaticity is a dimension that reflects the saturation of the colour. The three dimensions that together represent the property of colour have their roots in the human perceptual mechanism of vision, however, this need not always be the case; dimensions may also be abstract.

The concept "apple" may have domains taste, shape, colour etc. The thrust of Gärdenfors' proposal is that properties (and concepts) are represented geometrically as points (or regions) in a space of dimensions (or domains). Context is modelled as a weighting function on the domains, which expresses the dimensions' salience within a given context. 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).

Gärdenfors extends the notion of properties into concepts which are based on the concept of a *domain*- a domain being a set of integral dimensions in the sense that a value in one dimension(s) determines or affects the value in another dimension(s). By way of illustration, the dimensions used to establish colour are integral because a value cannot be assigned on one dimension without giving values to the other dimensions. For example, the brightness of a colour will affect its saturation (chromaticity).

A human encountering a new concept draws its meaning via an accumulation of experience of the contexts in which the concept appears. In parallel, for text machine learning, the meaning of a concept can be learnt through training lexical co-occurrence information in a corpus to obtain the history of contexts it experiences. Following this idea, Burgess and Lund developed a representational model of semantic memory, namely Hyperspace Analogue to Language (HAL) to automatically construct a high dimensional semantic space from a collection of text. Numeric vectors of concepts are produced to represent meanings of these concepts (Lund and Burgess 1996; Burgess, Livesay and Lund 1998).

A window is moved over the whole corpus by one word increment and all the words within the window are considered as co-occurring with each other with strengths inversely proportional to the distance between them. After traversing the corpus, an accumulated co-occurrence matrix for all the words in a target vocabulary is produced. Note that the word pair in HAL is direction sensitive, i.e. the co-occurrence information for words preceding every word and co-occurrence information for words following it are recorded separately by its row and column vectors. Given n-word vocabulary, the length of each vector is 2n.

We applied HAL method to the Reuters-21578 collection. The vocabulary is constructed by removing a list of stop words and also dropping some infrequent words which appears less than 25 times in the collection. The size of final vocabulary is 5403 words. Window size is set to be 6. A too small window leads to loss of potentially relevant correlations between words, whereas a too large window may compute irrelevant correlations. Burgess et al (1998) employed a window size of 8 in their experiments. We think 6-word window size is reasonable since precision is our major concern. Furthermore, for the purpose of this paper, we don't consider the direction sensitivity of word pair, and added the row and column vectors into one, thus the dimension of each vector is reduced to vocabulary size n. As an example, part of the HAL vector for *nec* is as follows:

nec = < analysts: 28, bull: 29, chip: 27, computer: 33, corp: 227, electronics: 40, information: 26, japan: 37, nec: 120, pct: 26, petition: 32, quantum: 27, series: 29, …… >

This example demonstrates how a word is represented as a weighted vector whose dimensions comprise words. The un-normalized weights represent how strongly words co-occur with "NEC" in the context of the sliding window, summed across the whole collection. Note that those highly weighted dimensions, e.g., "corp", would be expected by the average user to be useful term associations in relation to the term "NEC". We propose to use the HAL vectors as a means to prime the geometric representations inherent to Gärdenfors' conceptual spaces. HAL vectors are also interesting because semantic associations computed using these vectors correlate with semantic associations drawn from human subjects. Therefore, there is evidence that the HAL vectors approximate to cognitive representations of words. Another advantage of the HAL approach is that it is automatic. In the following we formally define a computational model of conceptual space based on HAL vectors.

Concept

A concept c_i is a vector representation:

$c_i = < w_{c_i p_1}, w_{c_i p_2}, ... w_{c_i p_n} >$

where $p_1, p_2, ..., p_n$ are called dimensions of c_i, n is number of dimensions of c_i, and $w_{c_i p_i}$ is the weight of p_i in vector of c_i. A dimension is termed a property if its weight is greater than zero. Using the illustration above (c_i = *NEC*), for the property *computer*, $w_{NEC.computer}$ = 33. Function $P(c_i)$ is used to represent the set of properties of concept c_i.

Quality Property

Some dimensions are more important than others depending on the context. In this account, we will not deal with context directly, but will approximate it by extracting those properties of sufficient weight as computed across the Reuters collection. In our experiments, various thresholds will be set for identifying quality properties.

