A Quantum-Inspired Matching Network with Linguistic Theories for Metaphor Detection

Wenbo Qiao[1], Peng Zhang[2]*, Zenglai Ma[1]
[1]School of New Media and Communication, Tianjin University, Tianjin, China
[2]College of Intelligence and Computing, Tianjin University, Tianjin, China
{qiaowb, pzhang, zenglaima}@tju.edu.cn

Abstract

Enabling machines with the capability to recognize and comprehend metaphors is a crucial step toward achieving artificial intelligence. In linguistic theories, metaphor can be identified through Metaphor Identification Procedure (MIP) or Selectional Preference Violation (SPV), both of which are typically considered as matching tasks in the field of natural language processing. However, the implementation of MIP poses a challenge due to the semantic uncertainty and ambiguity of literal meanings of words. Simultaneously, SPV often struggles to recognize conventional metaphors. Inspired by Quantum Language Model (QLM) for modeling semantic uncertainty and fine-grained feature matching, we propose a quantum-inspired matching network for metaphor detection. Specifically, we use the density matrix to explicitly characterize the literal meanings of the target word for MIP, in order to model the uncertainty and ambiguity of the literal meanings of words. This can make SPV effective even in the face of conventional metaphors. MIP and SPV are then achieved by fine-grained feature matching. The results of the experiment finally demonstrated our approach has strong competitiveness.

Keywords: metaphor detection, linguistic theories, quantum-inspired language models

1. Introduction

Metaphors are widely present in the language, thought and behavior of humans, serving as a unique means for cognition and understanding of unknown and abstract concepts. The Conceptual Metaphor Theory (CMT) (Lakoff and Johnson, 1980) argues that metaphor is essentially a mapping from a source domain concept to a target domain concept, which is a kind of analogical reasoning ability that reflects human cognitive processes. For example, the sentence *"words were music to my ears"* establishes a connection between *"music"* and *"words"*. This description is more vivid, making it easier for readers to understand the author's emotions and attitudes. Currently, detection and understanding of metaphors have become crucial tasks in natural language processing, and understanding metaphors can assist in improving the performance of related tasks such as machine translation (Mao et al., 2018), sentiment analysis (Dankers et al., 2019; Mao and Li, 2021), and even depression detection (Han et al., 2022).

Although it is difficult for machines to fundamentally understand and recognize metaphors, linguistic theory offers a middle way that treats metaphor detection as a matching problem. Selectional Preference Violation (SPV), for instance, assesses metaphors by comparing the incongruity between the target word and its context (Wilks, 1975). As shown in Figure 1, the significant incongruity between the target word and its context is likely to indicate a metaphor. However, this method often fails when dealing with conventional metaphors in which the target word and its context are created widely recognized fixed collocations (Maudslay and Teufel, 2022; Zhang and Liu, 2022). In other words, this approach fails to identify metaphorical expressions commonly used. Therefore, many studies (Mao et al., 2019; Su et al., 2021; Choi et al., 2021; Zhang and Liu, 2022) have combined Metaphor Identification Procedure (MIP) to deal with conventional metaphors. MIP method identifies metaphors by comparing the similarity between the meaning of the target word in the context and the literal meaning of the target word (Group, 2007; Steen et al., 2010). As shown in Figure 1, the two meanings are inconsistent and the target word will be considered as a metaphor. However, characterizing the literal meaning of the target word is often challenging due to the uncertainty of natural language phenomena such as polysemy (Zhang and Liu, 2022).

In this paper, we attempt to tackle the aforementioned challenges using Quantum Language Model (QLM). It's a semantic computing framework that draws inspiration from mathematical principles from quantum mechanics (Sordoni et al., 2013). It explicitly models the uncertain phenomena present in human language, such as polysemy (Meyer and Lewis, 2020; Zhang et al., 2022), and enables fine-grained semantic matching for information retrieval (Sordoni et al., 2013; Jiang et al., 2020) or question answering (Zhang et al., 2018; Chen et al., 2021). Recently, it has also demonstrated advantages in tasks related to the understanding of human

* Corresponding author

LREC-COLING 2024, pages 1435–1445

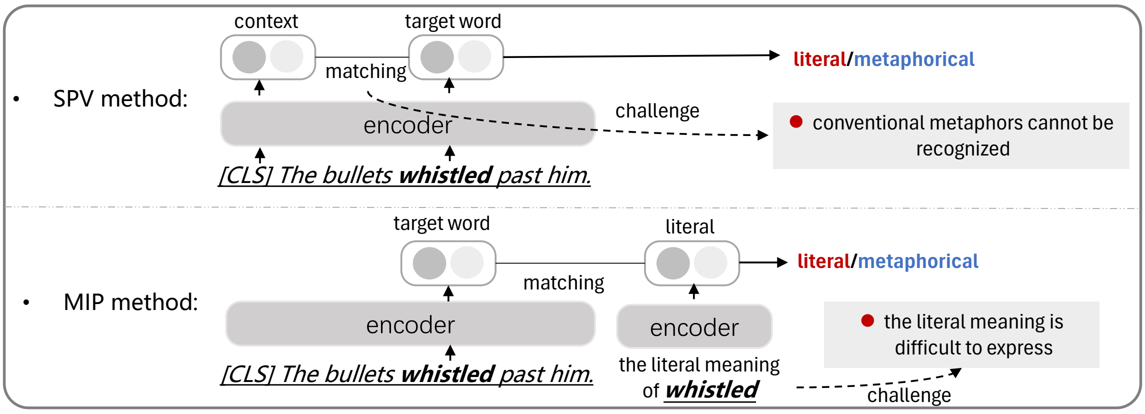

Figure 1: MIP and SPV methods are implemented respectively in related work, along with the existing challenges associated with these methods.

cognition, such as emotion recognition (Li et al., 2021a; Gkoumas et al., 2021) and sarcasm detection (Zhang et al., 2021; Liu et al., 2021). We note that metaphor detection (MIP or SPV methods) are also matching tasks, and their deeper semantic matching mechanisms align with the application scope of QLM. Furthermore, QLM can handle semantic uncertainties like polysemy. It may provide an elegant solution to the challenge of modeling the literal meaning of target words in MIP method.

