Middle East Technical University
Department of Computer Engineering

An Approach to Modeling Context Using Semantic Maps

Mehmet Çelik, Sinan Kalkan

METU-CENG-TR-2017-03

March 2017

Department of Computer Engineering
Middle East Technical University
Dumlupınar Bulvarı, 06531, Ankara
TURKEY

Technical Report

This page contains a Turkish translation of the title and the abstract of the report. The report continues on the next page.

Bağlam Modellemeye Semantik Harita Yaklaşımı

Mehmet Çelik, Sinan Kalkan

Bilgisayar Mühendisliği Bölümü
Ortadoğu Teknik Üniversitesi
Dumlupınar Bulvarı, 06531, Ankara
TÜRKİYE

Öz

Tam bir tanımını yapmak zor olsa da, bağlam, anlamın bağlı olduğu bilgi olarak görülebilir. İnsanlar ile etkileşmesini beklediğimiz yapay bir etmen için, çevresi ve insanlar ile olan etkileşimin başarısı büyük oranda bağlama bağlıdır. Bu rapor, bağlama semantik uzay modelleri ve komutatif ve dağıtık olmayan olasılık çerçeveleri ile yaklaşmayı araştırmaktadır.

Abstract

Although a precise definition is hard to pinpoint, context can be seen as the information that meaning is dependent on. For an artificial agent that is expected to interact with humans, the success of its interactions with both the environment and humans depends on context. This report aims to investigate a unified framework for context in robots, inspired from semantic space models and a non-commutative and non-distributive probability framework.

1 Introduction

As Harnad stated, categorization of objects, events, and actions, is quite essential for cognition:

> "To Cognize is to Categorize: Cognition is Categorization" – S. Harnad [16]

However, in order to be able to classify, preliminary information of what is being classified, an abstraction, must already exist. These abstractions, called concepts, make us understand and give meaning to categories: Put another way, we can understand words with concepts, identify objects, perceive events. We exist with cognitive mechanisms built on top of concepts [3, 6].

Although concepts are indispensable for cognition, it is suggested that concepts cannot be separated from the context they exist in [4]. Context is defined in the Oxford Dictionary as:

> "The circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood"
> "The parts of something written or spoken that immediately precede and follow a word or passage and clarify its meaning."

This and similar context definitions are available in many resources, however there are critical differences between those definitions [5]. The identified differences are:

- Defining context externally or internally.
- Defining context as a set of information or as processes.
- Defining context as static or dynamic.
- Defining context as a simple set of phenomenon or an organized network.

Although there are different definitions, the common point in those definitions is the importance of context while defining and interpreting objects, events and phenomena [34, 32]. For example, in what manner we carry a cup may depend on the temperature of the liquid in it, how much liquid it has, whether we are in a hurry, if there are obstacles around. Or the manner we speak with a neighbor we meet every day can be affected by whether we are in a hurry, our health at that moment or our mood. The sentences we construct to communicate and the meaning we derived from them can change in different contexts.

Certainly, the connection between context and concept, which is essential for understanding cognition, is also essential for both artificial intelligence and robotics. In the first studies on concepts, definitions of concept already suggest the similarities and differences between context and concepts. Psychological findings and recent studies that define context as a class of concept [9] also suggest the similarity of context and concept.

It can be argued that the established formalisms of context in AI are focused on representation of the background information that will affect the agent [28]. In this respect, they are static and descriptive systems. Also the recent success stories from AI generally follow an connectionist approach. For an autonomous robot as an intelligent agent in an environment with humans, the need for a unified framework of context that represent background knowledge, can incorporate learning and interface with connectionist systems is identified as a need and will be the goal of this study.

2 Background

This section introduces the necessary background, for the proposed study, on concepts, context and quantum probability.

2.1 Concepts

The definition and the nature of concept is discussed for centuries since the times of Aristotle and Plato. Different theories are suggested [12, 27] and these theories can be classified under three main categories[17]:

- Rule based concepts: In this approach, concepts are defined as a set of rules which are hard thresholds. For example, properties of an object is represented as the rule "color = yellow AND $10cm <$ height $< 12cm$" In this approach, the membership of a thing is determined by which rules are satisfied by the object properties.
- Prototype based concepts: In this approach, concepts are defined by a prototypical object that summarizes the properties of the whole set. In this concentrated representation, properties are represented as the distribution of possible properties such as "color = %40 yellow, %30 blue, %30 black". Membership of an object is determined by comparing it to the prototypical object.
- Example based concepts: In this approach, concepts are defined by a set of observed objects. A general definition is not defined. Membership of an object is determined by comparing it to the whole set of examples.

