Embedding Probabilities in Predication Space with Hermitian Holographic Reduced Representations

Trevor Cohen[1] and Dominic Widdows[2]

[1] University of Texas School of Biomedical Informatics at Houston
[2] Microsoft Bing

Appeared in 9th international conference on Quantum Interaction (QI-2015), Filzbach, Switzerland, July 14-17, 2015

Abstract. Predication-based Semantic Indexing (PSI) is an approach to generating high-dimensional vector representations of concept-relation-concept triplets. In this paper, we develop a variant of PSI that accommodates estimation of the probability of encountering a particular predication (such as $\mathrm{fluoxetine}$ TREATS $\mathrm{major\_depressive\_disorder}$) in a collection of predications concerning a concept of interest (such as major depressive disorder). PSI leverages reversible vector transformations provided by representational approaches known as Vector Symbolic Architectures (VSA). To embed probabilities we develop a novel VSA variant, Hermitian Holographic Reduced Representations, with improvements in predictive modeling experiments. The probabilistic interpretation this facilitates reveals previously unrecognized connections between PSI and quantum theory - perhaps most notably that PSI's estimation of relatedness across multiple reasoning pathways corresponds to the estimation of the probability of traversing indistinguishable pathways in accordance with the rules of quantum probability.

Keywords: Distributional Semantics, Vector Symbolic Architectures, Holographic Reduced Representations, Quantum Interactions

1 Introduction

The increasing availability of electronic text presents opportunities for automated acquisition of computer-interpretable knowledge. Two fundamental issues related to the application of such automatically extracted knowledge are how best to represent and reason with it. It differs from knowledge in manually curated resources in several respects, the most obvious being in scale, indicating a need for reduced-dimensional representations and scalable inference methods. In addition, it contains distributional information, as the same assertion may be extracted from multiple contexts, and is unlikely to be perfectly accurate. This suggests a need for a continuous-valued alternative to the discrete estimates of truth or falsehood provided by symbolic logical inference. Previously, we have addressed these issues using a geometric approach to representation known as Predication-based Semantic Indexing (PSI) [1] that mediates a scalable form of approximate inference [2]. However, the association strengths estimated by PSI cannot be interpreted probabilistically, which is desirable from a theoretical perspective, as well as for practical purposes such as their integration with probabilities from other sources. Consequently in this paper, we develop a probabilistic interpretation of PSI and the operators that it mediates, revealing a deeper relationship to the probabilistic calculus of quantum mechanics than had been elucidated previously.

2 Background

2.1 Distributional Semantics and Predication-based Semantic Indexing

Methods of distributional semantics learn the relatedness between terms from their distribution across large electronic text collections [3]. A commonly used methodological approach involves generating vector space models of corpora in which terms are represented as vectors derived from the contexts in which they occur, such that terms occurring in similar contexts will have similar representations. Such models have shown remarkable successes in simulating human behavior on certain cognitive tasks (see for example [4]). However, they do not encode the nature of the relationship between terms, and therefore are limited in their utility as a means to model analogical reasoning, or support logical operations. Predication-based Semantic Indexing (PSI) [1] was developed to model sets of concept-relation-concept triplets, known as *semantic predications*, extracted from the biomedical literature by a natural language processing system known as SemRep [5]. PSI concept vectors encode both the nature and the distribution of the predications in which a concept occurs. Consequently, a PSI space can provide answers to questions such as “what TREATS schizophrenia?”. Later iterations of PSI used reversible vector transformations to mediate analogical inference, used for predictive modeling in drug repurposing and other applications (for a review, see [2]).

2.2 Quantum-inspired Operations and Interpretations

Several authors have provided interpretations of vector space models of distributional semantics that relate to quantum theory. Widdows and Peters develop distributional models of negation and disjunction, using the connectives of quantum logic [6]. Aerts and Czachor draw parallels between the application of the Singular Value Decomposition in Latent Semantic Analysis [4] and spectral decomposition in quantum physics, and show how the relatedness between terms within a set of sentences can be represented as a density matrix [7]. Bruza and his colleagues draw an analogy between word vectors in distributional models and state vectors in quantum mechanics, from which it follows that context-specific associations of a term might be revealed by a process analogous to collapse of a state vector upon measurement [8].