Therefore, we propose a **Q**uantum-inspired **M**atching Network with Linguistic Theories for **M**etaphor Detection (QMM). Specifically, we encode each word as a quantum state representation and model semantic combinations by the density matrix. Then, we implement MIP and SPV by computing the similarity between the density matrices. QMM also encapsulates the polysemy information of the target word in a density matrix by introducing external knowledge and provides a solution to the problem of modeling the literal meaning of the target word in MIP method. The contributions of this paper can be summarized as follows:

- We design QLM based on linguistic theorie, which is the first work to achieve metaphor detection through QLM.
- We solve the problem of difficult modelling of literal meanings. The literal meaning of target words is modeled from a novel perspective in a more natural and reasonable way.
- We achieved highly competitive experimental results on multiple datasets. The code is available in the repository[1].

2. Related Work

In recent work, metaphor detection can be categorized into various task forms, including approaches based on sequence labeling, multi-task learning, classification and matching.

Sequence Labeling-based Methods. The methods primarily explore metaphorical information within the sequential relationships. These works mostly use recurrent neural networks to extract the sequential relationships of words and train models leveraging existing metaphor labels (Wu et al., 2018; Gong et al., 2020; Li et al., 2021b). Considering the lack of interpretability of simply using recurrent neural networks, Mao et al. (2019) implement metaphor detection theory MIP and SPV into sequence labeling models. However, these approaches rely on shallow neural networks and their performance still requires further enhancement.

Multi-task Learning-based Methods. The methods primarily investigate the relationships between metaphor detection and other tasks. For instance, Le et al. (2020) studied the relationship between semantic disambiguation and metaphor detection and discovered that these two tasks can mutually benefit from knowledge sharing. There is also a close relationship between sentiment recognition and metaphor detection (Dankers et al., 2019; Mao and Li, 2021). Nevertheless, these works often require the design of complex multi-task learning networks, and the establishment of the relevant datasets is also often challenging.

Context-based Classification Methods. The methods primarily study the relationship between different semantic objects in the context (Mu et al., 2019; Zayed et al., 2020). For example, Su et al. (2020) takes advantage of contextual features by treating metaphor detection as a reading comprehension problem. Some studies on metaphor detection have further investigated the specific grammatical relationships (Dankers et al., 2020) and the impact of verb-object relationships (Song et al., 2021) in context. These works heavily rely on labeled data and are not inherently dedicated to metaphor detection.

[1] https://github.com/QuaRobot/QMM

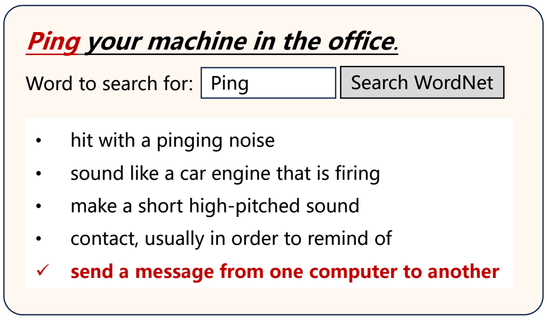

Figure 2: The gloss of the word *"ping"* in WordNet. If the first gloss is simply taken as the literal meaning of *"ping"* (*"ping"* actually means in context is what the last gloss described), MIP method may incorrectly judge the target word as metaphorical. Because the meaning expressed by the first gloss is clearly not similar to its meaning in context.

Matching Methods Based on Linguistic Theories. The methods primarily study the relationship between target words and context based on linguistic theories SPV or the relationship between the meaning of target words in context and their literal meanings based on linguistic theories MIP. Some works simply assume that the literal meaning of the target word is obtained through pre-training word vectors (Mao et al., 2019; Choi et al., 2021), while others contend that the first gloss (the interpretation or definition of a word) in the dictionary for the target word represents its literal meaning (Zhang and Liu, 2022). These works are based on naive assumptions, leading to potentially arbitrary judgments made by models, as shown in Figure 2. Some researchers have attempted to utilize concatenate operations (Su et al., 2021) or weight aggregation operations (Wan et al., 2021) to fuse multiple glosses information. Obviously, the simple fusion methods may introduce more noise, resulting in limited or even decreased performance.

3. Preliminaries

The section will briefly introduce the basic knowledge of quantum theory, then provide the implementation ideas of QLM in the relevant literature.

3.1. Quantum Preliminary

The mathematical foundation of quantum theory is based on Hilbert space, which is typically a complex vector space $\mathcal{H} \in \mathbb{C}^d$. Quantum states are abstracted as column vectors (e.g., $\boldsymbol{u}$) on this space, represented by Dirac notation as a ket $|u\rangle$. Its conjugate transpose $\boldsymbol{u}^\dagger$ is represented as a bra $\langle u|$. Given any two quantum states $|u\rangle$ and $|v\rangle$, the inner product and outer product can be defined as $\langle u|v\rangle = \boldsymbol{u}^\dagger \boldsymbol{v}$ and $|u\rangle\langle v| = \boldsymbol{u}\boldsymbol{v}^\dagger$, respectively. Mathematically, quantum states be represented as a linear combination of a set of basis vectors: $|u\rangle = \sum_{i=1}^{d} c_i |\boldsymbol{e_i}\rangle$, where $\{|\boldsymbol{e_i}\rangle\}_{i=1}^{d}$ present a set of basis vectors of $\mathcal{H}$, corresponding to a set of basis states and $\{c_i\}_{i=1}^{d}$ refer to a set of coefficients known as probability amplitudes, which must satisfy the normalization condition $\sum_{i=1}^{d} |c_i|^2 = 1$.

A quantum state describes a quantum system in a pure state. While more complex mixed quantum systems can be formulated by density matrices $\rho = \sum_{i=1}^{n} p_i |u_i\rangle\langle u_i|$, while $\{p_i\}_{i=1}^{n}$ represent a classical probability distribution that describe the probability of each quantum state $|u_i\rangle$ appearing. Finally, we can employ Gleason's theorem (Gleason, 1975) in quantum measurement to derive the probability of an observation from a quantum system. Additionally, von Neumann entropy (Umegaki, 1954) can be utilized to quantify the correlation between two quantum systems.

