Quantum-Inspired Interactive Networks for Conversational Sentiment Analysis

Yazhou Zhang[1,3], **Qiuchi Li**[4], **Dawei Song**[2,5*], **Peng Zhang**[1] and **Panpan Wang**[1]
[1]College of Intelligence and Computing, Tianjin University, Tianjin, China
[2]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[3]Zhejiang Lab, Hangzhou, China
[4]Department of Information Engineering, University of Padua, Padua, Italy
[5]School of Computing and Communications, The Open University, United Kingdom
{dawei.song2010}@gmail.com

Abstract

Conversational sentiment analysis is an emerging, yet challenging Artificial Intelligence (AI) subtask. It aims to discover the affective state of each participant in a conversation. There exists a wealth of interaction information that affects the sentiment of speakers. However, the existing sentiment analysis approaches are insufficient in dealing with this task due to ignoring the interactions and dependency relationships between utterances. In this paper, we aim to address this issue by modeling intra-utterance and inter-utterance interaction dynamics. We propose an approach called quantum-inspired interactive networks (QIN), which leverages the mathematical formalism of quantum theory (QT) and the long short term memory (LSTM) network, to learn such interaction dynamics. Specifically, a density matrix based convolutional neural network (DM-CNN) is proposed to capture the interactions within each utterance (i.e., the correlations between words), and a strong-weak influence model inspired by quantum measurement theory is developed to learn the interactions between adjacent utterances (i.e., how one speaker influences another). Extensive experiments are conducted on the MELD and IEMOCAP datasets. The experimental results demonstrate the effectiveness of the QIN model.

1 Introduction

Sentiment analysis (SA) targets at judging sentiment polarities for various types of texts at document, sentence or aspect levels [Tripathy *et al.*, 2017; Yang and Cardie, 2014; Pontiki *et al.*, 2016]. The recent boom of social network services produces a huge volume of textual records of communications between humans. Such data carry a rich source of information including sentiments or opinions, which often evolve during the conversation. It brings forth a new challenge of judging the evolving sentiment polarities of different people in a conversational discourse. Therefore, the research on *conversational sentiment analysis* has attracted an increasing attention from both academia and industry.

*The corresponding author.

Conversational sentiment analysis aims to detect the affective states of multiple speakers during and after an conversation, and study the sentimental evolution of each speaker in the course of the interaction. The interaction dynamics in a conversation mainly consists of intra- and inter-utterance interactions. Intra-utterance interaction refers to the correlations between terms within an utterance, while inter-utterance interaction involves repeated interactions between the speakers' utterances. Fig. 1 provides an example from the MELD dataset [Poria *et al.*, 2018], from which, we can notice that the evolution of *Jen* and *Ross*'s affective states is influenced by both intra- and inter-utterance interactions.

Existing research in conversational sentiment analysis mainly focused on leveraging intra-utterance interactions, e.g., learning relations between words, extracting effective features, etc., to judge sentiment, while the inter-utterance interactions are largely neglected. For instance, Ojamaa et al. [2015] used a lexicon-based method to extract speakers' attitudes from conversational texts. However, they neglected the interactions and used only 23 dialogue files. Bhaskar et al. [2015] proposed to combine acoustic and textual features in audio conversations to enhance the efficiency of emotion classification. However, they did not consider interactions among speakers. Huijzer et al. [2017] performed affective analysis of emails. They did notice, but did not model the interaction between customer support and customers.

In recent years, quantum theory (QT), as a mathematical formalism to model the complex interactions and dynamics in quantum physics, has been adopted for constructing text representation in various information retrieval (IR) and NLP tasks [Sordoni *et al.*, 2013; Wang *et al.*, 2018; Li *et al.*, 2018; Zhang *et al.*, 2018c]. For instance, the Quantum Language Model (QLM) [Sordoni *et al.*, 2013] represents a query or document as a density matrix on a quantum probabilistic space, and computes density matrix-based metrics as ranking function. the Neural Network based QLM (NNQLM) [Zhang *et al.*, 2018a] builds an end-to-end network for question answering (QA) to jointly model a question-answer pair based on their density matrix representations. Zhang et al. [2018b] leverages an improved version of QLM for twitter sentiment analysis. Such QT-based models can be considered as a generalization of classical approaches in that they are capable of capturing inherent intricacies in interactions. This motivates us to explore the use of quantum theory as a theoretical basis