A property p_i of a concept c_i is a *quality property* iff $w_{c_i p_i} > \partial$, where ∂ is a non-zero threshold value. Function $QP_i(c_i)$ is used to represent the set of quality properties of concept c_i with respective to ∂ set for c_i.

Conceptual Space

A conceptual space **S** is set of all concepts in a collection and the properties of a concept are all the concepts in the space. Formally, let $c_1, c_2, ... c_n$ be the concepts in a space **S**, i.e. $c_1, c_2, ... c_n \in \mathrm{S}$. For each concept c_i, $c_i = < w_{c_i c_1}, w_{c_i c_2}, ..., w_{c_i c_i}, ... w_{c_i c_n} >$.

Combining Concepts

Gärdenfors states that "our ability to combine concepts and, in particular, to *understand* new combinations of concepts is a remarkable feature of human thinking" (Gärdenfors 2000, p114). For example, most people can understand combinations such as *pink elephant* or *cubic soap bubble*. Gärdenfors presents a thoughtful account of how the combination of concepts is realized in terms of the geometric representations of the conceptual level. For example, "elephant" is a concept with many dimensions, one of which is colour. This colour is typically grey, which is a property represented as a region[1]. The concept "pink elephant" can be constructed by replacing this grey region with another region representing the property pink. Observe in this example that "pink" acts as a modifier for the concept "elephant", the latter being the more dominant of the two. In general, the combination of concepts cannot always be realized in such a straightforward fashion. As a consequence, Gärdenfors does not present a comprehensive theory from which an implementation of concept combination can be derived. Therefore, in the following we formalize a heuristic-based approach to combining concepts based on HAL vectors.

Given two concepts $c_1 = < w_{c_1 p_1}, w_{c_1 p_2}, ... w_{c_1 p_n} >$ and $c_2 = < w_{c_2 p_1}, w_{c_2 p_2}, ... w_{c_2 p_n} >$. Assume c_1 is dominant. The resulting combined concept is denoted $c_1 \oplus c_2$. For example, let c_1 be "NEC" and c_2 be "computer", then $c_1 \oplus c_2$ would denote the geometric representation underpinning the noun phrase compound "NEC computer".

An important intuition is to weight the dimensions in the dominant concept higher than in the other concept, and strengthen the weights of the dimensions in common. In the following, we restrict our attention to "meaningful" composition – the intersection between the sets of quality properties of c_1 and c_2 is not empty.

Note we did not normalize the property weights in the previous definitions. For the purpose of computation, in particular, the composition of two concepts, however, normalization is desirable. For example, the weights of a HAL vector depend on the overall history of co-occurrence of dimensions within the moving window. However, the significance of a dimension in a vector is relative. Thus by normalizing the vectors, they can be compared or computed at the same level. We propose to use the following cosine normalization algorithm for a dimension p_j in concept c_i:

$$w_{c_i p_j} = \frac{w_{c_i p_j}}{\sqrt{\sum_k w_{c_i p_k}^2}} \qquad (1)$$

Concept Combination Heuristic

Step 1: Re-weight c_1 and c_2 in order to assign higher weights to the quality properties in c_1.

$$w_{c_1 p_i} = \ell_1 + \frac{\ell_1 * w_{c_1 p_i}}{\underset{k}{Max}(w_{c_1 p_k})} \qquad (2)$$

$$w_{c_2 p_i} = \ell_2 + \frac{\ell_2 * w_{c_2 p_i}}{\underset{k}{Max}(w_{c_2 p_k})}$$

$\ell_1, \ \ell_2 \in (0.0, \ 1.0)$ and $\ell_1 > \ell_2$

For example, if $\ell_1 = 0.5$ and ℓ_2=0.4, then property weights of c_1 are transferred to interval [0.5, 1.0] and property weights of c_2 are transferred to interval [0.4, 0.8], thus scaling the dimensions of the dominant concept higher.

[1] Gärdenfors proposes that natural properties occupy a convex region of a domain (a set of integral dimensions).