3.2. Quantum Language Model

QLM are more generalized models for modeling the representation, composition, and matching of natural language, which can capture the uncertainty present in natural language (Liu et al., 2021).

Semantic Representation. In QLM, the semantic space of natural language is viewed as a Hilbert space $\mathcal{H}$ spanned by a set of orthonormal basis vectors $\{|\boldsymbol{e_i}\rangle\}_{i=1}^{d}$ (e.g., a set of one-hot vectors). $\{|\boldsymbol{e_i}\rangle\}_{i=1}^{d}$ be interpreted as a set of sememes, while a word w represents a superposition of sememes: $|w\rangle = \sum_{i=1}^{d} c_i |\boldsymbol{e_i}\rangle$, where $\sum_{i=1}^{d} |c_i|^2 = 1$. In the probability amplitude $c_i = r_i \mathtt{e}^{\mathtt{i}\phi_i}$, $\mathtt{e}$ is natural logarithm, $\mathtt{i}$ is the imaginary number with $\mathtt{i}^2 = -1$. $\{r_i\}_{i=1}^{d}$ are real scalar values that describe the probability distribution of the word w with sememes combinations. $\phi \in [-\pi, +\pi]$ is the phase, which may contain some important information such as emotions, sarcasm, metaphors, or positional information of language sequences (Zhang et al., 2022; Gkoumas et al., 2021).

Semantic Composition. More complex semantic combinational relationships, such as paragraphs, n-grams, and documents, can be modeled using density matrices. For example, given a paragraph $\{w_1, w_2, ...w_n\}$, its semantic combination can be represented as: $\rho_d = \sum_{i=1}^{n} p_i |w_i\rangle\langle w_i|$, where $|w_i\rangle$ is the semantic representation of the word w_i. $\{p_i\}_{i=1}^{n}$ represent the weight information of each word, corresponding to a classical probability distribution (i.e., $\sum_{i=1}^{n} p_i = 1$). It is important to note that the diagonal elements in the density matrix represent classical semantic probability distributions, while the off-diagonal elements describe correlations between different sememes, which are usually assumed to be independent in classical methods (Li et al., 2019). In addition, the density matrix is a tool for modeling uncertainty and fuzziness (Meyer and Lewis, 2020). This will help us to model the literal meanings of the words elegantly.

Semantic Matching. The similarity between two semantic objects u and v in QLM can be measured by either the quantum relative entropy (i.e., $S(\rho_u, \rho_v) = -tr(\rho_u log \rho_v)$) (Sordoni et al., 2013) or the trace inner product (i.e., $S(\rho_u, \rho_v) = tr(\rho_u \rho_v)$) (Van Rijsbergen, 2004) of their corresponding density matrices. The trace inner product can be encoded into the joint matrix $\rho_{uv} = \rho_u \rho_v$, where the diagonal elements of the joint matrix correspond to semantic overlap, while the off-diagonal elements encode semantic interactions between different sememes. Therefore, some QLMs enhance the semantic matching effect by using convolutional neural network to extract off-diagonal elements (Zhang et al., 2018, 2022).

4. Proposed Model

In linguistic theories, metaphor detection has been reduced to a matching task. The task can be formalized as follows: given textual data $D = [w_1, w_2, ..., w_t, ..., w_n]$, determine whether the target word w_t is a metaphor. For MIP method, it is necessary to compute the similarity between the target word w_t and its literal meaning w_l, while SPV method computes the similarity between the target word w_t and its context $C = [w_1, w_2, ..., w_n]$. We propose a metaphor detection framework QMM that combines MIP and SPV as shown in Figure 3.

4.1. Semantic Representation and Composition

4.1.1. Contextual Density Matrix

We use BERT to obtain the encoding representation of the context $C = [w_1, w_2, ..., w_n]$:

$$[\boldsymbol{r}_1^c, \boldsymbol{r}_2^c, ..., \boldsymbol{r}_n^c] = BERT([w_1, w_2, ..., w_n]). \quad (1)$$

Here, $\boldsymbol{r}_i^c \in \mathbb{R}^d$ represents the column vector of the context word w_i.

The positional relationship of the target word relative to other words is an important feature for metaphor detection. We assign the target word the index number 1. Subsequently, the index numbers of the words to the left and right of the target word gradually increase. Then, QMM encodes the index of w_i through an encoding layer $f_{index}^c(.)$ with randomly initialized values:

$$[\boldsymbol{\phi}_1^c, \boldsymbol{\phi}_2^c, .., \boldsymbol{\phi}_n^c] = f_{index}^c([w_1, w_2, ..., w_n]), \quad (2)$$

where $\boldsymbol{\phi}_i^c \in \mathbb{R}^d$. This is similar to the global and local features considered in works (Su et al., 2020; Choi et al., 2021; Zhang and Liu, 2022), but the method used by QMM is clearly fine-grained and flexible.

Eq.1 corresponds to the real part, and Eq.2 corresponds to the imaginary part of the semantic representation. Thus the word w_i is denoted by the complex-valued vector:

$$\begin{aligned}\boldsymbol{w_i} &= [r_{i,1}^c + \mathtt{i}\phi_{i,1}^c, r_{i,2}^c + \mathtt{i}\phi_{i,2}^c, ..., r_{i,d}^c + \mathtt{i}\phi_{i,d}^c]^T \\ &= \boldsymbol{r}_i^c + \mathtt{i}\boldsymbol{\phi}_i^c.\end{aligned} \quad (3)$$

By Eq.4 we can obtain a normalized complex-valued vector:

$$|w_i\rangle = \frac{\boldsymbol{w_i}}{||\boldsymbol{w_i}||}, \quad (4)$$

where $|w_i\rangle \in \mathbb{C}^d$ represents the quantum-like representation of the word w_i.