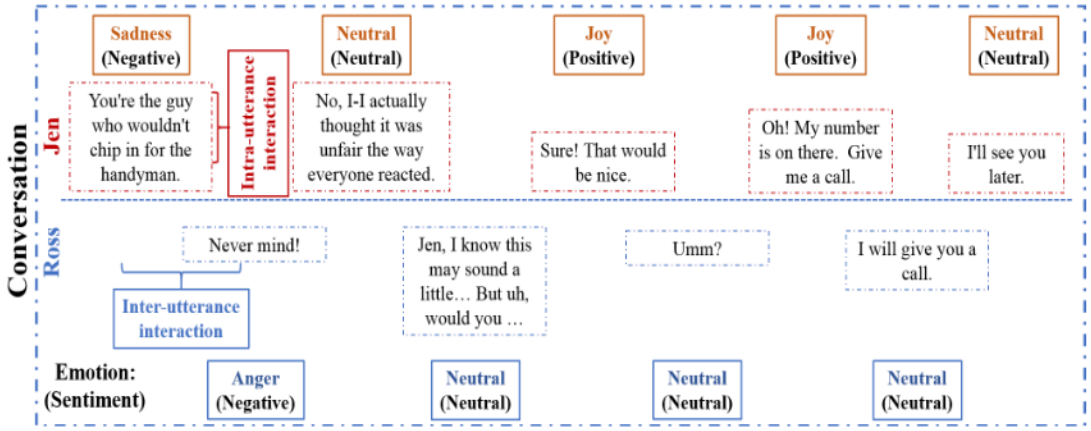

Figure 1: Two interaction dynamics in a conversation. Red and blue colors are used to show the emotion shift of Jen and Ross respectively.

for capturing the intra- and inter-utterance interaction dynamics, both of which are complex in nature.

In this paper, we propose quantum-inspired interactive networks (QIN) that jointly captures intra- and inter-utterance interactions for conversational sentiment analysis. The model extracts textual features with a density matrix based CNN (DM-CNN) to capture the intra-utterance correlations between words. The inter-utterance interaction is extracted by a strong-weak influence model, which is inspired by quantum measurement theory, to measure the influence between speakers across utterances. The influence is integrated into the output gates of an LSTM with textual features as inputs. The obtained hidden states of the LSTM are fed to a *softmax* function to determine the affective state for each utterance.

We have designed and carried extensive experiments on the MELD and IEMOCAP datasets to evaluate the QIN model, in comparison with a wide range of baselines, including six deep neural network approaches: a deep convolutional neural network (CNN) and five variants of long short term memory (LSTM) networks. The experimental results demonstrate the effectiveness of the QIN model.

2 Quantum Theory Preliminaries

2.1 Basic Notations and Concepts in Quantum Theory

In quantum theory, the probabilistic space for quantum theory is an infinite Hilbert Space [Bourbaki, 1966] [1], noted as $\mathbb{H}$. For simplicity and in line with previous quantum-inspired models [Zhang *et al.*, 2018a; Wang *et al.*, 2018], we restrict our model to finite vector spaces over real numbers in $\mathbb{R}$.

With Dirac's notation, a state vector φ and its transpose are expressed as a Ket $|\varphi\rangle$ and a Bra $\langle\varphi|$ respectively. The inner product between two state vectors $|\varphi_1\rangle$ and $|\varphi_2\rangle$, is represented as $\langle\varphi_1|\varphi_2\rangle$. Similarly, the representation of the wave function (which is also a mathematical description of the quantum state) in Hilbert Space is given by the inner product $\varphi(x) = \langle x|\varphi\rangle$.

Quantum Probability (QP) is a generalization of the classical probability theory. In QP, an event is a subspace of Hilbert Space represented by an orthogonal projector Π. Assume $|u\rangle$ is a unit vector, i.e., $\|\vec{u}\|_2 = 1$, the projector Π on the direction u is written as $|u\rangle\langle u|$. $\rho = \sum_i p_i |u\rangle\langle u|$ can represent a density matrix. Density matrix ρ is symmetric, $\rho = \rho^T$, positive semi-definite ($\rho \geq 0$), and of trace 1. The quantum probability measure μ is associated with the density matrix. It satisfies two conditions: (1) for each projector $|u\rangle\langle u|$, $\mu(|u\rangle\langle u|) \in [0, 1]$, and (2) for any orthonormal basis $\{|e_i\rangle\}$, $\sum_{i=1}^{n} \mu(|e_i\rangle\langle e_i|) = 1$. The Gleason's Theorem [Gudder, 2014] has proven the existence of a mapping function $\mu(|u\rangle\langle u|) = tr(\rho|u\rangle\langle u|)$ for any $|u\rangle$.

The density matrix representation in QP provides a comprehensive mathematical formalism to capture the intra-utterance interactions, which will be detailed in Section 3.