Step 2: Strengthen the weights of properties appearing in both c_1 and c_2 - these will form important dimensions in the resulting combination.

$$\forall (p_i \in P(c_1) \wedge p_i \in P(c_2)) \mid (w_{c_1 p_i} = \alpha * w_{c_1 p_i} \text{ and } w_{c_2 p_i} = \alpha * w_{c_2 p_i}) \quad \text{where } \alpha > 1.0. \quad (3)$$

Step 3: Compute property weights in the composition $c_1 \oplus c_2$:

$$w_{(c_1 \oplus c_2) p_i} = w_{c_1 p_i} + w_{c_2 p_i} \quad (4)$$

Step 4: Normalize the vector $c_1 \oplus c_2$. The resultant vector can then be considered as a new concept, which, in turn, can be composed to other concepts by applying the same heuristic.

To illustrate the above heursitic, the following vectors of *nec* and *corp* contain only quality dimensions. (In this case, those dimensions above two standard deviations of the mean):

NEC = < bull: 0.098, computer: 0.111, corp: 0.766, electronics: 0.135, japan: 0.125, nec: 0.405, petition: 0.108, series: 0.098 >

corp = < acquisition: 0.069, agreed: 0.042, air: 0.045, america: 0.039, **american: 0.123**, **bank: 0.100**, banking: 0.049, board: 0.051, boston: 0.066, business: 0.055, capital: 0.044, chairman: 0.053, chemical: 0.043, chrysler: 0.080, communications: 0.039, company: 0.072, **computer: 0.047**, corp: 0.688, credit: 0.046, debt: 0.054, development: 0.059, dividend: 0.070, dlrs: 0.097, electric: 0.042, energy: 0.042, expects: 0.045, federal: 0.055, filed: 0.043, financial: 0.146, general: 0.163, group: 0.040, insurance: 0.062, international: 0.163, loss: 0.064, machines: 0.048, mln: 0.128, motors: 0.138, national: 0.086, net: 0.154, offer: 0.040, offering: 0.049, pacific: 0.056, pct: 0.085, petroleum: 0.060, poor: 0.146, qtly: 0.053, qtr: 0.255, quarterly: 0.041, sales: 0.042, securities: 0.053, sets: 0.081, shares: 0.074, shr: 0.084, standard: 0.104, stock: 0.067, subsidiary: 0.117, systems: 0.052, union: 0.048 >

The combination of *NEC* and *corp* is (ℓ_1 , ℓ_2 and α are set to be 0.5, 0.3 and 2.0 respectively):

NEC ⊕ corp = < acquisition: 0.014, agreed: 0.009, air: 0.009, america: 0.008, **american: 0.025**, **bank: 0.021**, banking: 0.010, board: 0.011, boston: 0.013, bull: 0.297, business: 0.011, capital: 0.009, chairman: 0.011, chemical: 0.009, chrysler: 0.016, communications: 0.008, company: 0.015, **computer: 0.308**, corp: 0.555, credit: 0.009, debt: 0.011, development: 0.012, dividend: 0.014, dlrs: 0.020, electric: 0.009, **electronics: 0.305**, energy: 0.009, expects: 0.009, federal: 0.011, filed: 0.009, financial: 0.030, general: 0.034, group: 0.008, insurance: 0.013, international: 0.033, **japan: 0.303**, loss: 0.013, machines: 0.010, mln: 0.026, motors: 0.028, national: 0.018, nec: 0.359, net: 0.032, offer: 0.008, offering: 0.010, pacific: 0.011, pct: 0.017, petition: 0.299, petroleum: 0.012, poor: 0.030, qtly: 0.010, qtr: 0.052, quarterly: 0.008, sales: 0.009, securities: 0.011, series: 0.297, sets: 0.017, shares: 0.015, shr: 0.017, standard: 0.021, stock: 0.014, subsidiary: 0.024, systems: 0.011, union: 0.010 >

The above example illustrates that the composition heuristic assigns high weights to intersecting quality properties (e.g., "computer"), and assigns the other properties appearing in the dominant concept relatively higher weights than those in the non-dominant concept. *Corp* is a concept with properties typically relevant to corporate issues, such as "bank", "finance", "sales", "shares", etc. When it is combined with *NEC*, intersecting properties such as *computer* and *corp* are strengthened[2]. Also, the other quality properties of *NEC* are merged into the *corp* vector. The weights of other quality properties are weakened. Observe how in the *corp* vector that "american" has a high weight reflecting that in the Reuters collection there is a strong co-occurrence relationship between "american" and "corp". After the composition, "american" has a relatively low weight, and in contrast "japan" has a high weight. This illustrates desirable nonmontonic behaviour with respect to concepts (Gärdenfors 2000, p126). As a consequence, the resultant *NEC ⊕ corp* vector reflects a Japanese computer and electronics corporation as would be expected.