Finally, QMM utilizes density matrices for semantic composition:

$$\rho_c = \sum_{i=1}^{n} p_i \, |w_i\rangle \, \langle w_i|, \quad (5)$$

where $\rho_c \in \mathbb{C}^{d \times d}$. We obtain weights $\{p_i\}_{i=1}^n$ by $[p_1, p_2, ..., p_n] = \sigma(f([\boldsymbol{r}_1^c, \boldsymbol{r}_2^c, ..., \boldsymbol{r}_n^c]))$, where the $f(.)$ denotes a two-layer perceptron, the $\sigma(.)$ denotes Softmax activation function. The calculation process of this weight is similar to an attention mechanism. The contextual density matrix fully models the semantic combination and achieves a fine-grained feature interaction in a way often overlooked by classical approaches.

4.1.2. Target Word Density Matrix

For the target word, we adopt a similar approach. We obtain the real part representation of the target word w_t according to:

$$\boldsymbol{r}^t = BERT(w_t | [w_1, w_2, ..., w_n]), \quad (6)$$

and get the imaginary part representation by:

$$\boldsymbol{\phi}^t = f_{pos}^t(w_t | [w_1, w_2, ..., w_n]), \quad (7)$$

where $\boldsymbol{r}^t, \boldsymbol{\phi}^t \in \mathbb{R}^d$. $f_{pos}^t(.)$ is a randomly initialized encoder, used for coding Part of Speech (POS) of target words. A great deal of previous work has shown POS of the target word plays a very important role in metaphor detection (Su et al., 2020; Zhang and Liu, 2022). Similar to Eq.3 and Eq.4, we also obtain the semantic quantum representation $|w_t\rangle$ of the target word and a pure-state density matrix $\rho_t \in \mathbb{C}^{d \times d}$:

$$\rho_t = |w_t\rangle \, \langle w_t| \, . \quad (8)$$

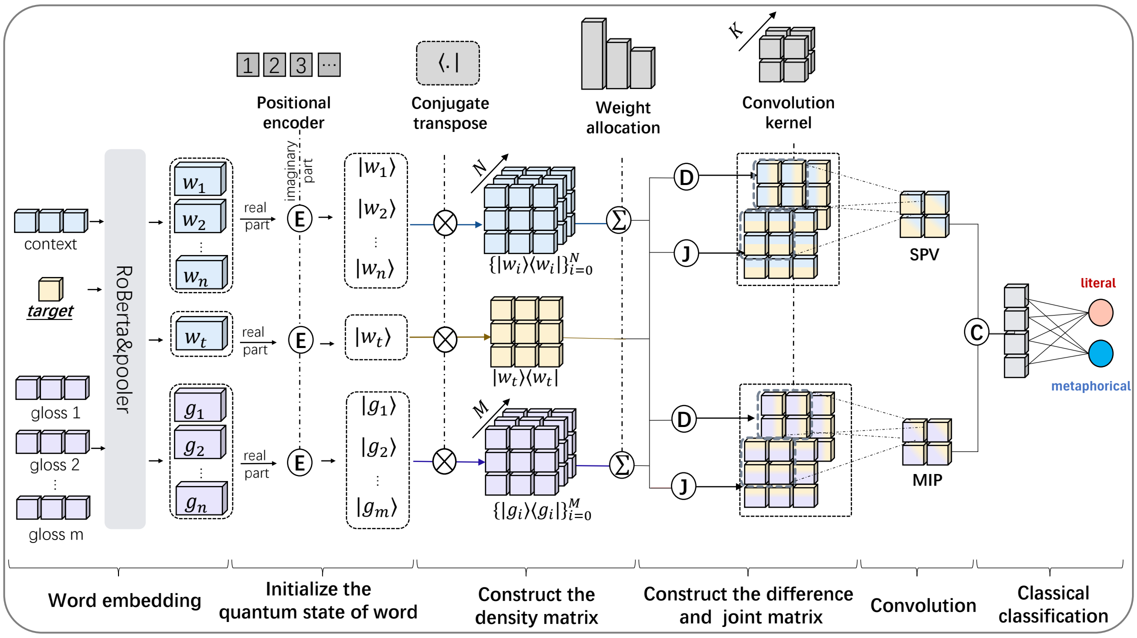

Figure 3: The framework of QMM. Here, **E** represents Eq.3-4 and indicates the initialization of a quantum state. $\bigotimes$ represents the outer product, $\sum$ represents Eq.5/ Eq.11, while **J** and **D** represent Eq.12 and 13, respectively. **C** represents feature unfolding and connecting operations.

4.1.3. Literal Semantic Density Matrix

QMM introduces the glosses $G_t = [g_1, g_2, ..., g_m]$ from external knowledge (e.g., the dictionaries or WordNet), where each gloss is also a sequence of words (i.e., $g_i = [w_1, w_2, ..., w_n]$). Using BERT and a pooling layer, we obtain the global encoding representation of m glosses of the word w_t:

$$[\boldsymbol{r}_1^g, \boldsymbol{r}_2^g, .., \boldsymbol{r}_m^g] = BERT([g_1, g_2, ..., g_m]|w_t). \quad (9)$$

Here, $\boldsymbol{r}_i^g \in \mathbb{R}^d$ will be represented as the real part of gloss g_i.

The ordering relationship of glosses in a dictionary is an important prior knowledge, reflecting the frequency with which humans use different meanings of a word (Miller, 1995). Intuitively, the more forward the position, the more likely it is to be the literal meaning of the target word. Therefore, we index each gloss in the order of its appearance. Then, we encode the index of g_i through an encoding layer $f_{index}^g(.)$:

$$[\boldsymbol{\phi}_1^g, \boldsymbol{\phi}_2^g, .., \boldsymbol{\phi}_m^g] = f_{index}^g([g_1, g_2, ..., g_m]|w_t), \quad (10)$$

where $\boldsymbol{\phi}_i^g \in \mathbb{R}^d$ will be represented as the imaginary part of gloss g_i.

Similar to Eq.3 and Eq.4, we can obtain the semantic quantum representation $|g_i\rangle \in \mathbb{C}^d$. Consequently, we have a literal semantic density matrix $\rho_l \in \mathbb{C}^{d\times d}$:

$$\rho_l = \sum_{i=1}^{m} p_i \, |g_i\rangle \, \langle g_i|, \quad (11)$$

where the weight p_i is assigned by calculating the similarity between the context and each gloss: $[p_1, p_2, ..., p_m] = \sigma([\boldsymbol{r}_1^g, \boldsymbol{r}_2^g, .., \boldsymbol{r}_m^g]^T \boldsymbol{r}_1^c)$. $\boldsymbol{r}_1^c$ is the first word $[CLS]$ in the context. $\sigma(.)$ denotes Softmax activation function. Therefore, QMM represents the multiple meanings of a word as a density matrix naturally and harmoniously, while the classical method is lacking such a theoretical guidance.