Regarding PSI, the utility of quantum-logical operators has been clearly demonstrated across several experiments, and aspects of the model have been interpreted with respect to their relationship to quantum mechanics. Vector space equivalents of disjunction and negation, described as quantum logic by Birkhoff and von Neumann [9], and applied previously in information retrieval experiments [6], have been applied to PSI spaces to direct search toward concepts of interest [10], and evaluate the relatedness between concepts across multiple predicate paths with improved performance in predictive modeling experiments [11]. In analogical retrieval experiments, performance was improved with compound cue vectors generated by superposing cue vectors derived from several pairs of cue concepts [12]. This phenomenon was interpreted with respect to its relationship to entanglement. Implicit in this interpretation is the notion that the semantic concept vectors generated by PSI are analogous to the state vectors used in quantum mechanics to estimate the probabilities of different outcomes. In addition, we have developed a complex vector based implementation of PSI [11], which bears some relation to quantum mechanics also. However, PSI was not conceived with quantum mechanics in mind, and to date our adoption of quantum-related mathematics has been primarily pragmatically motivated, and arguably somewhat ad-hoc.

In this paper we aim to provide a more extensive account of the relationships between PSI and quantum mechanics. To do so, we develop a probabilistic interpretation of PSI, through which the semantic distance metrics derived by the method can be interpreted with respect to the probabilities of encountering particular predications in the collection from which the space was derived. In the vector space model for information retrieval, the cosine similarity $q \cdot d$ between a query q and a document d can naturally be interpreted as the probability that d is relevant to q, and the correspondence between this interpretation and Born's rule in quantum mechanics has been recognized for some years [13]. This paper takes this parallel further with the key mathematical observation that when searching over multiple relationships at once in a vector representation, the superposition operation used naturally in vector spaces gives rise to a probabilistic interpretation based on squared amplitudes, and so the way probabilities are combined follows the rules of quantum rather than classical probability.

The remainder of the paper proceeds as follows. First, we describe the complex vector based implementation of PSI, and the modifications required to facilitate a probabilistic interpretation. We then provide probabilistic interpretations of PSI-based methods of retrieval and inference, present an evaluation of our modified implementation, and conclude with a discussion of the implications of this work.

3 Mathematical Structure and Methods

3.1 Circular Holographic Reduced Representations (CHRR)

PSI derives vector representations of concepts by superposing vector products representing concept-predicate pairs, such as "ISA fluoxetine". These vector products are composed using reversible vector transformations provided by a family of representational approaches known as Vector Symbolic Architectures or VSAs [14], and are commonly referred to as "binding", represented with the symbol $\otimes$, and the inverse of binding (or "release") represented with the symbol $\oslash$. A range of VSAs with suitable binding and superposition operations can be implemented over various number fields including the real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, and binary numbers $\mathbb{Z}_2$. As we have argued previously, the use of complex vectors is standard in quantum theory, but comparatively unexplored in information retrieval and distributional semantics [15].

The longest-established VSA using complex numbers is Plate's Circular Holographic Reduced Representation (CHRR) [16, 17]. CHRR uses complex vectors each of whose coordinates is a number on the unit circle $U(1)$ in the complex plane, thus the space of available vectors in n dimensions is $U(1)^n \subset \mathbb{C}^n$. We will refer to such a complex vector as a *circular vector*. For our present purposes, this architecture provides both a binding operator with an exact inverse such that $A \otimes B \oslash A = B$ (which is not the case for Holographic Reduced Representations in general), and a continuous-valued vector space representation conducive to a probabilistic interpretation (this is more difficult with the Binary Spatter Code (BSC) [18], another influential VSA).

Binding ($\otimes$) in CHRR is accomplished by pairwise multiplication, the natural group operation on $U(1)^n$, so that $X \otimes Y = \{X_1Y_1, X_2Y_2, ...X_{n-1}Y_{n-1}, X_nY_n\}$. This is equivalent to addition of the phase angles of the unit length circular vectors concerned. Release ($\oslash$) is accomplished by binding to the inverse of a circular vector, where the inverse of a vector is its complex conjugate. PSI also requires a superposition operator ($+$). In CHRR this is accomplished by pairwise addition of unit circle vectors, with subsequent normalization of each circular component back to unit length. In practice, for pairwise addition, this is just the average of the phase angles, which as a group addition operation suffers from the obvious objection that it is not associative, so that the elements added later have more significance than those added earlier. In some cases, this preference for more recent items may be desirable, for example, in modelling short-term memory [19]. Where it is not desirable, its effects can be mitigated by storing several vectors and superposing them in a batch: for example, in the case of PSI, normalization occurs after training concludes, so the sequence in which superposition occurs is not relevant. The benefits of CHRR in this form are partly computational: only one number needs to be stored for each complex dimension, binding and superposition are fast and simple operations, and sparse representations can be supported without undue difficulty [17, 15]. However, the restriction from $\mathbb{C}^n$ to $U(1)^n$ has been found unsuitable for a fuller probabilistic interpretation, and as described later in the paper, we have extended the CHRR model by relaxing the requirement that coordinates lie on the unit circle.