3. INFORMATION FLOW

We view the HAL-based vectors of concepts to be cognitively motivated representations of "meaning". The token such as "NEC" is not simply a sequence of three characters, but is underpinned by a much richer vector representation which embodies associations with other concepts. In classical logic, the connection between the level of tokens (syntax) and the level of meaning (semantics) is established by model theory. A parallel can be found between the symbolic and conceptual levels in the sense that tokens at the symbolic level are related to each other via their representations at the conceptual level. Barwise & Seligman (1997) have proposed an account of information flow that provides a theoretical basis for establishing such a connection by the use of information state spaces. The conceptual spaces constructed from HAL vectors are a particular example of the state spaces that Barwise and Seligman propose. The connection between the symbolic level and state space is formalized as follows:

Definition 1 (Barwise-Seligman's Information Flow)

$$i_1, \ldots, i_n \vdash j \text{ iff } \bigcap_{0 \le k \le n} s(i_k) \subseteq s(j)$$

The left hand side of the formula describes an relationship between a set of types (tokens) i_1 , i_2 , … i_n and a type (token) *j*. The intuition behind the above formula is that it establishes the information described by the combination of tokens i_1 to i_n carries the information described *j*. For example,

NEC, computer |- technology

Barwise & Seligman refer to it as a "constraint" between the respective sets of types; the relationship can be conceptualised as one of information flow between the conjunction of i_1 to i_n to *j*.

[2] Note that the above **NEC ⊕ corp** vector has been normalized. Thus the absolute weights of its properties might be less than those in *NEC* or *corp* vectors, even though their relative significances in **NEC ⊕ corp** may have increased.

We consider Barwise & Seligman's definition of information flow to particular formalization of informational inference at a symbolic level.

The right hand side of Barwise & Seligman's definition describe how the inference relationship is defined in terms of state spaces, where $s(i_k)$ denotes the state space associated with token $i_k, 0 \le k \le n$. We view that a HAL vector represents the information "state" of a particular concept (or combination of concepts) with respect to a given collection. For example, the token "NEC" is underpinned by a particular vector representation as shown in the previous section. As a consequence, the intersection and inclusion of states needs to be defined appropriately in order to compute the right hand side of the above formula. For intersection, we propose the concept combination heuristic detailed in the previous section. Furthermore, as HAL vectors are not perfect representations of the associated concept, inclusion should not be defined in strict sense as in set theory. We propose to compute a degree of inclusion based on how many dimensions of one concept are present in another concept. The informational inference at the symbolic level is deemed to hold if the degree of inclusion is deemed to be sufficient:

Definition 2 (HAL-based information flow)

$$i_{1,} \ldots, i_n \mid\!- j \text{ iff degree}(\oplus c_i \subseteq c_j) > \partial$$

where c_i denotes the conceptual representation of token i, and δ is a threshold value. (For ease of exposition, $\oplus c_i$ will be referred to as c_i (combinations of concepts are also concepts).

Inclusion is a relation over concepts (i.e., ⊆ ∈ S×S), which models that one concept is included in another one.

Given two concepts c_i and c_j the degree of inclusion is defined as follows:

$$\text{degree}(c_i \subseteq c_j) = \frac{\sum_{p_l \in (QP_i(c_i) \wedge QP_j(c_j))} w_{c_i p_l}}{\sum_{p_k \in QP_i(c_i)} w_{c_i p_k}} \tag{5}$$

The underlying idea of this definition is to make sure that a majority of the most important quality properties of c_i appear in c_j. The numerator calculates the accumulation of weights of those quality properties appearing in both c_i and c_j. The denominator is the sum of all quality properties weights of c_i. According to Barwise and Seligman's definition, the other quality properties of c_j need not to be considered[3]. When a threshold value 1.0 is set for degree($c_i \subseteq c_j$), the HAL-based information flow definition equates to Barwise & Seligman's one.

Our definition of information flow shows some similarity to the use of fuzzy inclusion for computing broader terms (Miyamoto 1990). However, this work does not deal with concept combinations, and moreover, we feel that informational inference goes beyond the notion of broader term.

4. EXAMPLES

4.1 Single-Concept Information Flow

For the case of single concepts, we select a fairly typical concept "NEC". "NEC" is a company that appears in a number of business contexts. As a consequence, it's HAL vector has 241 properties. The results detailed in this section are similar to the results we achieved investigating other concepts drawn from the Reuters collection.