4.2. Semantic Matching for Metaphor Detection

Given ρ_t, ρ_l, and ρ_c, we use the joint matrix (Zhang et al., 2018, 2022) for semantic matching. QMM implements SPV and MIP methods by the joint matrix:

$$\begin{aligned} \rho_{SPV}^j &= \rho_t \rho_c \\ \rho_{MIP}^j &= \rho_t \rho_l. \end{aligned} \quad (12)$$

The joint matrix is a more generalized feature interaction method for the matching pairs and is also a representation method close to the quantum relative entropy (Zhang et al., 2018). Intuitively, the joint matrix is better at capturing features in terms of 'angle', which may ignore features in terms of 'distance'. Inspired by the definition of the trace distance (Nielsen and Chuang, 2010), we propose a difference matrix for SPV and MIP:

$$\begin{aligned} \rho_{SPV}^d &= |\rho_t - \rho_c| \\ \rho_{MIP}^d &= |\rho_t - \rho_l|. \end{aligned} \quad (13)$$

Here $|.|$ denotes the computation of L1 paradigm. Finally, based on MIP and SPV, we obtain the interaction features:

$$H_{SPV} = [\rho^j_{SPV}; \rho^d_{SPV}]$$
$$H_{MIP} = [\rho^j_{MIP}; \rho^d_{MIP}], \tag{14}$$

where $H_{SPV} \in \mathbb{C}^{2\times d\times d}$ and $H_{MIP} \in \mathbb{C}^{2\times d\times d}$ can be considered as two "images" with two channels.

4.3. Training and Optimization

The complex-valued features H_{SPV} and H_{MIP} encapsulate the semantic interaction information from the perspectives of MIP and SPV. QMM further introduces complex convolutional layers (Chiheb et al., 2017). It can be represented as:

$$Z = \begin{bmatrix} \Re(\mathbf{W} * H) \\ \Im(\mathbf{W} * H) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} * \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \tag{15}$$

where the real and imaginary parts of the convolution kernel $\mathbf{W}$ are represented by A and B, respectively. The real and imaginary parts of the semantic interaction feature H_{SPV} or H_{MIP} are represented by $\mathbf{x}$ and $\mathbf{y}$, respectively. $\Re(.)$ denotes the operation of taking the real part, and $\Im(.)$ denotes the operation of taking the imaginary part. The convolved results are unfolded and concatenated, serving as the input to a linear layer, which generates an output y_i. The network parameters are optimized using cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i} y_i \log(\hat{y}_i), \tag{16}$$

where the convolution kernel parameters are updated by $\nabla_L(\mathbf{W}) = \nabla_L(\Re(\mathbf{W})) + i\nabla_L(\Im(\mathbf{W}))$.

5. Experiments and Analysis

In this section, we conducted experiments to validate the effectiveness of QMM.

5.1. Experimental Settings

5.1.1. Datasets

We validated QMM on three most well-known datasets:

- **VUA ALL** (Steen et al., 2010) is the largest metaphor dataset extracted from the VU Amsterdam Metaphor Corpus (VUA).
- **VUA Verb** (Steen et al., 2010) is a subset of VUA ALL, which only involves metaphor annotations for verbs.
- **MOH-X** (Mohammad et al., 2016) also only involves metaphor annotations for verbs. It collects usages of various verbs from WordNet and annotates them for metaphor.

5.1.2. Implementation Details

All the gloss data used in this experiment is extracted from WordNet 3.1. We set the size of the density matrix to 32×32. We used 32 convolution kernels with a size of 3×3 for convolution. The kernels were slid one step at a time. We employed pre-trained RoBERTa and fine-tuned it 200 epochs. For VUA ALL and VUA Verb, we set the learning rate of the pre-trained model to 3e-6, and the learning rate of QMM to 3e-3. We report the best results on the test set. For MOH-X, we set the learning rates of the pre-trained model and QMM to 3e-6. Since this dataset is relatively small, we added L1 regularization with a weight of 1e-6 and reported the results of 10-fold cross-validation. All experiments are implemented with PyTorch v1.7.0 and run on a single NVIDIA Tesla 3090 GPU.

5.2. Baselines

We select several representative methods from the sequence labeling-based methods, the multi-task learning-based methods, the classification based methods and the matching-based methods respectively as baselines.

- **RNN_ELMo** (Gao et al., 2018) and **RNN_BERT** (Mao et al., 2019) encode semantic representations using ELMo and BERT, respectively.
- **RNN_HG** and **RNN_MHCA** (Mao et al., 2019) consider the sequence labeling task paradigm, and they further combine linguistic theories MIP and SPV, respectively.
- **RoBERTa_SEQ** (Leong et al., 2020) is a baseline model in the ACL 2020 metaphor detection shared task.
- **MUL_GCN** (Le et al., 2020) is based on a graph convolutional neural network, which can perform metaphor detection and semantic disambiguation.
- **DeepMet** (Su et al., 2020) regards metaphor detection as a reading comprehension tasks.
- **MrBERT** (Song et al., 2021) regards metaphor detection as a relation classification task between verbs and nouns and then uses BERT for fine-tuning.
- **MelBERT** (Choi et al., 2021) and **MisNet** (Zhang and Liu, 2022) are both based on linguistic theories. In MIP module, MelBERT uses an independent encoder to encode the literal meaning of words, while MisNet uses the basic usage of words (the first gloss) in the dictionary as the literal meaning of words.

Table 1: The overall experimental results. The best result is marked in boldface and the second best result is underlined.