[1] complex vector space possessing the structure of inner product

	Strong Measurement		Weak Measurement	
Variance	$\sigma < \lvert eigenvalue \rvert$		$\sigma \geqslant \lvert eigenvalue \rvert$	
Position in Eq.4	left side	right side	left side	right side
Suppose x_0 is around 1	$\frac{-(x-0)^2}{4\sigma^2} \to -\infty$ $e^{\frac{-(x-0)^2}{4\sigma^2}} \to 0$	$\frac{-(x-1)^2}{4\sigma^2} \to 0$ $e^{\frac{-(x-1)^2}{4\sigma^2}} \to 1$	$\frac{-(x-0)^2}{4\sigma^2} \to 0$ $e^{\frac{-(x-0)^2}{4\sigma^2}} \to 1$	$\frac{-(x-1)^2}{4\sigma^2} \to 0$ $e^{\frac{-(x-1)^2}{4\sigma^2}} \to 1$
The effect on quantum state	collapsed to $\lvert 1 \rangle$		biased a little	

Table 1: The parameter analysis for Equation 4

2.2 Preliminaries of Quantum Measurements

Quantum Measurement (QM) theory includes ordinary quantum measurements (i.e., strong measurements) and weak measurements. Quantum measurement consists of two steps: (i) the measurement device is weakly coupled to the quantum system; (ii) the measurement device is strongly measured, and its collapsed state is referred to as the outcome of the measurement process.

Let $|\phi_d\rangle$ denote the wave function of measurement device and represent the position basis. It could be written as:

$$|\phi_d\rangle = \int_x \phi(x)|x\rangle dx \tag{1}$$

$$\phi(x) = (2\pi\sigma^2)^{-\frac{1}{4}} e^{-x^2/4\sigma^2} \tag{2}$$

where x is the pointer position, and σ is the standard deviation of a normal distribution around 0.

Suppose the quantum system being measured S is in a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, in which α and β are the complex probability amplitudes satisfying $|\alpha|^2 + |\beta|^2 = 1$. $|0\rangle$ and $|1\rangle$ are the eigenstates corresponding to the eigenvalues 0, 1 respectively. The system and the measurement device are then entangled as such:

$$\int_x [e^{-\frac{(x-0)^2}{4\sigma^2}} \alpha|0\rangle \otimes |x\rangle + e^{-\frac{(x-1)^2}{4\sigma^2}} \beta|1\rangle \otimes |x\rangle] dx \quad (3)$$

The details of entanglement process can be referred to [Von Neumann, 2018]. Next, a strong measurement is carried out on the pointer of the measuring device, resulting in a collapse of the pointer to the state $|x_0\rangle$. The entangled system is also collapsed to:

$$[e^{-\frac{(x_0-0)^2}{4\sigma^2}} \alpha|0\rangle + e^{-\frac{(x_0-1)^2}{4\sigma^2}} \beta|1\rangle] \otimes |x_0\rangle \quad (4)$$

where the eigenvalue x_0 could be anywhere around 0 or 1, or even further away. Whether the quantum measurement is strong or weak is determined by the $\Delta = \sigma^2$. The collapse of the pointer biases the system's vector. However, if σ is very big, the bias will be very small and the system's outcome will be very similar to the original vector. A detailed analysis is shown in Table 1.

QM provides a principled and effective mechanism to capture the inter-utterance interactions.

3 Learning Interaction Dynamics with the Quantum-Inspired Interactive Networks

3.1 Problem Formulation and Network Procedure

In this work, we target determining the attitude of each speaker at the utterance (sentence) level, in terms of positive, negative and neutral. The problem we investigate thus takes each utterance u as input and produces its sentiment label y as output. Hence, we formulate the problem as follows:

In a multi-turn conversation, how can we capture the interactions between speakers to determine their emotional changes during the conversation?

The architecture of the proposed quantum-inspired interactive network (QIN) is shown in Fig. 3. We first extract textual features of conversational discourses $\vec{x} = [\vec{r}_1, \vec{r}_2, ..., \vec{r}_n]$ through a density matrix based convolutional neural network (DM-CNN), which takes the semantic dependencies into consideration. Second, inspired by quantum measurement theory, a strong-weak influence model is developed to compute the inter-utterance influences between speakers within the whole conversation, denoted as R. Last, an LSTM variant is built on top of the extracted textual features $\vec{x}$ to model the evolution of sentiments in the conversation, with the output gate o_t combined with the inter-utterance influences R.