The basis of the experiment is to see if the inferences resulting from information flow have a different character to similarity metrics. To this end, we compare the information flow results with those computed using the cosine and Minkowski measures. The latter is claimed by Burgess et al (1998) to be a semantic distance measure between words.

$$\text{similarity-cosine}(c_i, c_j) = \frac{\sum_k w_{c_i p_k} * w_{c_j p_k}}{\sqrt{\sum_k w^2_{c_i p_k} * \sum_k w^2_{c_j p_k}}} \tag{6}$$

$$\text{minkowski}(c_i, c_j) = \sqrt[l]{\sum_k | w_{c_i p_k} - w_{c_j p_k} |^l} \tag{7}$$

$$\text{similarity-minkowski} = e^{-k \, * \, minkowski(c_i,\, c_j)} \tag{8}$$

In our experiments, l is set to be 2 (Burgess et al (1998) also used $l = 2$ in their experiments), and k is set to be 1/1500.

Table 1 depicts the results. Each column of a table lists top 10 concepts[4] resulting from the cosine-based similarity function, Minkowski distance based similarity function and the HAL-based information flow functions:

Similarity (Cosine)	Similarity (Minkowski-based)	Information Flow
nec : 1.0 intel : 0.6878 hospital : 0.6067 itt : 0.6055 republicbank:0.5922 unisys : 0.5872 gte : 0.5861 exxon : 0.5758 corp : 0.5717 usx : 0.5591	nec : 1.0 intel : 0.7768 unisys : 0.7532 republicbank:0.7504 itt : 0.7429 southland : 0.7417 ball : 0.7407 bellsouth:0.7403 nynex : 0.7392 gte : 0.7347	nec (55): 1.0 computer(271):0.9415 { corp (596), electronics(135), information(225)}: 0.8355 analysts (474): 0.8154 computers(129): 0.8007 { charging (66), chip (93), controls (166), high (413), japan (510), largest (311), maker (121), technology (295), supply (305) }: 0.7623

Table 1: Analysis of concept "NEC"

[3] We designed another algorithm considering both c_i and c_j in the denominator. It behaves similarly to the cosine function.

[4] The information flows grouped in braces all have the same associated degree

- We tried various threshold variables to determine quality properties used in the underlying representations of concepts. However, as the Reuters collection is small, there is insufficient basis for forming a theory in this regard. The selection of quality properties is still a research question.
- The number in brackets next to each concept is the number of quality properties of that concept.

Discussion:

- The cosine and Minkowski functions yield results of a different character than information flow. As the cosine and Minkowski functions measure similarity, they tend to compute similar concepts to NEC. For example, "Intel", "Unisys", "ITT" are all technology companies like NEC. On the other hand, information flow tends to uncover information carried by the source concept NEC, for example, "computer", "electronics", "industry", "information", "corp.", "japan" etc. Moreover, they are far beyond the "broader terms" of "nec".
- The resultant concepts flowing from a concept are not necessarily among the properties of that concept. This means information flow has a truly inferential character rather than simply being a product of term co-occurrence.
- There is a wide spread in the number of properties of the concepts being inferred (from 66 to 596). Results across a number of examples suggest that the information flow of Definition 2 is not biased towards inferring concepts with larger numbers of properties.

4.2 Multi-concept Information Flow

In this subsection, we give illustrations by showing some typical results from our concept combination model. The quality properties of the combined concept are selected by setting a threshold of 2 standard deviation above the average with the minimal number of quality properties being 8. The threshold of right hand side of information flow relation is set to be above the average. The parameters ℓ_1, ℓ_2 and α are set to be 0.5, 0.3 and 2.0 respectively. The experiments are classified into *vertical* and *horizontal* tests. A vertical test refers to the refinement of a concept by composing a number of other related concepts in order to specify its context. For example, combination of "arms" and "talks" into "arms talks" makes the concept "talks" more specific. A horizontal test is to specify different contexts of a general (i.e. multi-contextual) concept by composing different more specific concepts to it. Different information flows could be produced with respect to different contexts. For example, information flowing out of "arms talks" is different from "gatt[5] talks" because "arms" and "gatt" specify different contexts of "talks".