Model	MOH-X				VUA All				VUA Verb			
	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1	Acc.
RNN_ELMo	79.1	73.5	75.6	77.2	71.6	73.6	72.6	93.1	68.2	71.3	69.7	81.4
RNN_BERT	75.1	81.8	78.2	78.1	71.5	71.9	71.7	92.9	66.7	71.5	69.0	80.7
RNN_HG	79.7	79.8	79.8	79.7	71.8	76.3	74.0	93.6	69.3	72.3	70.8	82.1
RNN_MHCA	77.5	83.1	80.0	79.8	73.0	75.7	74.3	93.8	66.3	75.2	70.5	81.8
MUL_GCN	79.7	80.5	79.6	79.9	74.8	75.5	75.1	93.8	72.5	70.9	71.7	83.2
RoBERTa_SEQ	-	-	-	-	80.4	74.9	77.5	-	79.2	69.8	74.2	-
DeepMet	-	-	-	-	82.0	71.3	76.3	-	79.5	70.8	74.9	-
MelBERT	-	-	-	-	80.1	76.9	78.5	-	78.7	72.9	75.7	-
MrBERT	80.0	85.1	82.1	81.9	**82.7**	72.5	77.2	94.7	**80.8**	71.5	75.9	**86.4**
MisNet	84.2	84.0	83.4	83.6	80.4	**78.4**	**79.4**	94.9	78.3	73.6	75.9	86.0
QMM(ours)	**86.0**	**86.7**	**86.3**	**86.7**	80.9	77.8	79.3	**95.0**	73.9	**79.0**	**76.4**	85.3

5.3. Overall Results

The overall results of the experiment are shown in Table 1. We report the results of QMM on evaluation metrics Precision (Pre.), Recall (Rec.), F1-measure (F1), and Accuracy (Acc.) respectively. The QMM receives the highest scores of 86.0, 86.7, 86.3, and 86.7 on the four evaluation metrics for MOH-X. It achieved overwhelming performance, with improvements of 2.9% and 3.1% in F1-measure and Accuracy, respectively, compared to the state-of-the-art model. On VUA All dataset, QMM achieved the highest Accuracy score of 95. On VUA Verb dataset, QMM achieved the highest Recall score of 79.0 and the highest F1-measure score of 76.4. These results indicate that our method is highly competitive and can achieve state-of-the-art performance on some datasets.

However, it is worth noting that QMM exhibits inferior performance compared to MOH-X on VUA. We possess a fundamental comprehension regarding this matter. In dataset MOH-X, the majority of target words can be covered by WordNet. The VUA Verb dataset, however, exhibits greater complexity as it encompasses target words that are not accounted for in WordNet. The VUA ALL dataset includes all possible POS. Since WordNet only covers nouns, verbs, adjectives, and adverbs, there might be more target words not covered by WordNet, resulting in a performance that is inferior to that achieved on the first two datasets. Therefore, using a dictionary that covers more target words might yield better results. However, considering the challenge of obtaining a larger dictionary, we ultimately chose WordNet for its broader influence and ease of use.

5.4. Ablation Study

To validate the role of each component in the proposed method, we performed ablation experiments. **-MIP** and **-SPV** indicate that MIP and SPV are not used, respectively. **-C** indicates that complex-valued features are eliminated. **-Q** indicates that QLM is not used, in which case the model will degenerate into a model with multiple glosses added by weights, similar to Wan et al. (2021) and Su et al. (2021).

5.4.1. Module Ablation

The results of ablation experiments are presented in Table 2. Firstly, the results demonstrate that MIP and SPV are complementary metaphor detection methods, and the combination of the two methods yields the best results. In addition, the use of QLM resulted in a significant improvement, which may be due to the density matrix implementing more fine-grained feature interaction and uncertainty modeling of a literal meaning. After adding the complex-valued feature, QMM further improved indicating that the positional relationship of the target word relative to other words is an important feature for metaphor detection. At the same time, we also observed that the performance of MIP surpassed SPV when complex-valued features were introduced, indicating that the relative order of glosses in MIP method is a significant factor that can help us model the literal meaning of the target word more effectively.

5.4.2. Matching Mode Ablation

We also analyzed the matching mode, and the experimental results are shown in Table 3. We found that the proposed difference matrix outperformed the joint matrix proposed in previous works (Zhang et al., 2018, 2022). When combining both the joint matrix and the difference matrix, QMM achieved the best performance.

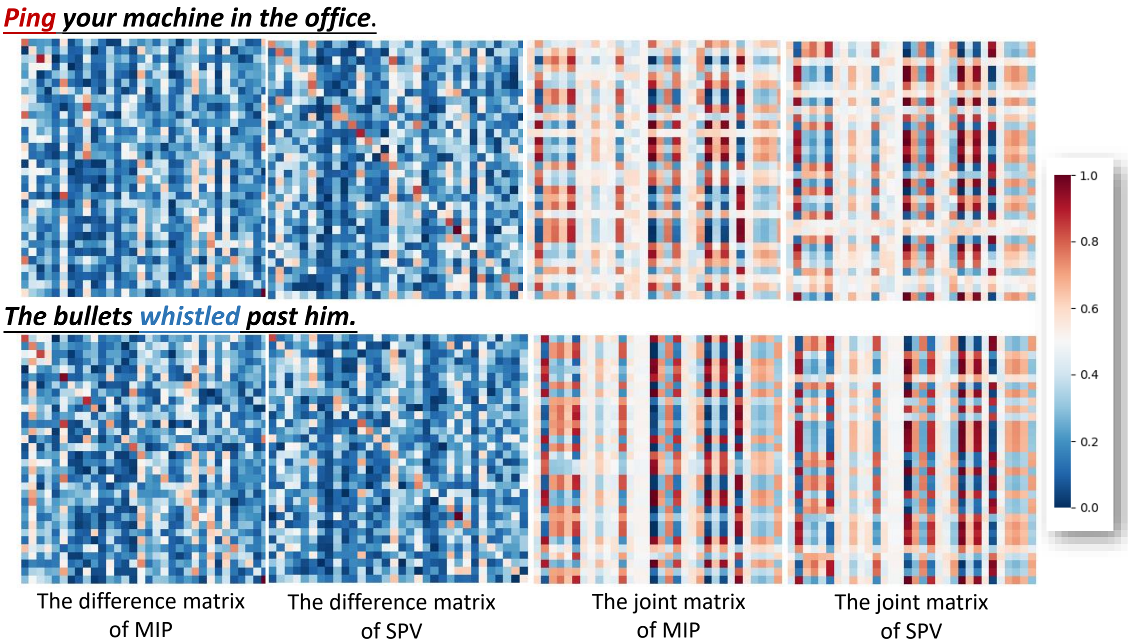

Figure 4: The difference matrix and joint matrix are formed by MIP and SPV on two sentences, where *"whistled"* is a metaphorical and *"Ping"* is literal. The more red areas in the difference matrix, the greater the semantic difference and likelihood of it being a metaphor. The joint matrix is the opposite.