Different studies and approaches suggests that concepts are not isolated definitions but has a connected nature. This can be as a hierarchical structure of concepts [13, 19, 11], or as a more general graph structure where different kinds of connections are present [8]. For example, in a hierarchical structure "horse" and "sheep" concepts are connected under "vertebrate" concept. In a more complex concept web, "cup" and "plate" concepts can be connected as both carry the property "hard".

2.2 Context

Most generally, context can be defined as any and all information, in any time frame, that can affect the meaning/understanding of the symbols or objects that has been observed.

2.2.1 Context in AI

Two approaches for context put forward by Ghidini [14] and McCarthy and his team [21] can be argued to be the two mature approaches in AI for context [28].

Propositional logic of context is the framework that is developed by McCarthy, who tried to define context for agents in a logical formalism [21]. McCarthy claims that context have following properties:

- Contexts are formal objects.
- Context can be defined exactly.
- There are connection between contexts and contextual function, and a new context can be constructed from another by changing certain properties such as time, place, and state.

The second approach is local model semantics or multi-context systems [14]. In this approach, context is mainly a subset of the global state of the individual's belief/information about the world. They identify two natures of context, namely metaphysical context and cognitive context. Metaphysical context is the structure of the world modeled around a speaker and a listener. On the other hand cognitive context is the representation of the world within an individual. The idea is, reasoning happens locally in a cognitive context that is also a subset of agents global state. Here, different contexts are not unrelated belief sets, but different representations of the world. This difference may be difference in perspective or in difference in levels of detail. Also the relations between context can only be partial, meaning there is no complete translation between representations of two different contexts.

2.2.2 Context in Computer Vision and Visual Cognition

In computer vision, the nature of context can be simplified by a relation between foreground and background of an image. Also context is noted as any and all information that influence the perception of a scene [31]. With this definition, physical, photogrammetric and computational context is defined. Physical context is the real world situation the image is taken in, ranging from the geometry of the scene to the weather the image is taken. The photogrammetric context is the parameters of the device that is taking the picture. These can be exposure of a simple camera or the position and angle of the satellite that is taking the picture. Computational context is the information about the state of processing, requirements and information on both hardware and software.

In cognition, the context definition also takes into account the expectations of the observer [2]. Here the expectation can be constructed from the background knowledge that is learned from first hand experience or from a secondary source. The time frame can also vary, it can be an observation that has been made few seconds ago or a piece of knowledge acquired years ago.

2.2.3 Context in Linguistics

Context in the context of language can be defined by two different focuses. One is the focus on the linguistic expression itself and the second is the person who is using this expression, the utterance and the speaker. Although there is no clear distinction in their definitions, first, the expression, can be described as semantics, and the second is pragmatics [30].

In semantics, the context is the information that can be used for, for example, the disambiguation of indexicals. Information exists in utterance, the text that is been read or the entry in the dictionary as background knowledge defines the context. The word-sense disambiguation is also can be seen as the effect of context. However, a similar definition can be given for near-side pragmatics.

In pragmatics, context also includes the state of the parties that is in communication. The intension of the speaker also affects the context and defines the meaning of an utterance [18]. Both the speaker and the listener or the reader and the writer can be counted within the context that the meaning emerges from.

3 Quantum Probability

Giving a comprehensive story of quantum theory is way more than one computer major to handle. Following is summarized from the background information given in [7]. To give few prominent points in Quantum Physics's history, we can start with the quantization of energy by Max Planck. To unify the existing radiation laws, Planck introduced discrete energies. Another step was the interpretation of matter as a wave or as a particle. There was existing experimental evidence for the wave nature of light by Thomas Young, when Albert Einstein describe the photoelectric effect with the particle nature of photons. Later, Neils Bohr used the quantized discrete energy levels to explain electron levels of hydrogen atom.

Louis de Broglie used the Einstein's idea to view light as particles rather than waves and defined matter as waves rather than particles. In 1920s, Erwin Schrödinger put de Broglie's ideas into the wave function of matter known as Schrödinger equation. On the other hand, Werner Heisenberg developed a matrix formulation of quantum theory.

In 1932, John von Neumann proved the wave function and matrix formulation is mathematically equivalent. In 1930s, von Neumann, Francis Murray and Garrett Birkhoff axiomatized the quantum theory with a set of fundamental principles. With von Neumann's axiomatization, a non-commutative and non-distributive logic is defined that entails a generalized theory of probability.