3.2 Density Matrix-Based CNN

Nowadays, a series of pioneering studies provide the evidence that density matrix, which is defined on the quantum probabilistic space, could be applied in natural language processing as an excellent representation method [Sordoni *et al.*, 2013; Zhang *et al.*, 2018a; Li *et al.*, 2018]. Compared with embedding vector, density matrix could encode more semantic dependencies. Motivated by Zhang's work [Zhang *et al.*, 2018a], we develop a density matrix based convolutional neural network (DM-CNN) to represent utterances. The representation procedure is described below.

Suppose $|w_i\rangle = (w_{i1}, w_{i2}, ..., w_{id})^T$ is a normalized word vector. The projector Π_i for a single word w_i is formulated in Eq. (5). One-hot representation of words over other words is known to suffer from the curse of dimensionality and difficulty in representing ambiguous words. In this work, we use word embeddings to construct projectors in semantic space.

$$\Pi_i = |w_i\rangle\langle w_i| \quad (5)$$

Based on word projectors Π_i, we represent an utterance with a density matrix ρ_u, which is formulated as:

$$\rho_u = \sum_i p_i \Pi_i = \sum_i p_i |w_i\rangle\langle w_i| = \begin{bmatrix} \sum_i p_i (w_{i1})^2 & \sum_i p_i w_{i1} w_{i2} & ... & \sum_i p_i w_{i1} w_{id} \\ \sum_i p_i w_{i2} w_{i1} & \sum_i p_i (w_{i2})^2 & ... & \sum_i p_i w_{i2} w_{id} \\ \vdots & & ... & \\ \sum_i p_i w_{id} w_{i1} & \sum_i p_i w_{id} w_{i2} & ... & \sum_i p_i (w_{id})^2 \end{bmatrix} \quad (6)$$

where p_i is the probability of event (word) Π_i satisfying $\sum_i p_i = 1$. In this work, we set p_i with equal probabilities $p_i = \frac{1}{D}$, where D is the number of words in a document.

The utterance density matrix ρ_u is then fed into a deep CNN architecture to obtain abstract textual features. The CNN consists of two convolutional layers with max pooling, a fully connected layers and a softmax layer. The first convolutional layer has eight 5×5 filters. The second convolutional layer has sixteen 3×3 filters. The obtained textual features $\vec{x} = [\vec{r}_1, \vec{r}_2, ..., \vec{r}_n]$ are used as inputs of our QIN model.

3.3 A Quantum Measurement-Inspired Strong-Weak Influence Model

Influence is an indirect and invisible way of altering the state of an entity, and is thus difficult to model. When one talks to another person, he is often influenced by their style of interaction. We posit that the intensity of interaction determines whether a speaker's affective state might or might not change. In particular, a **strong interaction** is in effect if his affective state changes, indicating he is greatly affected by others. On the other hand, **weak interaction** is used to describe such influence as too weak to offer changes to a speaker's affective state.

In QT, quantum measurement describes the interaction (coupling) between a quantum system and the measurement device. Strong measurement leads to the collapse of the quantum system while weak measurement disturbs the quantum system very little. The variance of pointer readings of the measurement device could distinguish whether the interaction is strong or weak. In this work, we treat each speaker as a learning system. Accordingly, the interaction could be characterized as a coupling between two systems. The interaction between quantum system and the measurement device is exactly similar to the interaction between speakers. Inspired by this, we associate strong and weak interaction with quantum measurement, and develop a strong-weak influence model.

Specifically, our model is an extension of the dynamical "influence model", which generalizes HMM by constructing the influence matrices for describing the influence each Markov chain has on the others [Pan *et al.*, 2012]. Suppose there are C entities in the system, and each entity e is associated with a finite set of possible states $\{1, 2, ..., S\}$. At different time t, each entity e is in one of the states, denoted by $q_t^e \in \{1, 2, ..., S\}$. Influence is treated as the conditional dependence between each entity's current state q_t^e at time t and the previous states of all entities $q_{t-1}^1, q_{t-1}^2, ..., q_{t-1}^C$ at time $t-1$. q_t^e is only influenced by all entities at time $t-1$. Therefore, the conditional probability can be formulated as:

$$\begin{aligned} &P\left(q_t^e \middle| q_{t-1}^1, q_{t-1}^2, ..., q_{t-1}^e, ..., q_{t-1}^C\right) \\ &= \sum_{c \in 1,2,...,C} R\left(r_t\right)_{e,c} \times P\left(q_t^e | q_{t-1}^c\right) \end{aligned} \tag{7}$$

where $R\left(r_t\right)$ is a $C \times C$ matrix ($R\left(r_t\right)_{e,c}$ represents the element at the eth row and the cth column), $r_t \in \{1, 2, 3, ..., J\}$, $t = 1, ..., T$, and J is a hyperparameter set by users to define the number of influence matrices $R\left(r_t\right)$. $P\left(q_t^e | q_{t-1}^c\right)$ is the transition probability from state q_{t-1}^c to q_t^e, controlled by an $S \times S$ matrix $M^{c,e}$ specific to a pair of entities (c, e): $P\left(q_t^e | q_{t-1}^c\right) = M^{c,e}_{q_{t-1}^c, q_t^e}$, where $M^{c,e}_{q_{t-1}^c, q_t^e}$ represents the element at the q_{t-1}^cth row and q_t^eth column of matrix $M^{c,e}$.