3.2 Predication-based Semantic Indexing (PSI))

PSI is based upon the random indexing paradigm [20], in which basic *elemental vectors* representing terms, concepts or documents are superposed to generate *semantic vector* representations. In high dimensions, randomly-chosen vectors make suitable elemental vectors thanks to their high probability of being mutually almost orthogonal. With CHRR, elemental vectors are initialized by randomly assigning a phase angle to each of a user-defined number of dimensions which determine the dimensionality of the resulting PSI space. Elemental vectors are generated for each concept $E(\text{concept})$, and each relation type $E(\textsc{predicate})$ and its inverse $E(\textsc{predicate-inv})$. Semantic vectors are learned gradually by superposing the bound products of elemental vectors representing predicate-argument pairs. For example, encoding a single instance of the predication "prozac ISA fluoxetine" is accomplished as follows:

$$S(\text{prozac}) \mathrel{+}= E(\text{ISA}) \otimes E(\text{fluoxetine})$$
$$S(\text{fluoxetine}) \mathrel{+}= E(\text{ISA-INV}) \otimes E(\text{prozac})$$

Thus, the semantic vector for prozac encodes the assertion that it is (the brand name of) fluoxetine, and the semantic vector for fluoxetine encodes the assertion that it has the hyponym prozac. As the same predication may be extracted from many documents, the generated vectors encode distributional information. So statistical weighting metrics are often applied such that the extent to which a predication contributes to a semantic vector is some function of its frequency and the global frequency of the other concept and/or predicate concerned. After training is complete, the net result is a set of semantic concept vectors derived from the predications each concept occurs in. Elemental predicate vectors for each concept and predicate (and its inverse) are also retained.

4 Embedding Probabilities in Predication Space

This section describes how the CHRR model using vectors in $U(1)^n$ was extended to use more general vectors in $\mathbb{C}^n$. This supports the probabilistic operations we depend on, and also makes the architecture much more similar to the Hilbert space models used standardly in quantum mechanics.

Scaling: In accordance with Plate's original description of CHRR [16], in previous work we have normalized semantic vectors at the component level after training. This ensures that the vectors produced during training consist of numbers on the unit circle in the complex plane. To accommodate a probabilistic interpretation, we instead apply normalization to the vector as a whole, such that $\|S(\mathrm{C})\| = 1$ after training. A consequence of this modification is that the components of our vectors do not necessarily fall on the unit circle of the complex plane.

Binding: Binding and release are still conducted by adding the phase angles of the components concerned (as though they were unit circle vectors). However, with semantic vectors the radii of the circular components are likely to differ on account of the training process, and are retained. During binding and release, the magnitudes of corresponding circular components are multiplied. As elemental vectors are constructed with unit length circular components, this particularly affects operations involving pairs of semantic vectors, or their superpositions.

Vector Comparison: In Plate's original model, and in our previous work, comparison between vectors is conducted at the component level, such that for two k-dimensional vectors V_1 and V_2, $\mathrm{sim}(V_1, V_2) = \frac{1}{k}\sum_i^k \cos(\mathrm{V}_1^{\mathrm{i}}, \mathrm{V}_2^{\mathrm{i}})$, the mean pairwise cosine distance between components. For our current purposes, we instead employ the hermitian inner product between normalized circular vectors.

Weighting: Instead of the heuristically-motivated approaches we have utilized previously, such as predication-level TF-IDF weighting, the square root of the number of times a predication occurs is used as a weighting metric. As we will subsequently illustrate, this step is a prerequisite to the recovery of the probabilities of events of interest using quantum mechanical operators. The training operations for a predication C_1 PRED C_2 with *pf* instances in the knowledge base then becomes:

$$S(\mathrm{C}_1) \mathrel{+}= E(\mathrm{PRED}) \otimes E(\mathrm{C}_2) \times \sqrt{pf}$$
$$S(\mathrm{C}_2) \mathrel{+}= E(\mathrm{PRED\text{-}INV}) \otimes E(\mathrm{C}_1) \times \sqrt{pf}$$

For the remainder of the paper we will refer to CHRR with these modifications as Hermitian Holographic Reduced Representations (HHRR), and refer to HHRR-based PSI with the modified weighting metric we have described as probabilistic PSI (pPSI).