3.1 Notation and Definitions from Linear Algebra

3.1.1 Vector notation

To explain quantum probability, there are a few definitions that are needed from linear algebra. But first, the notation will be introduced. The physicists use Dirac notation, named after Paul

Dirac, to represent quantum states. $|W\rangle$ is called a ket vector and can be represented as a column vector:

$$|W\rangle \mapsto \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}. \tag{1}$$

A bra vector $\langle R|$ can also be represented as a row vector:

$$\langle R| \mapsto \begin{bmatrix} r_1 & r_2 & \cdots & r_n \end{bmatrix}. \tag{2}$$

The notation does not imply a basis, so $|W\rangle$ is an abstract representation of a vector. However the "coordinates" of this vector depend on the basis that is used to define it, and there can be more than one representation for a given vector $|W\rangle$. A $n \times 1$ matrix represents the coordinates with a certain basis vector set[1]. For example, with the basis $\{|X\rangle, |Y\rangle, |Z\rangle\}$ that spans a 3D space, a vector $|S\rangle$ can be represented by $\mathbf{s}$ in Equation 3. The same vector $|S\rangle$ is $\mathbf{s}'$ with the basis $\{|K\rangle, |L\rangle, |M\rangle\}$ shown in Equation 4. Clearly $\mathbf{s} \neq \mathbf{s}'$; however, they still represent the same vector.

$$|S\rangle = s_1 |X\rangle + s_2 |Y\rangle + s_3 |Z\rangle, \tag{3}$$

$$|S\rangle = s'_1 |K\rangle + s'_2 |L\rangle + s'_3 |M\rangle, \tag{4}$$

$$|S\rangle \mapsto \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \mathbf{s}, \tag{5}$$

$$|S\rangle \mapsto \begin{bmatrix} s'_1 \\ s'_2 \\ s'_3 \end{bmatrix} = \mathbf{s}'. \tag{6}$$

Turning a bra vector to a ket vector is a linear function in the complex space:

$$|W\rangle \mapsto \langle W| = \begin{bmatrix} w_1^* & w_2^* & \cdots & w_n^* \end{bmatrix}. \tag{7}$$

Note that each component of the vector is complex-conjugated[2].

3.1.2 Inner Product

For n dimensional vectors $|W\rangle$, $|S\rangle$, inner product is shown with a braket and defined as follows:

$$\langle W|S\rangle = \sum_{i=1}^{n} w_i^* s_i. \tag{8}$$

3.1.3 Hilbert Space

Hilbert space is a generalization of Euclidean space. Named after David Hilbert, it is a possibly infinite dimensional, real or complex inner product space.

By definition, in a Hilbert space $\mathcal{H}$, for any $\lambda \in \mathbb{C}$, any pair of elements $\mathbf{w}, \mathbf{s} \in \mathcal{H}$, it also holds that $\mathbf{w} + \lambda\mathbf{s} \in \mathcal{H}$. The following properties have to hold for the inner product $\langle W|S\rangle$:

1. *Linearity:* Linear in the second argument and anti-linear in the first argument. Also called sesquilinear.

$$\begin{aligned} \langle W|c_1 S_1 + c_2 S_2\rangle &= c_1 \langle W|S_1\rangle + c_2 \langle W|S_2\rangle, \\ \langle c_1 W_1 + c_2 W_2|S\rangle &= c_1^* \langle W_1|S\rangle + c_2^* \langle W_1|S\rangle. \end{aligned} \tag{9}$$

[1]Basis is a set of vectors that are linearly independent and any vector in the space can be represented as a linear combination of these vectors.

[2]For a complex number $x = a + bi$, $x^* = a - bi$. x^*can also be shown as $\overline{x}$

2. *Conjugate symmetry:*

$$\langle W|S\rangle = \langle S|W\rangle^* . \tag{10}$$

3. *Positive-definiteness:*

$$\langle W|W\rangle = \sum_{i=1}^{n} w_i^* w_i = \sum_{i=1}^{n} |w_i|^2 \geq 0. \tag{11}$$

3.1.4 Quantum Probability

In Kolmogorovian or classical probability, events are represented as members of a set, and probabilities are given by a function that maps a set to a real number in range $[0,1]$. Quantum probability is defined on different axioms and in a way generalizes classical probability [26].

3.1.5 Events

In classical probability, a sample space contains finite number of points. Each point represents an elementary event that can be observed. A more complex event E_i is a subset of this sample space. Conjunction of two events E_1 and E_2 is represented by the set intersection $(E_1 \cap E_2)$, disjunction of two events is a set union $(E_1 \cup E_2)$. Conceptually there is no difference between E_1, E_2 or $E_1 \cap E_2$.