However, in a turn-taking conversation, the speakers' states in each turn are influenced by both the current states of speakers who speak in front of e at turn (time) t, i.e., $q_t^1, q_t^2, ..., q_t^{e-1}$ and the previous states of other speakers who have not yet spoken (including the current speaker under concern) in the current round, i.e., $q_{t-1}^e, q_{t-1}^{e+1}, ..., q_{t-1}^C$. In particular, the state of the first speaker is influenced solely by previous states of all entities. The conditional probability then becomes

$$P\left(q_t^e | q_t^1, q_t^2, ..., q_t^{e-1}, q_{t-1}^e, q_{t-1}^{e+1}, ..., q_{t-1}^C\right) \tag{8}$$

Referring to the example shown in Fig. 1, i.e., $C = \{Jen\left(J\right), Ross\left(R\right)\}$. Each speaker is in one of three affective states, which are positive, negative and neutral, e.g., $S = 3$, and $q_t^R, q_t^J \in \{-1, 0, 1\}$. The conditional probability is measured as:

$$\begin{cases} P\left(q_t^J | q_{t-1}^J, q_{t-1}^R\right) \\ = R\left(r_t\right)_{JJ} \cdot P(q_t^J | q_{t-1}^J) + R\left(r_t\right)_{JR} \cdot P(q_t^J | q_{t-1}^R) \\ P\left(q_t^R | q_t^J, q_{t-1}^R\right) \\ = R\left(r_t\right)_{RJ} \cdot P(q_t^R | q_t^J) + R\left(r_t\right)_{RR} \cdot P(q_t^R | q_{t-1}^R) \end{cases} \tag{9}$$

where $R\left(r_t\right)_{JJ}$, $R\left(r_t\right)_{JR}$, $R\left(r_t\right)_{RJ}$, $R\left(r_t\right)_{RR}$ are four elements of the influence matrix $R\left(r_t\right)$, denoting how Jen influences Jen, how $Ross$ influences Jen, how Jen influences $Ross$, and how $Ross$ influences $Ross$.

Inspired by quantum measurement, we use two influence matrices (i.e., $J = 2, r_t \in \{1, 2\}$) to represent strong and weak influences. The switching of r_t is determined by the average standard deviation of speakers' sentimental scores σ_{avg}. We set the eigenvalues of speaker's affective state to -1, 0 and 1, i.e., $x \in \{-1, 0, 1\}$. Hence, we introduce the following prior for r_t:

$$\begin{cases} r_t = 1 & \text{if } \sigma_{avg} \geq \sum_x p\left(x\right) |x| \ \textit{weak influence} \\ r_t = 2 & \text{if } \sigma_{avg} < \sum_x p\left(x\right) |x| \ \textit{strong influence} \end{cases} \tag{10}$$

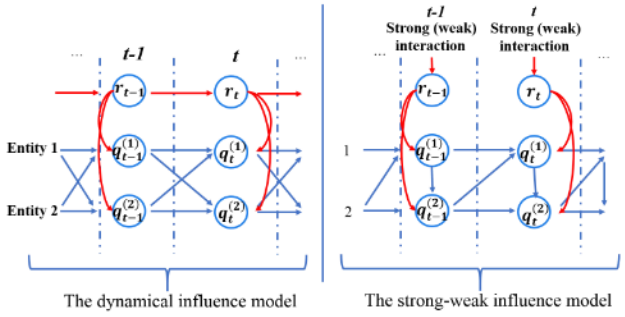

Figure 2: The difference between the dynamical influence model and the strong-weak influence model. The blue lines show the dependence, and the red lines indicate the switching influence.

where $p\left(x\right) = \left(2\sigma^2\pi\right)^{-\frac{1}{2}} e^{-(x-\mu_{avg})^2/2\sigma^2}$, denoting the probability density to get x, and μ_{avg} is set to the average of all expectations in this work.