Table 1. "mdd" = Major Depressive Disorder. "ssri" = Selective Serotonin Reuptake Inhibitor.

predication	frequency	predication	frequency
ssri TREATS mdd	64	prozac ISA ssri	16
fluoxetine TREATS mdd	36	prozac ISA antidepressive_agents	4
prozac ISA fluoxetine	25	prozac TREATS anxiety_disorders	4

Table 2. Non-zero entries of mdd vector in unreduced concept-concept-predicate matrix.

	TREATS-INV ssri	TREATS-INV fluoxetine
mdd	$\sqrt{\frac{64}{100}} = 0.8$	$\sqrt{\frac{36}{100}} = 0.6$

4.1 Estimating probabilities

A pPSI space can be considered as a reduced-dimensional approximation of a weighted and normalized term-by-term-by-predicate matrix, which we will refer to directly for illustrative purposes. Consider the small collection of predications in Table 1. In the case of this collection and $S(\text{mdd})$, the non-zero values of the matrix that is being approximated after normalization of the semantic vectors are shown in Table 2. The probability of drawing the predication "ssri TREATS mdd" from the pool of predications concerning "mdd" is then equal to the squared length of the projection of $S(\text{mdd})$ on the "TREATS-INV ssri" axis. In symbols, the correspondence between this probability, $|S\rangle = S(\text{mdd})$, and the basis vectors $|X\rangle$ representing TREATS-INV ssri and $|Y\rangle$ representing $\text{TREATS-INV fluoxetine}$ can be expressed as follows:

$$
\begin{aligned}
|S\rangle &= (\sqrt{\tfrac{64}{100}}).\,|X\rangle + (\sqrt{\tfrac{36}{100}}).\,|Y\rangle \\
P(\text{ssri TREATS mdd}) = &\quad \|\,|X\rangle\langle X|S\rangle\,\|^2 = \tfrac{64}{100}
\end{aligned}
$$

To derive a pPSI space from the collection of predications in Table 1, $S(\text{mdd})$ is generated by superposing the bound products of high-dimensional elemental circular vectors representing TREATS-INV ssri and $\text{TREATS-INV fluoxetine}$. Unlike $|X\rangle$ and $|Y\rangle$, these bound products may only be approximately orthogonal to one another. Nonetheless, the dimensionality of the space concerned ensures that random elemental vectors, and hence their bound products, are mutually orthogonal or close-to-orthogonal with high probability [2]. So, for example, over 1,000 simulations with different random initializations of a 1,000-dimensional pPSI space the squared length of the projection of $S(\text{mdd})$ onto the vector product $E(\text{TREATS-INV}) \otimes E(\text{ssri})$ approximates $\frac{64}{100}$, as illustrated in Table 3. This is, of course, the probability of drawing "ssri TREATS mdd" from the collection of 100 mdd-related predications in Table 1.

4.2 Disjunction and Negation

Similarly, the probability of drawing one of a specified set of predications can be determined by the squared length of the projection of a semantic (cf. state) vector onto the subspace spanned by the vectors representing the component predicate-argument pairs of interest. If $|S\rangle$ represents $S(\text{prozac})$ and the basis vectors $|W\rangle$, $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ represent predicate-argument pairs ISA fluoxetine; ISA ssri; ISA $\text{antidepressive agents}$ and TREATS anxiety disorders respectively, then:

$$|S\rangle = (0.7143).\,|W\rangle + (0.5714).\,|X\rangle + (0.2857).\,|Y\rangle + (0.2857).\,|Z\rangle$$

The probability of drawing either ISA fluoxetine or ISA ssri is the square of the length of the projection of $|S\rangle$ onto the subspace spanned by $|W\rangle$ and $|X\rangle$. This projection can be expressed as $(|W\rangle\langle W| + |X\rangle\langle X|)\,|S\rangle = |W\rangle\langle W|S\rangle + |X\rangle\langle X|S\rangle = (0.7143).\,|W\rangle + (0.5714).\,|X\rangle$, with a squared length of $0.7143^2 + 0.5714^2 = 0.8367 = \frac{(25+16)}{49}$, the probability of drawing either "prozac ISA fluoxetine" OR "prozac ISA ssri" from the collection of 49 prozac-related predications in Table 1. This probability can also be estimated from a pPSI space derived from the predications in Table 1 as the squared length of the projection of $S(\text{prozac})$ onto an orthonormal subspace constructed by applying the Gram-Schmidt procedure to bound products $E(\text{ISA}) \otimes E(\text{fluoxetine})$ and $E(\text{ISA}) \otimes E(\text{prozac})$. Over 1,000 simulations with different random initializations the mean squared length of this projection was 0.8370 ± 0.0048.