Table 2: The results of the ablation experiment.

Ablation	Pre.	Rec.	F1	Acc.
-SPV-C-Q	79.8	80.9	79.8	80.0
-MIP-C-Q	79.6	83.2	80.6	80.6
-C-Q	81.9	83.2	82.5	82.9
-SPV-C	79.7	83.7	81.3	81.2
-MIP-C	81.0	82.3	81.4	81.6
-C	84.5	83.3	83.7	84.5
-SPV	85.3	86.4	85.6	85.9
-MIP	**86.1**	85.5	85.5	85.9
QMM(ours)	86.0	**86.7**	**86.3**	**86.7**

Table 3: The results of the matching mode.

Matching mode	Pre.	Rec.	F1	Acc.
Joint matrix	85.8	86.5	85.8	85.9
Difference matrix	85.2	**87.6**	86.1	86.2
QMM(ours)	**86.0**	86.7	**86.3**	**86.7**

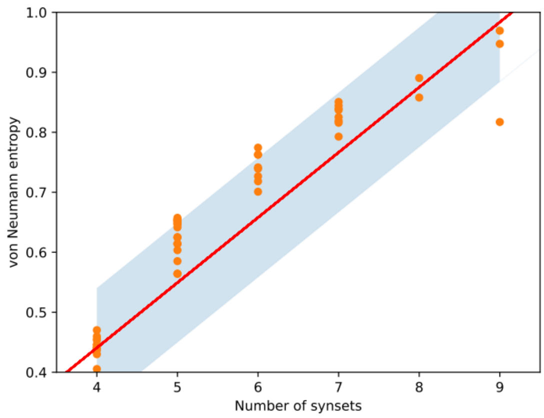

Figure 5: The literal semantic density matrix is formed by random sampling of 100 samples.

5.5. Effectiveness Study

In order to interpret the literal semantic density matrix and matching mode, we perform visual analysis.

5.5.1. Density Matrix Analysis

To verify that the learned literal semantic density matrix can model the phenomenon of polysemy and reflect the ambiguity of the literal meaning of the word, we calculated the von Neumann entropy according to the formula: $S(\rho) = -tr(\rho \log \rho)$. The final results are shown in Figure 5. It shows that there is a certain degree of positive correlation between the von Neumann entropy of the density matrix and the number of synsets corresponding to the word, which indicates that QMM reflect the phenomenon of polysem (Meyer and Lewis, 2020) and is a reasonable method for characterizing the literal meaning of words.

5.5.2. Matching Mode Analysis

Finally, the visualization analysis of the matching mode is shown in Figure 4. We found the difference matrix produces a diagonal effect, and most of the semantic differences are concentrated on the diagonal elements. In contrast, the joint matrix produces horizontal or vertical textures, and most of semantic differences are concentrated on the horizontal and vertical textures, intermittently. These results indirectly explain why QLM (Zhang et al., 2018, 2022) can not obtain optimal performance by directly extracting the traces of the joint matrix. Because the semantic overlap of the joint density matrix is not concentrated on the diagonal, which is why we need convolutional neural networks. In addition, we observe that the second sentence has more red areas on the diagonal in the difference density matrix of MIP, suggesting that the sentence is more likely to be a metaphor, which is indeed the case. Similar phenomena can also be observed in the joint matrix of MIP, where there are more blue areas in the second example sentence, indicating that the sentence is more likely to be a metaphor.

6. Conclusions

This paper proposes a quantum-inspired matching network with linguistic theories for metaphor detection (QMM). It not only realizes more granular semantic interaction but also models the literal meaning of uncertainty of target words. The experimental results show the competitiveness of this method. To our knowledge, this is the first work to use QLM for metaphor detection. This provides a new application direction and innovative ideas for the intersection of quantum mechanics and natural language processing.

7. Ethics Statement

Our research aims to advance the development of natural language processing technologies to address specific real-world problems. The data used in the experiments are all public datasets, and we will follow the highest ethical standards. Furthermore, we commit to open-sourcing our research findings to promote sharing and collaboration within the field.

8. Acknowledgements

This work is supported in part by the Natural Science Foundation of China (grant No. 62276188), TJU-Wenge joint laboratory funding.

9. Bibliographical References

Yiwei Chen, Yu Pan, and Daoyi Dong. 2021. Quantum language model with entanglement embedding for question answering. *IEEE Transactions on Cybernetics*.

Trabelsi Chiheb, O Bilaniuk, D Serdyuk, et al. 2017. Deep complex networks. In *International Conference on Learning Representations*.

Minjin Choi, Sunkyung Lee, Eunseong Choi, Heesoo Park, Junhyuk Lee, Dongwon Lee, and Jongwuk Lee. 2021. Melbert: Metaphor detection via contextualized late interaction using metaphorical identification theories. *arXiv preprint arXiv:2104.13615*.

Verna Dankers, Karan Malhotra, Gaurav Kudva, Volodymyr Medentsiy, and Ekaterina Shutova. 2020. Being neighbourly: Neural metaphor identification in discourse. In *Proceedings of the second workshop on figurative language processing*, pages 227–234.

Verna Dankers, Marek Rei, Martha Lewis, and Ekaterina Shutova. 2019. Modelling the interplay of metaphor and emotion through multitask learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2218–2229.

Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. 2018. Neural metaphor detection in context. In *Conference on Empirical Methods in Natural Language Processing*.

Dimitris Gkoumas, Qiuchi Li, Shahram Dehdashti, Massimo Melucci, Yijun Yu, and Dawei Song. 2021. Quantum cognitively motivated decision fusion for video sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 827–835.