We illustrate the difference between the dynamical influence model and the strong-weak influence model in Fig. 2. Finally, we obtain two influence matrices, which capture the strong and weak influences, i.e., $R\left(2\right)$ and $R\left(1\right)$, from one speaker over another speaker under different interactive environments[2].

3.4 Quantum-Inspired Interactive Networks

Since we have learned the interaction information (including interactions between terms and interactions between speakers), then we incorporate them into the quantum-inspired interactive networks (QIN), which is a variant of LSTM.

The QIN model is proposed for conversational sentiment analysis. The main idea is: (1) for each LSTM unit, combining the output gate o_t with the learned influence matrices $\mathbf{R}$ to constitute new output gate, describing what information we're going to output. The new output gate has considered the previous speakers' influences. (2) Taking textual vectors that are built by DM-CNN as inputs, obtaining their hidden states h_t, and thus making decisions. Fig. 3 represents the overall architecture of the QIN model.

Let $\vec{x}_t^{e_i} = [\vec{r}_1, \vec{r}_2, ..., \vec{r}_n]$ represents the input of speaker e_i, which has been learned by DM-CNN. $h_t^{e_i}$ represents the outputs of speaker e_i, where $t = \{1, 2, ..., T\}$. We put $h_t^{e_i}$ into a $softmax$ layer to obtain the sentiment label. That is,

$$y_t^{e_i} = softmax\left(W_s h_t^{e_i} + b_s\right) \tag{11}$$

where W_s and b_s are the parameters.

In the conversation, the influence that one speaker has on the other speaker would control the affected speaker's response. In Fig. 3, for two adjacent speakers (denoted as $e1$ and $e2$) at turn $t = 1$ (i.e., $Sp_{t=1}^{e1}$, $Sp_{t=1}^{e2}$), $Sp_{t=1}^{e1}$ actually determines how $Sp_{t=1}^{e2}$ is constructed. Furthermore, at the next turn $t = 2$, the construction of $Sp_{t=2}^{e1}$ would be influenced by both $Sp_{t=1}^{e1}$ and $Sp_{t=1}^{e2}$, and the construction of $Sp_{t=2}^{e2}$ would be influenced by both $Sp_{t=1}^{e2}$ and $Sp_{t=2}^{e1}$. Influence controls what information one speaker is going to flow out,

[2]The detailed inference process is given on https://github.com/anonymityanonymity/influence-model.git

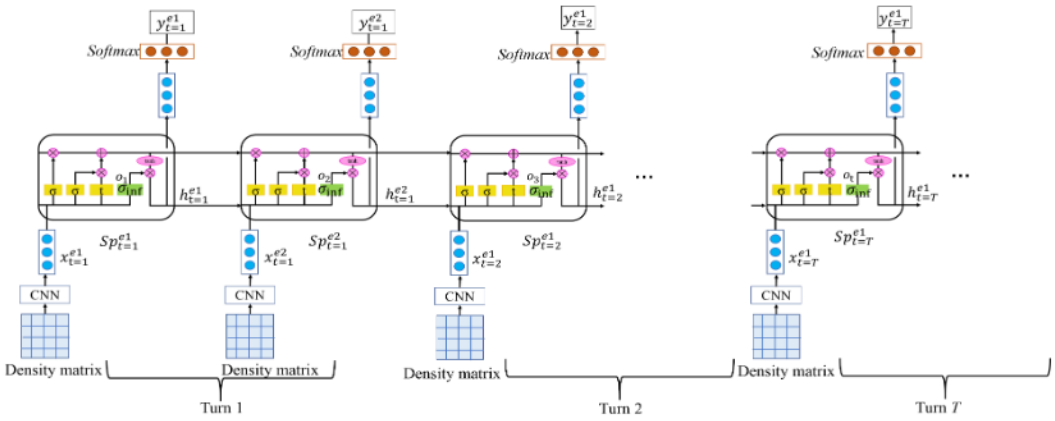

Figure 3: The architecture of quantum-inspired interactive networks.

which is similar to the role of the output gate. This influence has already been described by the influence matrix **R** (in Section 3.3). Hence, we consider the influences of the previous speakers on the current speaker by incorporating the influence scores into the computation of the output gates in the LSTM of QIN, which can be formulated as:

$$
\begin{aligned}
o^{e1}_{t|t=1} &= \sigma\left(W_{xo}\vec{x}^{e1}_t + b_o\right) \\
o^{e2}_{t|t=1} &= \sigma\left(W_{xo}\vec{x}^{e2}_t + W_{ho}h^{e1}_t + b_o\right) + \sigma(R_{e2,e1}\cdot\vec{x}^{e2}_t) \\
o^{e1}_{t|t\geq 2} &= \sigma\left(W_{xo}\vec{x}^{e1}_t + W_{ho}h^{e2}_{t-1} + b_o\right) \\
&\quad + \sigma(W_{e1}[R_{e1,e1}, R_{e1,e2}]\cdot\vec{x}^{e1}_t) \\
o^{e2}_{t|t\geq 2} &= \sigma\left(W_{xo}\vec{x}^{e2}_t + W_{ho}h^{e1}_t + b_o\right) \\
&\quad + \sigma(W_{e2}[R_{e2,e2}, R_{e2,e1}]\cdot\vec{x}^{e2}_t)
\end{aligned} \tag{12}
$$

where W_{e1} and W_{e2} are the normalized weights. $R_{e1,e1}$, $R_{e1,e2}$, $R_{e2,e1}$, $R_{e2,e2}$ are elements in $R(r_t)$.

Model training. In QIN model, cross entropy with L_2 regularization is used as the loss function, which is defined as:

$$
J = -\frac{1}{N}\sum_i\sum_j y^j_i log\hat{y}^j_i + \lambda_r \left\|\theta\right\|^2 \tag{13}
$$

where y_i denotes the ground truth, $\hat{y}_i$ is the predicted sentiment distribution. i is the index of utterance, j is the index of class. λ_r is the coefficient for $L2$ regularization. We use the back propagation method to compute the gradients and update all the parameters.

4 Experiments

4.1 Experimental Settings

Datasets. We conduct experiments on the MELD[3] and IEMOCAP[4] datasets to validate the effectiveness of QIN model. MELD contains 13,708 utterances from 1433 dialogues. The utterances in each dialogue are annotated with three sentiments (which are *positive*, *negative* and *neutral*) and seven emotions (which are *anger*, *disgust*, *fear*, *joy*, *neutral*, *sadness* and *surprise*). IEMOCAP is a multimodal database of ten speakers involved in two-way dyadic conversations. Each utterance is annotated using the following emotion categories: *anger*, *happiness*, *sadness*, *neutral*, *excitement*, *frustration*, *fear*, *surprise*, and *others*.

Evaluation metrics. Considering the imbalanced sample problem, we adopt *weighted* **F1 score**, **Accuracy** as the evaluation metrics to evaluate the classification performance. We employ t-test to perform the significance test.

Hyperparameters setting. In this work, we use the GloVe word vectors[5] [Pennington *et al.*, 2014] to find word embeddings. The dimensionality is set to 300. All weight matrices are given their initial values by sampling from a uniform distribution $U(-0.1, 0.1)$, and all biases are set to zeros. We set the initial learning rate to 0.001. The batch size is 60. The coefficient of $L2$ normalization in the objective function is set to 10^{-5}, and the dropout rate is set to 0.5.

4.2 Comparative Models

In order for a comprehensive evaluation of the QIN model, we include a range of baselines for comparison. They are listed as follows.

CNN. We employ a CNN [Kim, 2014] including three convolutional layers and a fully connected layer. It is trained on top of word embeddings for utterance-level classification.

LSTM & biLSTM. We implement a standard LSTM and bi-directional LSTM. They take word embeddings as input so as to get the hidden representation of each word.

ATAE-LSTM. We implement an attention based LSTM with aspect embedding [Wang *et al.*, 2016]. We obtain aspect embeddings by averaging the vectors of words, and append it with each word embedding vector.

Contextual biLSTM & Hierarchical biLSTM. We implement a contextual biLSTM [Poria *et al.*, 2017] and hierarchical contextual biLSTM to model semantic dependency among the utterances.

[3]This dataset is available on https://affective-meld.github.io/.

[4]http://sail.usc.edu/iemocap/.

[5]Pre-trained word embedding of GloVe can download from: https://nlp.stanford.edu/projects/glove/

MELD dataset	Models	Metrics	
		F1	Accuracy
Sentiments (3-class)	CNN	0.604	0.609
	LSTM	0.626	0.630
	biLSTM	0.611	0.624
	ATAE-LSTM	0.615	0.628
	contextual biLSTM	0.632	0.643
	Hierarchical biLSTM	0.638	0.652
	QIN	**0.662**†	**0.679**†
Emotions (7-class)	CNN	0.537	0.560
	LSTM	0.546	0.575
	biLSTM	0.536	0.557
	ATAE-LSTM	0.517	0.579
	contextual biLSTM	0.554	0.597
	Hierarchical biLSTM	0.563	0.608
	QIN	**0.578**	**0.619**
IEMOCAP dataset	**Models**	**Metrics**	
		F1	Accuracy
Emotions (9-class)	CNN	0.239	0.333
	LSTM	0.318	0.322
	biLSTM	0.275	0.331
	ATAE-LSTM	0.316	0.326
	contextual biLSTM	0.329	0.344
	Hierarchical biLSTM	0.335	0.351
	QIN	**0.343**	**0.376**†

Table 2: Comparison with baselines. Best performances are in **bold**. The symbol † indicates the improvement of QIN model over the baselines are statistically significant.