It follows that the probability of drawing something other than ISA fluoxetine or ISA ssri is the squared length of the vector $|S\rangle - (|W\rangle\langle W| + |X\rangle\langle X|)\,|S\rangle$, the projection of $|S\rangle$ onto a subspace orthogonal to that spanned by $|W\rangle$ and $|X\rangle$. The squared length in this case is $\|(0.2857).\,|Y\rangle + (0.2857).\,|Z\rangle\|^2 = 2 \times 0.2857^2 = 0.1632 = \frac{8}{49}$.

4.3 Logical Leaps

An appealing feature of PSI is the capacity for efficient yet approximate inference across multiple reasoning pathways simultaneously. This is accomplished by transforming the task of exploring these pathways into the task of comparing the similarity between concept vector representations. Consider once again the collection of predications in Table 1, and the task of estimating the strength of the indirect relationship between the concepts "prozac" and "mdd" across the predicate path ISA:TREATS. One way to think of this task is as the estimation of the probabilities of a set of transitions, shown in Figure 1. Though this is a schematic representation, it is worth noting that any vector in the PSI space can be interpreted as a position, and so the notion of a journey from concept to concept is arguably more literal than analogical.

One way of estimating the probability of "getting from" prozac "to" mdd along this pathway would be to use a Markovian approach in accordance with the laws of classical probability. Specifically, we assume the probability of reaching a destination

Table 3. Simulations in 1000-dimensional Complex PSI Space. "mdd" = Major Depressive Disorder. "ssri" = Selective Serotonin Reuptake Inhibitor.

vector	bound product	$\overline{x}$ similarity	$\overline{x}$ probability
$S(\text{mdd})$	$E(\text{TREATS}_{\text{INV}}) \otimes S(\text{ssri})$	0.8001 ± 0.0049	0.6402 ± 0.0078
$S(\text{mdd})$	$E(\text{TREATS}_{\text{INV}}) \otimes S(\text{fluoxetine})$	0.6002 ± 0.0116	0.3603 ± 0.0139

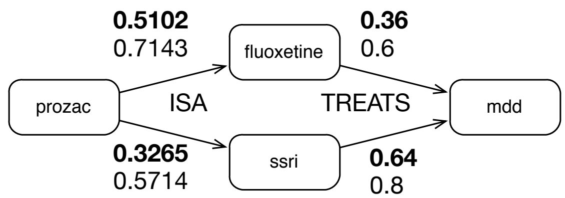

Fig. 1. Transition **Probabilities** and Amplitudes

reachable by two disjoint alternative paths is the sum of the probabilities of each of the paths, where the probability of a path is the product of the probabilities of the transitions along this path.[3] The probability of our semantic journey would then be (0.5102×0.36)+(0.3265×0.64) = 0.3926. Alternately, we could utilize Feynman's rules for calculating probabilities across multiple indistinguishable paths [21], applied to model cognitive phenomena in [22]. In contrast to the Markov approach, probability would then be estimated as the square of the sum of the path *amplitudes*, which are the products of the component amplitudes in each path - the lengths of the projections of the (semantic) state vectors onto the relevant (elemental) basis vectors. With $|S_p\rangle$ representing the vector product $S(\text{prozac}) \oslash E(\text{ISA})$ and $|S_m\rangle$ representing the vector product $S(\text{mdd}) \oslash E(\text{TREATS} - \text{INV})$, and the basis vectors $|E\rangle$, $|F\rangle$ representing elemental vectors $E(\text{ssri})$ and $E(\text{fluoxetine})$ respectively, the component amplitudes are $\| \, |E\rangle\langle E|S_p\rangle \, \| = 0.5714$; $\| \, |E\rangle\langle E|S_m\rangle \, \| = 0.8$; $\| \, |F\rangle\langle F|S_p\rangle \, \| = 0.7143$; and $\| \, |F\rangle\langle F|S_m\rangle \, \| = 0.6$. The combined probability of our semantic journey would then be $\big((0.5714 \times 0.8) + (0.7143 \times 0.6)\big)^2 = 0.7845$.

Which, if any, of these probabilities correspond to the estimation of distance in pPSI? Distance across a specified predicate path in PSI is measured by binding the semantic concept vector for one concept to the inverse of the product of the "release" operation on elemental vectors representing the predicates concerned, and comparing the result to the other semantic concept vector. With pPSI, for our example this would be accomplished by measuring the inner product of the semantic vector $|S_m\rangle = S(\text{mdd})$ and the vector $|S_{pit}\rangle = S(\text{prozac}) \oslash E(\text{ISA}) \otimes E(\text{TREATS-INV})$. This gives the length of the projection of one vector onto the other, $\langle S_{pit}|S_m\rangle = \langle S_m|S_{pit}\rangle$, and squaring this length gives an estimate of the probability of the journey from one concept to another across this path, $\langle S_{pit}|S_m\rangle^2$. Over 1,000 simulations with different random initializations, the mean squared length of this projection was 0.7843 ± 0.0075, corresponding to the probability estimated using rules of quantum probability.