In quantum theory, an event is a subspace of a Hilbert space $\mathcal{X}$. A set of basis vectors $V = \{|V_i\rangle, i = 1, 2, .., n\}$ spans the space and each basis vector corresponds to an elementary outcome. For an event E_1 that is spanned by the $V_{E_1} \subseteq V$ and event E_2 spanned by $V_{E_2} \subseteq V$, two projectors are defined:

$$\mathbf{P}_{E_1} = \sum_{V_i \in V_{E_1}} |V_i\rangle \langle V_i| , \tag{12}$$

$$\mathbf{P}_{E_2} = \sum_{V_i \in V_{E_2}} |V_i\rangle \langle V_i| . \tag{13}$$

Conjunction of two events, $E_1 \wedge E_2$, is an intersection of subspaces, $V_{E_1} \cap V_{E_2}$, and disjunction of two events, $E_1 \vee E_2$, is span of subspaces, $V_{E_1} \cup V_{E_2}$.

3.1.6 System State

Classical probability theory defines a function that maps points of sample space to a number in interval $[0,1]$. The function defines the state of the system. The empty space is mapped to 0 and sample space is mapped to 1. For events E_1 and E_2, the probability of their occurrence can be defined as the relation $p(E_2 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$.

In quantum probability theory, state is a unit vector $|S\rangle$ in a Hilbert space. Probability of an event E_i with basis $V_{E_i} \subseteq V$ is $q(E_i) = \|\mathbf{P}_{E_i} |S\rangle\|^2$ where

$$\|\mathbf{P}_{E_i} |S\rangle\|^2 = \langle S| \mathbf{P}_{E_i}^{\dagger} \cdot \mathbf{P}_{E_i} |S\rangle = \langle S| \mathbf{P}_{E_i} |S\rangle . \tag{14}$$

For events E_1 and E_2, that are mutually exclusive, $E_1 \cap E_2 = \emptyset$, their probability is defined, from orthogonality, as follows:

$$\begin{aligned} q(E_1 \vee E_2) =& \|\mathbf{P}_{E_1} + \mathbf{P}_{E_2} |S\rangle\|^2 && (15)\\ =& \|\mathbf{P}_{E_1} |S\rangle\|^2 + \|\mathbf{P}_{E_2} |S\rangle\|^2 && (16)\\ =& q(E_1) + q(E_2). && (17) \end{aligned}$$

3.1.7 State Revision

Revision of the system state is the conditional probability function in classical probability. For the occurrence of an event E_i, the probability function is defined as:

$$p(X|E_i) = \frac{p(X \cap E_i)}{p(E_i)}. \tag{18}$$

In quantum probability, after the event E_i is observed the state $|S\rangle$ is updated with following relation:

$$|S_{E_i}\rangle = \frac{\mathbf{P}_{E_i}|S\rangle}{\|\mathbf{P}_{E_i}|S\rangle\|}. \tag{19}$$

3.1.8 Compatibility

In classical probability, a single probability function (also called the probability density function) defines the complete set of possibilities. All events and their combinations are defined and mapped by this probability function. This is called the principle of unicity by [15].

In quantum probability all events are contained within a single Hilbert space. However, events are of two kinds, compatible events and incompatible events.

If two events, E_1 and E_2, are spanned by subsets of same basis set, $V_{E_1} \subset V$ and $V_{E_2} \subset V$, then two events are compatible. As mentioned earlier in section 3.1.5, disjunction is union of basis vectors of the event and conjunction is intersection of those sets for compatible events.

If two events are defined over two different basis sets then, the events are incompatible in quantum probability, which is a condition that does not exist in classical probability. The two event have to be considered in some sequence. Suppose two basis sets are defined for the same Hilbert space $\mathcal{X}$, $V = \{|V_i\rangle, i = 1, 2, .., n\}$, $W = \{|W_i\rangle, i = 1, 2, .., n\}$. Event E_V is spanned by $V_{E_V} \subseteq V$ and event E_W is spanned by $W_{E_W} \subseteq W$. For the sequence of events, E_V occurring before E_W, probability of event E_V occurring is

$$q(E_V) = \|\mathbf{P}_{E_V}|S\rangle\|^2. \tag{20}$$

Then the state conditioned on observing event E_V is

$$|S'\rangle = \frac{\mathbf{P}_{E_V}|S\rangle}{\|\mathbf{P}_{E_V}|S\rangle\|}. \tag{21}$$

With the updated state vector $|S'\rangle$, the probability of event E_W is equals to

$$q(E_W|E_V) = \|\mathbf{P}_{E_W}|S'\rangle\|^2. \tag{22}$$

Then the probability of sequence of events E_V and E_W is

$$\begin{aligned} q(E_V) \cdot q(E_W|E_V) &= \|\mathbf{P}_{E_V}|S\rangle\|^2 \cdot \|\mathbf{P}_{E_W}|S'\rangle\|^2 \\ &= \|\mathbf{P}_{E_V}|S\rangle\|^2 \cdot \left\| \mathbf{P}_{E_W} \frac{\mathbf{P}_{E_V}|S\rangle}{\|\mathbf{P}_{E_V}|S\rangle\|} \right\|^2 \\ &= \|\mathbf{P}_{E_V}|S\rangle\|^2 \cdot \frac{1}{\|\mathbf{P}_{E_V}|S\rangle\|^2} \|\mathbf{P}_{E_W}\mathbf{P}_{E_V}|S\rangle\|^2 \\ &= \|\mathbf{P}_{E_W}\mathbf{P}_{E_V}|S\rangle\|^2. \end{aligned} \tag{23}$$