4.2.1 Vertical Test

Using HAL-based information flow, the top 20 inferences from the concept "talks" are:

talks |- {**talks** (387): 1.000, **negotiations** (302): 0.865, **agreement** (498): 0.832, meeting (343): 0.784, international (502): 0.758, plans (482): 0.741, government (710): 0.739, set (423): 0.734, bank (611): 0.733, major (595): 0.713, company (776): 0.704, trade (550): 0.685, officials (477): 0.682, banks (510): 0.680, companies (543): 0.676, official (445): 0.672, national (421): 0.665, countries (406): 0.664, conference (247): 0.661, continue (467): 0.660 }

Observe that the information flowing from "talks" includes general concepts such as "negotiations", "agreement", etc.

Similarly, for the concept "arms":

arms |- { **arms** (92): 1.000, **iran** (212): 0.864, officials (475): 0.766, **reagan** (326): 0.758, **american** (436): 0.723, set (436): 0.723, **soviet** (244): 0.718, major (603): 0.690, result (353): 0.690, agreed (442): 0.679, oil (542): 0.665, strongly (155): 0.654, export (410): 0.653, **contra** (23): 0.649, **sale** (366): 0.646, agency (313): 0.644, states (348): 0.644, china (359): 0.642, equipment (261): 0.640, details (223): 0.638 }

In the Reuters collection, "arms" has two major contexts. The first is the "arms scandal" that the Reagan administration suffered when secretly selling arms to Nicaraguan rebels (the so-called Iran-contra affair). The other context is arms control – a series of talks were held with the Soviet Union about medium distance nuclear missiles. Both of these contexts are reflected in the above vector.

Using the concept composition heuristic presented earlier, "arms⊕ talks" yields (assuming "arms" to be dominant):

arms ⊕ talks = < agreement: 0.254692, arms: 0.259444, control: 0.121992, deal: 0.105783, iran: 0.189955, negotiator: 0.104788, profits: 0.103651, reagan: 0.257337, sale: 0.109196, sales: 0.127537, scandal: 0.132513, secret: 0.106352, soviet: 0.269334, talks: 0.331767, trade: 0.104741, >

Now, the top 20 informational inferences from arms ⊕ talks are:

arms ⊕ talks |- { **arms** (92): 1.0, officials (475): 0.816, **iran** (212): 0.802, **reagan** (326): 0.773, **american** (436): 0.771, set (436): 0.771, **soviet** (244): 0.770, oil (542): 0.764, **agreement** (505): 0.738, future (356): 0.736, **talks** (387): 0.728, agreed (442): 0.726, details (223): 0.721, made (558): 0.720, **union** (350): 0.719, **moscow** (120): 0.718, major (603): 0.703, result (353): 0.703, **government** (724): 0.693, make (509): 0.685 }

The above results highlight relevant inferences such as "American", "Soviet", "Union", "Moscow", etc. These convey information about the negotiations surrounding the control of nuclear missiles. Analysis revealed that the reason why desirable inferences "nuclear", "missile" were not appearing was due to the window producing the HAL-vectors was too narrow.

The above results also demonstrate some unsound inferences such as "Iran", etc. The problem here is that the Iran-contra context dominates the arms control context. This suggests that the combination algorithm is not sufficient to smooth out all the variations in context. Depending on the context, certain properties will be highly weighted and others less so. These weights will shifts as the context shifts. Gärdenfors (2000) proposes that context can be modelled as a weighting function over the

[5] Gatt (General Agreement on Tariffs & Trade) a forum for global trade talks.

properties, however, the practical research challenge is the automatic acquisition and maintenance of this weighting function.

4.2.2 Horizontal Test

This experiment compares the inferences drawn from arms$\oplus$talks versus those from gatt$\oplus$talks.