5 Evaluating HHRR

That HHRR more accurately preserves probability amplitudes during encoding suggests it may offer advantages over CHRR in predictive modeling experiments. To eval-

[3] We consider paths to be disjoint if they do not share any points along the path except for the beginning and end. Combining the probabilities of paths with intersections is more difficult, but in the case of 2-step paths, distinct paths are always disjoint in this sense.

uate this hypothesis, we repeated part of the experiments documented in [23] in which the length of the projection of a drug's semantic vector into a subspace derived from vector representations of ten PSI reasoning pathways was used to rank order a set of 1398 pharmaceutical agents with respect to their likely activity against prostate cancer cells in high-throughput screening experiments. In the original experiments, which used the BSC [18] as a VSA, reasoning pathways were inferred from known therapeutic relationships. We re-used the pathways from the "Knowledge Withheld" condition, in which predications directly linking a pharmaceutical agent and a type of cancer were withheld from the model. Unlike HRR, the binding operator in the BSC is its own inverse, so directionality of a predicate pathway is not encoded. Consequently, we used the pathway directionalities suggested in Figure 2 of the original paper. We generated 2000, 4000 and 8000-dimensional CHRR and HHRR spaces from version 24.2 of the publicly available SemMedDB database of semantic predications [24], containing 70,364,020 predications extracted from 23,921,088 MEDLINE citations. Concepts occurring 500,000 times or more in the set were excluded from the analysis, and only predications involving predicates in the set {AFFECTS, ASSOCIATED_WITH, AUGMENTS, CAUSES, COEXISTS_WITH, DISRUPTS, INHIBITS, INTERACTS_WITH, ISA, PREDISPOSES, PREVENTS, SAME_AS, STIMULATES, TREATS} were encoded. To accommodate inference across triple-predicate pathways, we also created *second-degree semantic vectors* [25] for the cue concepts "hormone-refractory_prostate_cancer" and "prostate_carcinoma" as weighted superpositions of the semantic vectors for all concepts that occurred as the subject of an ASSOCIATED_WITH relationship with them. These superposition operations, and those occurring during training, were weighted using the square root of the predication frequency, which was kept constant across models to isolate differences between CHRR and HHRR. Elemental vectors were generated using the deterministic procedure described in [26], so were identical across models with common dimensionality.
For each cue concept, a subspace was generated by applying the Gram-Schmidt procedure to a set of ten vectors, each constructed following the pattern $S(\text{disease}) \oslash E(\text{PRED}1) \otimes E(\text{PRED}2)$, where $S(\text{disease})$ represents either the first- or second-order semantic vector for one of the cue concepts. The semantic vectors of the 1398 pharmaceutical agents in the set were then projected into these subspaces, to measure the strength of their relatedness to the cue concepts across the reasoning pathways from [23]. The rank of each of the 68 agents that were active against PC3 cells with a growth rate of 1.5 standard deviations or more below the average across agents was then evaluated. The results of these experiments are shown in Table 2, which gives the area under the receiver operator characteristic curve (AUROC) for each model, with an AUROC of 0.5 anticipated with a random ordering of the agents. HHRR outperforms CHRR across both cue concepts and at all dimensionalities, supporting the hypothesis that the additional representational power it provides offers advantages for predictive modeling.

6 Discussion

In this paper, we developed a method through which the probabilities of encountering particular predications in a collection are embedded in a reduced-dimensional geomet-

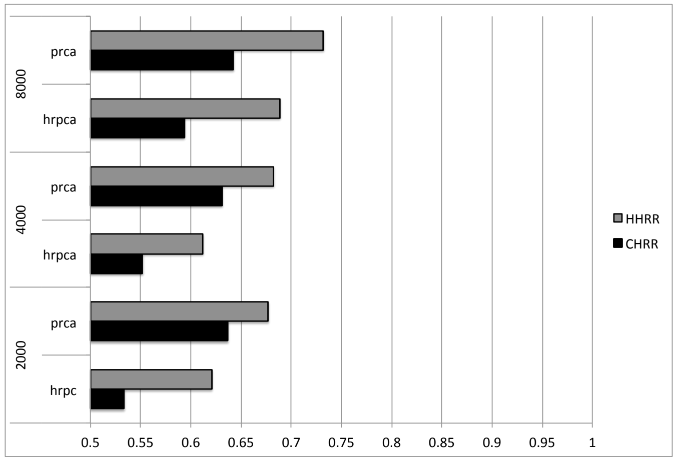

Fig. 2. Predictive modeling experiments. X axis = AUROC (≈ 0.5 with random selection). Y axis = dimensionality. prca="prostate carcinoma". hrpca="hormone-refractory_prostate_cancer".