Since the projectors $\mathbf{P}_{E_V}$ and $\mathbf{P}_{E_W}$ do not share a common set of eigenvectors which implies $\mathbf{P}_{E_W}\mathbf{P}_{E_V} \neq \mathbf{P}_{E_V}\mathbf{P}_{E_W}$ then $q(E_V) \cdot q(E_W|E_V) \neq q(E_W) \cdot q(E_V|E_W)$.

In classical probability, $p(D|E_1, E_2, ...E_n)$ of event D does not depend on the order of events E_i. Because, it is based on sets, events E_1 through E_n is the set $E_1 \cap E_2 \cap \cdots \cap E_n$ as a single event. However, in quantum probability, event order is important as shown with equation 23.

3.1.9 Total Probability and Interference

Suppose two events, E_1 and E_2, are being observed and $q(E_1)$ is the main focus of the experiment. For the single event E_1 the probability can be calculated as $q(E_1) = \|\mathbf{P}_{E_1}|S\rangle\|^2$. If E_1 is observed after E_2, then the total probability for E_1 is:

$$q_T(E_1) = \|\mathbf{P}_{E_1}\mathbf{P}_{E_2}|S\rangle\|^2 + \|\mathbf{P}_{E_1}\mathbf{P}_{\overline{E_2}}|S\rangle\|^2, \tag{24}$$

where $\overline{E_2}$ is used to show the negation of event E_2. However, if the expression for the probability of the single event is manipulated as follows:

$$
\begin{aligned}
q(E_1) =& \| \mathbf{P}_{E_1} |S\rangle \|^2 \\
=& \| \mathbf{P}_{E_1} \mathbf{I} |S\rangle \|^2 \\
=& \| \mathbf{P}_{E_1} (\mathbf{P}_{E_2} + \mathbf{P}_{\overline{E_2}}) |S\rangle \|^2 \\
=& \langle S| (\mathbf{P}_{E_2} + \mathbf{P}_{\overline{E_2}}) \mathbf{P}_{E_1} \mathbf{P}_{E_1} (\mathbf{P}_{E_2} + \mathbf{P}_{\overline{E_2}}) |S\rangle \\
=& \langle S| (\mathbf{P}_{E_2} + \mathbf{P}_{\overline{E_2}}) \mathbf{P}_{E_1} (\mathbf{P}_{E_2} + \mathbf{P}_{\overline{E_2}}) |S\rangle \\
=& \langle S| \mathbf{P}_{E_2} \mathbf{P}_{E_1} \mathbf{P}_{E_2} |S\rangle + \langle S| \mathbf{P}_{E_2} \mathbf{P}_{E_1} \mathbf{P}_{\overline{E_2}} |S\rangle \\
&+ \langle S| \mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_1} \mathbf{P}_{E_2} |S\rangle + \langle S| \mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_1} \mathbf{P}_{\overline{E_2}} |S\rangle \\
=& \| \mathbf{P}_{E_1} \mathbf{P}_{E_2} |S\rangle \|^2 + \| \mathbf{P}_{E_1} \mathbf{P}_{\overline{E_2}} |S\rangle \|^2 \\
&+ \langle S| \mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_1} \mathbf{P}_{E_2} |S\rangle + \langle S| \mathbf{P}_{E_2} \mathbf{P}_{E_1} \mathbf{P}_{\overline{E_2}} |S\rangle \\
=& q_T(E_1) + \langle S| \mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_1} \mathbf{P}_{E_2} |S\rangle + \langle S| \mathbf{P}_{E_2} \mathbf{P}_{E_1} \mathbf{P}_{\overline{E_2}} |S\rangle
\end{aligned}
\tag{25}
$$

It can be shown that the following equality holds:

$$q_T(E_1) = q_T(E_1) + int_{E_1}, \tag{26}$$

where $int_{E_1} = \langle S| \mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_1} \mathbf{P}_{E_2} |S\rangle + \langle S| \mathbf{P}_{E_2} \mathbf{P}_{E_1} \mathbf{P}_{\overline{E_2}} |S\rangle$ represents the interference. If E_1 and E_2 are compatible, ie. $\mathbf{P}_{E_1} \mathbf{P}_{E_2} = \mathbf{P}_{E_2} \mathbf{P}_{E_1}$ then, $int_{E_1} = 0$ and $q_T(E_1) = q_T(E_1)$. Because,

$$\mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_1} \mathbf{P}_{E_2} = \mathbf{P}_{\overline{E_2}} \mathbf{P}_{E_2} \mathbf{P}_{E_1} = 0 \cdot \mathbf{P}_{E_1} = 0. \tag{27}$$

However, int_{E_1} can have a positive or negative value and violate the total probability.