Assuming “gatt” to be dominant:

gatt $\oplus$ talks = < agreement: 0.282, agricultural: 0.106, body: 0.117, china: 0.121, council: 0.109, farm: 0.261, gatt: 0.279, member: 0.108, negotiations: 0.108, round: 0.312, rules: 0.134, talks: 0.360, tariffs: 0.114, trade: 0.432, world: 0.114,>

The top 20 informational inferences from gatt $\oplus$ talks are:

gatt $\oplus$ talks |- { **gatt** (119): 1.000, **trade** (540): 0.963, **agreement** (505): 0.961, **world** (426): 0.856, **negotiations** (307): 0.850, **talks** (387): 0.843, set (436): 0.822, states (348): 0.819, **ec** (371): 0.814, japan (499): 0.782, general (371): 0.778, **farm** (273): 0.776, include (354): 0.767, **rules** (225): 0.763, round (107): 0.763, members (338): 0.736, council (177): 0.734, **agriculture** (211): 0.731, officials (475): 0.724, government (724): 0.718 }

Gatt$\oplus$talks mainly carries information of negotiations and agreements about the rules of international agricultural (farm) trade between different countries, especially European countries. When contrasted with the inferences drawn from arms$\oplus$ talks, some measure of context sensitivity is revealed, meaning that the inference mechanism is sensitive to “gatt” or “arms” in the context of “talks”.

The above horizontal and vertical tests are only some of typical examples selected for illustration. They do suggest that HAL-based model of information flow can realize desirable behaviour in the form of non-monotonicity and context sensitivity.

5. SUMMARY AND CONCLUSION

The main contribution of this article is the realization of an informational inference mechanism via the use of a high dimensional conceptual space. The dimensional space offers two advantages. Firstly, it offers a cognitively motivated model theory on which to found inference at a symbolic level. Secondly, as this model theory is expressed in terms of vectors, the computational complexity of inference at the symbolic level can be side-stepped. More specifically, we claim to have operationalized important aspects of Gärdenfors’s three-level cognitive model. The connectionist level is primed with the HAL approach which produces vector representations for use at the conceptual level. We show how inference at the symbolic level can implemented by employing Barwise and Seligman’s theory of information flow: The real valued state spaces advocated by them are realized by HAL vectors to represent the information “state” of a word in the context of a collection of words. Cognitive studies have revealed that HAL vectors correlate with the cognitive representations of words, therefore, by employing them, we aim to promote informational inferences that correlate with human judgements about information. This article also features heuristics:

1. for enhancing HAL-based representations via the use of quality properties.
2. determining concept inclusion
3. computing concept composition.

The worth of these heuristics in underpinning informational inference are suggested via a series of small experiments. These experiments, though small in scale, show that informational inference proposed in this article has a very different character to the semantic associations produced by the Minkowski distance metric and concept similarity computed via the cosine coefficient. In short, informational inference generally uncovers concepts that are carried, or, in some cases, implied by another concept, (or combination of concepts). Though early days, our results suggest that the informational inference mechanism defined in this article could possibly be used to deduce nested information relationships (Van Rijsbergen, 1989) Such relationships, combined with semantic, and other automatically computed associations open the door to the automatic construction of ontologies. Our ultimate goal is to produce information processing devices which have some sense of the "meaning" of the information they are processing. Moreover, the inferences they draw in relation to information will correlate with inferences that human agents would draw modulo the context. These devices, termed semiotic-cognitive information processing systems, will enhance our cognitive firepower and thus help us become more aware in our ever more complex information environment.

6. ACKNOWLEDGEMENTS

The work reported in this paper has been funded in part by the Cooperative Research Centres Program through the Department of the Prime Minister and Cabinet of Australia.

7. REFERENCES

Barwise, J. and Seligman, J. (1997) Information Flow: The Logic of Distributed Systems. Cambridge Tracts in Theoretical Computer Science, 44.

Bruza, P.D., Song, D. and Wong, K.F. (2000) Aboutness from a Commonsense Perspective. Journal of the American Society for Information Science, 51(12), 1067-1156.

Bruza, P.D. and Song, D. (2001) Informational Inference via Information Flow. To appear.in the Proceedings of the second International workshop on logical and uncertainty models for information systems (LUMIS’01).

Burgess, C., Livesay, K. and Lund K. (1998) Explorations in Context Space: Words, Sentences, Discourse. Discourse Processing, 25(2&3), 211-257.

Gärdenfors, P. (2000) Conceptual Spaces: The Geometry of Thought. MIT Press.

Lund, K. and Burgess C. (1996) Producing High-dimensional Semantic Spaces from Lexical Co-occurrence. Behavior Research Methods, Instruments, & Computers, 28(2), 203-208.

Miyamoto, S. (1990) Fuzzy Sets in Information Retrieval and Cluster analysis. Kluwer Academic Publishers.

Rijsbergen, C.J. van (1989) Towards and information logic. In Proceedings of the 12th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp 77-86.