ric space. This reveals a new perspective on PSI, in which the logical operations it mediates can be interpreted probabilistically, providing a principled way to combine PSI's estimates with estimates of probability derived from other sources. It also provides the means to define the meaning of the measures of association that PSI produces. Predicate-specific strength of association between concepts that are directly related (e.g. between $S(\mathrm{mdd})$ and $E(\mathrm{fluoxetine}) \otimes E(\mathrm{TREATS})$ can now be defined as the magnitude of the probability amplitude of drawing a particular predication from the set of predications involving a concept. Disjunction and negation operators applied to PSI spaces can now also be interpreted probabilistically, providing a closer correspondence to Birkhoff and von Neumann's quantum logic of "experimental propositions" [9] than when these operators yield abstract similarity metrics that permit a geometric interpretation only. In addition, the association strengths measured using "logical leaps" across dual-predicate paths that subsume multiple relationships can be interpreted probabilistically based on squared amplitudes, in accordance with the rules of quantum probability. Of note, these measurements are conducted without recourse to the elemental vectors representing the middle terms that lie along a predicate path. Consequently, the possible pathways are indistinguishable at the time the estimation is made, a prerequisite to the application of squared sum of amplitudes to estimate path probabilities [21].

To mediate this interpretation, we developed a novel VSA variant, HHRR. HHRR is a modification of CHRR, and shows advantages over it in predictive modeling experiments. These advantages may be attributed to additional representational power provided by relaxing the constraint that semantic vector coordinates lie on the unit circle of the complex plane. As elemental vector coordinates still maintain this constraint, this results in a delegation of representational duties in which semantic and distributional information are encoded by the phase angles and magnitudes of circular vectors respectively. This raises implementation issues, such as the extent to which the discretized representation of phase angles (described in [17]) affects the resolution with which semantic information is encoded, which we will explore in future work. As noted by De Vine, superposition in CHRR suggests an interpretation involving interference effects [17] - superposing circular components with similar phase angles will result in positive interference, while components with dissimilar phase angles will exhibit negative interference. The effects of these interference effects on the relative magnitudes of the circular components are preserved by HHRR, but not by CHRR, after normalization. Our empirical results suggest HHRR is better positioned to encode distributional information. This was not the primary concern of Plate's original work, which focused on the encoding and retrieval of representations of combinations of discrete symbols [16], but is important for applications in distributional informatics and information retrieval. Predictive modeling may be further improved through application of supervised machine learning approaches to pPSI vectors, which we will explore in future work.

7 Conclusion

We developed a probabilistic interpretation of PSI, revealing previously unrecognized connections to quantum theory. This development required generation of a novel VSA variant, with improved performance in predictive modeling experiments.

Acknowledgments: This research was supported by US National Library of Medicine grant R01 LM011563. We would like to acknowledge Lance De Vine, for contributing the CHRR implementation that was adapted in the current research.

References

1. T. Cohen, R. Schvaneveldt, and T. Rindflesch, "Predication-based semantic indexing: Permutations as a means to encode predications in semantic space," *AMIA Annu Symp Proc.*, pp. 114–8, 2009.
2. D. Widdows and T. Cohen, "Reasoning with vectors: a continuous model for fast robust inference," *Logic Journal of IGPL*, p. jzu028, Nov. 2014.
3. T. Cohen and D. Widdows, "Empirical distributional semantics: methods and biomedical applications," *Journal of Biomedical Informatics*, vol. 42, pp. 390–405, Apr. 2009.
4. T. K. Landauer and S. T. Dumais, "A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, pp. 211–240, 1997.
5. T. C. Rindflesch and M. Fiszman, "The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text," *Journal of Biomedical Informatics*, vol. 36, pp. 462–477, 2003.
6. D. Widdows and S. Peters, "Word Vectors and Quantum Logic Experiments with negation and disjunction," in *Proceedings of Mathematics of Language 8* (R. T. Oehrle and J. Rogers, eds.), vol. 8, 2003.
7. D. Aerts and M. Czachor, "Quantum aspects of semantic analysis and symbolic artificial intelligence," *J. Phys. A: Math. Gen.*, vol. 37, pp. L123–L132, 2004.