4 Semantic Space

One of the ways to describe the meaning of a word is defining it as a point in a multidimensional space, called semantic space [10]. This space can be constructed from the co-occurrence of words within a corpus. It is observed that the meaning of a word can be represented by the distribution of words around it in a corpus. This representation can be simply the frequencies of the word occurrences [20]. The occurrences also can be used to generate word embeddings [24, 22]. Also apart from co occurrences syntactic dependency can be used to generate a semantic space for words [23].

A word w has co-occurrence relations with other words in the corpus $C = \{D_1, D_2...D_m\}$ where m is the number of documents, $V = \{w_1, w_2, ...w_n\}$ where n is number of word in the vocabulary. Co-occurrence can be calculated around a word within a window in each document D_i. This way certain number of words before and after the target word represents the context of the word. Dependency relations can be represented by a similar vector with same size. This dependency based vector can represent modifiers and affordances of a word, verbs and adjectives, used in relation to the target word. More complicated definition of context while generating word embeddings are shown to give marginal improvements [1].

5 Proposal

The quantum probability that is developed and mainly used by physicists has been proposed to be a good candidate for the modeling context [7]. There is evidence from cognitive science that the framework has better descriptive qualities for explaining cognitive science processes. Moreover, studies in information retrieval have demonstrated its practical applications, such as summarization of documents [25] and ranking [29].

5.1 Quantum Probabilistic Nature of Decision Making

Observations from a decision experiments will be explained to make a case that the quantum probability framework is more suitable to model decision processes of humans.

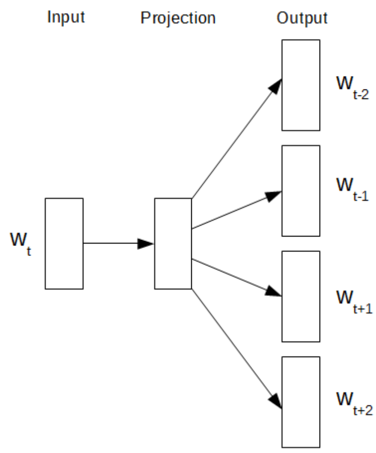

Figure 1: Word embedding using context. A naive representation of a word is projected to its representation, representation is mapped to its surrounding words.

5.1.1 The Conjunction Fallacy

The conjunction fallacy is a judgment error first described by Tversky and Kahneman in [33]. In the experiments which this judgment fallacy is observed, participants are given a text about the previous life of a woman, named Linda, and asked questions about her. The questions are: Linda is now:

1. active in the feminist movement,
2. a bank teller,
3. active in the feminist movement and a bank teller,
4. active in the feminist movement or a bank teller.

The conjunction fallacy occurs when option 3 is judged to be more likely than option 2. In classical interpretation, both options are sets and the option 3 is intersection of options 1 and 2. So the probability of option 3 has to be smaller than or equal to option 2. But the participants frequently judge option 3 more likely than option 2.

$$q(S_3) = \|\mathbf{P}_F \mathbf{P}_B \left|S\right\rangle\|^2 \tag{28}$$

$$q(S_1) = \|\mathbf{P}_B \left|S\right\rangle\|^2 \tag{29}$$

In quantum probability, option 3 can be weighed by $q(S_3)$ and option 1 can be weighed by $q(S_1)$. The value difference can be made to match the experiment results with careful selection of state vector values and the basis chosen for being a feminist $|F\rangle$ and being a bank teller $|B\rangle$. The figure 2 is a geometric representation that shows the difference between $|S_1\rangle$ and $|S_3\rangle$. As can be seen, quantum framework can model the behavior observed in the experiments. The two choices are represented by two different basis in two dimensional space. The non-compatibility of this two basis give quantum probability the ability to model the observation of the experiments done.

Motivated from these, this report proposes to study context from a quantum probability, as outlined below.

5.2 Modeling Context within Quantum Probability

Context can be modeled in two ways within quantum probability framework. First one is representing each context as a different basis set. Second one is representing context as subspaces of the same Hilbert space.