Andrew M Gleason. 1975. Measures on the closed subspaces of a hilbert space. In *The Logico-Algebraic Approach to Quantum Mechanics: Volume I: Historical Evolution*, pages 123–133. Springer.

Hongyu Gong, Kshitij Gupta, Akriti Jain, and Suma Bhat. 2020. Illinimet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 146–153.

Pragglejaz Group. 2007. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and symbol*, 22(1):1–39.

Sooji Han, Rui Mao, and Erik Cambria. 2022. Hierarchical attention network for explainable depression detection on twitter aided by metaphor concept mappings. *arXiv preprint arXiv:2209.07494*.

Yongyu Jiang, Peng Zhang, Hui Gao, and Dawei Song. 2020. A quantum interference inspired neural matching model for ad-hoc retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 19–28.

George Lakoff and Mark Johnson. 1980. Metaphors we live by. *University of Chicago, Chicago, IL*.

Duong Le, My Thai, and Thien Nguyen. 2020. Multi-task learning for metaphor detection with graph convolutional neural networks and word sense disambiguation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8139–8146.

Chee Wee Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale, and Xianyang Chen. 2020. A report on the 2020 vua and toefl metaphor detection shared task. In *Proceedings of the second workshop on figurative language processing*, pages 18–29.

Qiuchi Li, Dimitris Gkoumas, Alessandro Sordoni, Jian-Yun Nie, and Massimo Melucci. 2021a. Quantum-inspired neural network for conversational emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13270–13278.

Qiuchi Li, Benyou Wang, and Massimo Melucci. 2019. Cnm: An interpretable complex-valued network for matching. *arXiv preprint arXiv:1904.05298*.

Shuqun Li, Liang Yang, Weidong He, Shiqi Zhang, Jingjie Zeng, and Hongfei Lin. 2021b. Label-enhanced hierarchical contextualized representation for sequential metaphor identification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3533–3543.

Yaochen Liu, Yazhou Zhang, Qiuchi Li, Benyou Wang, and Dawei Song. 2021. What does your smile mean? jointly detecting multi-modal sarcasm and sentiment using quantum probability. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 871–880.

Rui Mao and Xiao Li. 2021. Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13534–13542.

Rui Mao, Chenghua Lin, and Frank Guerin. 2018. Word embedding and wordnet based metaphor identification and interpretation. In *Proceedings of the 56th annual meeting of the association for computational linguistics*. Association for Computational Linguistics (ACL).

Rui Mao, Chenghua Lin, and Frank Guerin. 2019. End-to-end sequential metaphor identification inspired by linguistic theories. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 3888–3898.

Rowan Hall Maudslay and Simone Teufel. 2022. Metaphorical polysemy detection: Conventional metaphor meets word sense disambiguation. *arXiv preprint arXiv:2212.08395*.

Francois Meyer and Martha Lewis. 2020. Modelling lexical ambiguity with density matrices. *arXiv preprint arXiv:2010.05670*.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Saif Mohammad, Ekaterina Shutova, and Peter Turney. 2016. Metaphor as a medium for emotion: An empirical study. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 23–33.

Jesse Mu, Helen Yannakoudakis, and Ekaterina Shutova. 2019. Learning outside the box: Discourse-level features improve metaphor identification. *arXiv preprint arXiv:1904.02246*.

Michael A Nielsen and Isaac L Chuang. 2010. *Quantum computation and quantum information*. Cambridge university press.

Wei Song, Shuhui Zhou, Ruiji Fu, Ting Liu, and Lizhen Liu. 2021. Verb metaphor detection via contextual relation learning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4240–4251.

Alessandro Sordoni, Jian-Yun Nie, and Yoshua Bengio. 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*, pages 653–662.

Gerard Steen, Aletta G Dorst, J Berenike Herrmann, Anna Kaal, Tina Krennmayr, Trijntje Pasma, et al. 2010. A method for linguistic metaphor identification. *Amsterdam: Benjamins*.

Chang Su, Kechun Wu, and Yijiang Chen. 2021. Enhanced metaphor detection via incorporation of external knowledge based on linguistic theories. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1280–1287.

Chuandong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang, and Zhiqun Chen. 2020. Deepmet: A reading comprehension paradigm for token-level metaphor detection. In *Proceedings of the second workshop on figurative language processing*, pages 30–39.

Hisaharu Umegaki. 1954. Conditional expectation in an operator algebra. *Tohoku Mathematical Journal, Second Series*, 6(2-3):177–181.

Cornelis Joost Van Rijsbergen. 2004. *The geometry of information retrieval*. Cambridge University Press.

Hai Wan, Jinxia Lin, Jianfeng Du, Dawei Shen, and Manrong Zhang. 2021. Enhancing metaphor detection by gloss-based interpretations. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1971–1981.

Yorick Wilks. 1975. A preferential, pattern-seeking, semantics for natural language inference. *Artificial intelligence*, 6(1):53–74.

Chuhan Wu, Fangzhao Wu, Yubo Chen, Sixing Wu, Zhigang Yuan, and Yongfeng Huang. 2018. Neural metaphor detecting with cnn-lstm model. In *Proceedings of the workshop on figurative language processing*, pages 110–114.

Omnia Zayed, John Philip McCrae, and Paul Buitelaar. 2020. Contextual modulation for relation-level metaphor identification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 388–406.

Peng Zhang, Wenjie Hui, Benyou Wang, Donghao Zhao, Dawei Song, Christina Lioma, and Jakob Grue Simonsen. 2022. Complex-valued neural network-based quantum language models. *ACM Transactions on Information Systems (TOIS)*, 40(4):1–31.

Peng Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and Dawei Song. 2018. End-to-end quantum-like language models with application to question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Shenglong Zhang and Ying Liu. 2022. Metaphor detection via linguistics enhanced siamese network. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4149–4159.

Yazhou Zhang, Yaochen Liu, Qiuchi Li, Prayag Tiwari, Benyou Wang, Yuhua Li, Hari Mohan Pandey, Peng Zhang, and Dawei Song. 2021. Cfn: a complex-valued fuzzy network for sarcasm detection in conversations. *IEEE Transactions on Fuzzy Systems*, 29(12):3696–3710.