8. P. D. Bruza and R. J. Cole, "Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning," in *Artemov, S., Barringer, H., d'Avila Garcez, A.S., and Woods, J.H. (Eds.). We Will Show Them: Essays in Honour of Dov Gabbay.*, pp. 339–361, College Publications, London, UK., 2005.
9. G. Birkhoff and J. von Neumann, "The logic of quantum mechanics," *Annals of Mathematics*, vol. 37, pp. 823–843, 1936.
10. T. Cohen, D. Widdows, S. R., and T. Rindflesch, "Logical leaps and quantum connectives: Forging paths through predication space," in *Proc AAAI Fall Symposium on Quantum Informatics for Cognitive, Social, and Semantic Processes (QI 2010)*, 2010.
11. T. Cohen, D. Widdows, L. De Vine, R. Schvaneveldt, and T. C. Rindflesch, "Many Paths Lead to Discovery: Analogical Retrieval of Cancer Therapies," in *Quantum Interaction* (J. R. Busemeyer, F. Dubois, A. Lambert-Mogiliansky, and M. Melucci, eds.), no. 7620 in Lecture Notes in Computer Science, pp. 90–101, Springer Berlin Heidelberg, Jan. 2012.
12. T. Cohen, D. Widdows, R. Schvaneveldt, and T. Rindflesch, "Finding schizophrenia's prozac: Emergent relational similarity in predication space," *QI'11. Proc 5th International Symposium on Quantum Interactions. Springer-Verlag Berlin, Heidelberg.*, 2011.
13. K. van Rijsbergen, *The Geometry of Information Retrieval*. Cambridge Univ Press, 2004.
14. R. W. Gayler, "Vector symbolic architectures answer jackendoff's challenges for cognitive neuroscience," in *In Peter Slezak (Ed.), ICCS/ASCS International Conference on Cognitive Science*, (Sydney, Australia. University of New South Wales.), pp. 133–138, 2004.
15. Widdows, D, Cohen, T, and DeVine, L, "Real, Complex, and Binary Semantic Vectors," in *Quantum Interaction* (J. R. Busemeyer, F. Dubois, A. Lambert-Mogiliansky, and M. Melucci, eds.), no. 7620 in LNCS, Paris, France: Springer Berlin Heidelberg, 2012.
16. T. A. Plate, *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. Stanford, CA.: CSLI Publications, 2003.
17. L. De Vine and P. Bruza, "Semantic oscillations: Encoding context and structure in complex valued holographic vectors," *Quantum Informatics for Cognitive, Social, and Semantic Processes (QI 2010)*, 2010.
18. P. Kanerva, "The spatter code for encoding concepts at many levels," in *ICANN*, (London), pp. 226–229, Springer, 1994.
19. P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cognitive Computation*, vol. 1, no. 2, pp. 139–159, 2009.
20. P. Kanerva, J. Kristofersson, and A. Holst, "Random indexing of text samples for latent semantic analysis," *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, vol. 1036, 2000.
21. R. Feynman and A. Hibbs, *Quantum mechanics and path integrals*. MacGraw Hill; New York, 1965.
22. J. R. Busemeyer, Z. Wang, and A. Lambert-Mogiliansky, "Empirical comparison of markov and quantum models of decision making," *Journal of Mathematical Psychology*, vol. 53, no. 5, pp. 423–433, 2009.
23. T. Cohen, D. Widdows, C. Stephan, R. Zinner, J. Kim, T. Rindflesch, and P. Davies, "Predicting high-throughput screening results with scalable literature-based discovery methods," *CPT: pharmacometrics & systems pharmacology*, vol. 3, no. 10, p. e140, 2014.
24. H. Kilicoglu, D. Shin, M. Fiszman, G. Rosemblat, and T. C. Rindflesch, "Semmeddb: a pubmed-scale repository of biomedical semantic predications," *Bioinformatics*, vol. 28, no. 23, pp. 3158–3160, 2012.
25. T. Cohen, D. Widdows, R. Schvaneveldt, and T. Rindflesch, "Discovery at a distance: Farther journeys in predication space.," in *Proc First International Workshop on the role of Semantic Web in Literature-Based Discovery (SWLBD2012)*, (Philadelphia, PA), Oct. 2012.

26. M. Wahle, D. Widdows, J. R. Herskovic, E. V. Bernstam, and T. Cohen, “Deterministic Binary Vectors for Efficient Automated Indexing of MEDLINE/PubMed Abstracts,” *AMIA Annual Symposium Proceedings*, vol. 2012, pp. 940–949, Nov. 2012.