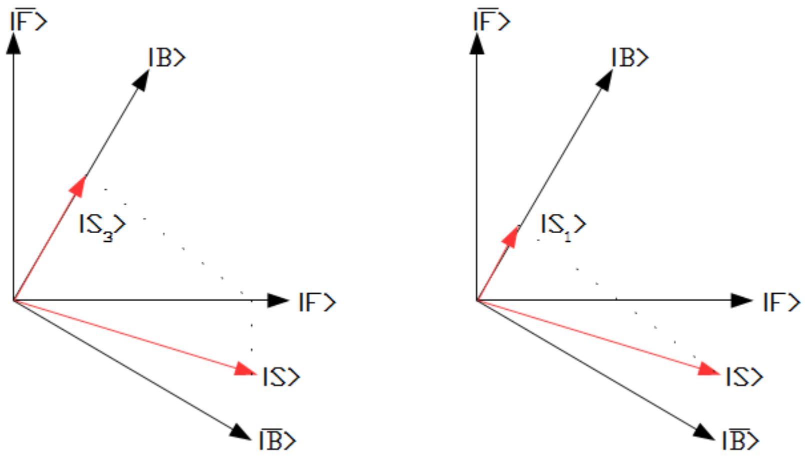

Figure 2: Same vector different basis sets

5.2.1 Context as a Basis

In the basis approach, each context C_i is a different set of basis vector for the Hilbert space $\mathcal{X}$ that the word w_i exist in. For contexts $C_1 = \{W_1^1, W_2^1, ..., W_n^1\}$ and $C_2 = \{W_1^2, W_2^2, ..., W_n^2\}$, the word $|w\rangle$ has following relations,

$$|w\rangle = a_1 \left|W_1^1\right\rangle + a_2 \left|W_2^1\right\rangle + \ldots + a_n \left|W_n^1\right\rangle, \tag{30}$$

$$|w\rangle = b_1 \left|W_1^2\right\rangle + b_2 \left|W_2^2\right\rangle + \ldots + b_n \left|W_n^2\right\rangle, \tag{31}$$

where a_i and b_i for $i \in \{1, 2, \ldots, n\}$ is associations of word $|w\rangle$ with word W_i. For example, if word vectors comes from co-occurrences, different context modifies the co-occurrence frequencies thus their meaning.

5.2.2 Context as a Subspace

After a corpus is processed with the aim to produce the mentioned frequencies, A word can be represented with a vector $w_i = (w_{i1}, w_{i2}, ..., w_{in})^T$. With basis vectors that represent each word in the vocabulary, that are not necessarily orthogonal, a vector that represent the “meaning” can be shown as $|w_i\rangle$. The density matrix can be calculated as $\rho_i = |w_i\rangle \langle w_i|$. Word w representing a concept

$$\rho_w = p_1\rho_1 + p_2\rho_2 +, \ldots, p_m\rho_m, \tag{32}$$

where there is m senses, or associations for the word w.

The effect of context on a word w can be defined as a quantum measurement, that projects a word state ρ_w into a subspace [7]. This projection is defined as:

$$\mathbf{P}_x \rho_w \mathbf{P}_x^\dagger = |w\rangle \langle x| \rho_w |x\rangle \langle w| = \rho_w^x, \tag{33}$$

where $\mathbf{P}_x = |w\rangle \langle x|$ is the operator representing the context x, which is not necessarily a projection operator. The probability p of this projection (collapse of state) is given by:

$$p = |x\rangle \rho_w \langle w|\,. \tag{34}$$

For a pure state vector $|w\rangle$, that represents the word w, the new state $|w'\rangle$ is given by:

$$|w'\rangle = \frac{\mathbf{P}_x |w\rangle}{\|\mathbf{P}_x |w\rangle\|}. \tag{35}$$

5.3 Modeling Word Embeddings with Quantum Probability

The probability of co-occurrence between two words can be modeled within a quantum probability framework. For a word w with embedding $|W\rangle$, its context can be defined as a set of words that are within a certain distance around the word w, $C = \{w_1, w_2, w_3, ..., w_t\}$. The probability of word w_i occurring within the context window of w, $q(w_i|w)$ is equal to:

$$q(w_i|w) = \|\mathbf{P}_{W_i} |W\rangle \|^2, \tag{36}$$

where $\mathbf{P}_{W_i} = |W_i\rangle \langle W_i|$.

With the probabilities calculated over a corpus, finding the vector representation of a given word becomes the following minimization problem.

$$\underset{|W_i\rangle, |W\rangle}{\arg\min}\, C(w, w_i) = \|\mathbf{P}_{W_i} |W\rangle \|^2 - P(w_i|w), \tag{37}$$

where the value $P(w_i|w)$ is the classical probability the word w_i occurring within the context window of word w, that is calculated over the corpus.

6 Conclusion and Future Work

In this report, we provide a semantic-space take on modeling context. For this end, we have started training word2vec embeddings that can be used as a semantic map for the approach provided in the report. However, the model requires unit length word representations, for which we wrote our own loss function for training a word2vec network. Our initial results suggest that a useful representation has been learned in this manner.

Acknowledgments

For the experiments, we acknowledge the use of the facilities provided by the the Modeling and Simulation Center of METU (MODSIMMER). This work is funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK) through project no 215E133.

References

[1] Arora, S., Li, Y., Liang, Y., Ma, T., and Risteski, A. (2015). Random Walks on Context Spaces: Towards an Explanation of the Mysteries of Semantic Word Embeddings. *arXiv preprint arXiv:1502.03520*, (February):1–23.

[2] Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5:617–629.

[3] Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, 22(4):577–660.

[4] Barsalou, L. (2008). Situating concepts. In P. Robbins, M. A., editor, *Cambridge handbook of situated cognition*, pages 236–263. New York: Cambridge University Press.

[5] Bazire, M. and Brézillon, P. (2005). Understanding context before using it. *Modeling and using context*, 3554:29–40.

[6] Borghi, A. M. (2007). Object concepts and embodiment: Why sensorimotor and cognitive processes cannot be separated. *La nuova critica*, 15(4):447–472.

[7] Busemeyer, J. and Bruza, P. (2012). *Quantum Models of Cognition and Decision.* Quantum Models of Cognition and Decision. Cambridge University Press.

[8] Celikkanat, H., Orhan, G., and Kalkan, S. (2015). A probabilistic concept web on a humanoid robot. *IEEE Transactions on Autonomous Mental Development*, 7(2):92–106.

[9] Celikkanat, H., Orhan, G., Pugeault, N., Guerin, F., Şahin, E., and Kalkan, S. (2016). Learning context on a humanoid robot using incremental latent dirichlet allocation. *IEEE Transactions on Cognitive and Developmental Systems*, 8(1):42–59.

[10] Firth, J. (1957). *A Synopsis of Linguistic Theory, 1930-1955.*

[11] Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine learning*, 2(2):139–172.

[12] Gabora, L., Rosch, E., and Aerts, D. (2008). Toward an ecological theory of concepts. *Ecological Psychology*, 20(1):84–116.

[13] Gennari, J. H., Langley, P., and Fisher, D. (1989). Models of incremental concept formation. *Artificial intelligence*, 40(1-3):11–61.

[14] Ghidini, C. and Giunchiglia, F. (2001). Local Models Semantics, or contextual reasoning = locality + compatibility. *Artificial Intelligence*, 127(2):221–259.

[15] Griffiths, R. B. (2014). The New Quantum Logic. *Foundations of Physics*, 44(6):610–640.

[16] Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. *Handbook of categorization in cognitive science*, pages 20–45.

[17] Kalkan, S., Yuruten, O., and Sahin, E. (2013). Relating affordances with verbs, nouns and adjectives. In *21st Signal Processing and Communications Applications Conference (SIU).*

[18] Korta, K. and Perry, J. (2015). Pragmatics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2015 edition.

[19] Lebowitz, M. (1987). Experiments with incremental concept formation: Unimem. *Machine learning*, 2(2):103–138.

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

[21] McCarthy, J. (1993). Notes on formalizing context. In *Proceedings of the 13th International Joint Conference on Artifical Intelligence - Volume 1*, IJCAI'93, pages 555–560, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[22] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Nips*, pages 1–9.

[23] Padó, S. and Lapata, M. (2007). Dependency-Based Construction of Semantic Space Models. *Computational Linguistics*, 33(2):161–199.

[24] Pennington, J., Socher, R., and Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543.

[25] Piwowarski, B. (2012). On using a quantum physics formalism for multidocument summarization. *Journal of the American ....*

[26] Rau, J. (2009). On quantum vs. classical probability. *Annals of Physics*, 324(12):2622–2637.

[27] Rosch, E. (1973). Natural categories. *Cognitive psychology*, 4(3):328–350.

[28] Serafini, L. and Bouquet, P. (2004). Comparing formal theories of context in AI. *Artificial Intelligence*, 155(1):41–67.

[29] Sordoni, A., Nie, J.-Y., and Bengio, Y. (2013). Modeling term dependencies with quantum language models for IR. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13*, page 653, New York, New York, USA. ACM Press.

[30] Speaks, J. (2016). Theories of meaning. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2016 edition.

[31] Strat, T. (1993). Employing Contextual Information in Computer Vision. *Proc. DARPA93*.

[32] Turner, J. C., Oakes, P. J., Haslam, S. A., and McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and social psychology bulletin*, 20:454–454.

[33] Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4):293.

[34] Yeh, W. and Barsalou, L. W. (2006). The situated nature of concepts. *The American journal of psychology*, pages 349–384.