4.3 Results and Analysis

Table 2 shows the performance comparison of QIN with other baselines. In the case of sentiment classification on MELD dataset, CNN, LSTM, biLSTM and ATAE-LSTM achieve worse performance against all neural network baselines, because they ignore the contextual dependencies among utterances. The complete meaning of an utterance might be determined by preceding utterances. Hence, the introduction of attention mechanism does not help improve the performance. This suggests the importance of contextual modeling. Moreover, on fine-grained (e.g., 7-class and 9-class) emotion classification tasks, all the above four baselines have their victory and defeat. This implies that distinguishing fine-grained emotions is a more difficult and intricate task. Through taking utterances as inputs, contextual biLSTM has extracted contextual features. Contextual biLSTM performs consistently better over other baselines. Our QIN takes a further step towards emphasizing the importance of modeling interactions. Through learning both the intra- and inter-utterance interaction dynamics, QIN achieves the best performance among all baselines.

In the case of emotion classification on IEMOCAP dataset, all models get very poor performance because of the large number of classes. However, QIN still achieves the best performance. Compared with contextual biLSTM, QIN improves the performance by 7.1% by accuracy. The main reason is that QIN has modelled more semantic dependencies and previous speakers' influence. The results demonstrate the effectiveness and necessity of modelling the interactions in conversational sentiment analysis.

Dataset	Models	Metrics	
		F1	Accuracy
MELD	DM-LSTM	0.654	0.663
	Influence-LSTM	0.628	0.635
	QIN	0.662	0.679
IEMOCAP	DM-LSTM	0.322	0.343
	Influence-LSTM	0.339	0.369
	QIN	0.343	0.376

Table 3: Ablated QIN for both MELD and IEMOCAP datasets.

5 Ablation Study

In this subsection, we design a series of sub-models to study the impact of different components of the QIN model: (1) DM-LSTM, which does not model influences, but only uses density matrix-based CNN; (2) Influence-LSTM, which only uses quantum measurement inspired by strong-weak influence model and incorporates influences into the output gate.

From Table 3, we observe that QIN achieves the best performance among all models. The results verify that modeling both intra- and inter-utterance interactions makes a positive contribution to judging the sentiment polarity of an utterance. Influence-LSTM is worse than DM-LSTM on MELD, but performs better on IEMOCAP. Because IEMOCAP only contains two-way dyadic conversations, and capturing the interactions between two speakers is easier. DM-LSTM achieves better performance than CNN, LSTM and ATAE-LSTM, showing that density matrix representation can more effectively encode the semantic dependencies and their probabilistic distribution information. Influence-LSTM outperforms the baselines, proving that modeling inter-utterance interactions benefits the sentiment classification performance.

6 Conclusions

In this paper, we propose the QIN model, which could capture the correlations between terms and measure the influence of the previous speakers. The main idea is to use a density matrix based CNN and a strong-weak influence model inspired by quantum measurement theory to model such interaction dynamics. The experimental results on MELD and IEMOCAP demonstrate that our proposed QIN largely outperforms a number of state-of-art sentiment analysis algorithms, and also prove the importance of modeling interactions.

Acknowledgments

This work is supported by The National Key Research and Development Program of China (grant No. 2018YFC0831704), Natural Science Foundation of China (grant No. U1636203, 61772363), Major Project of Zhejiang Lab (grant No. 2019DH0ZX01), and the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 721321.

References

[Bhaskar *et al.*, 2015] Jasmine Bhaskar, K Sruthi, and Prema Nedungadi. Hybrid approach for emotion classification of audio conversation based on text and speech mining. *Procedia Computer Science*, 46:635–643, 2015.

[Bourbaki, 1966] Nicolas Bourbaki. *Elements of Mathematics: General Topology. Part 1,[chapters 1-4]*. Hermann, 1966.

[Gudder, 2014] Stanley P Gudder. *Quantum probability*. Academic Press, 2014.

[Huijzer, 2017] Erwin Huijzer. Identifying effective affective email responses. 2017.

[Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[Li *et al.*, 2018] Qiuchi Li, Massimo Melucci, and Prayag Tiwari. Quantum language model-based query expansion. In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 183–186. ACM, 2018.

[Ojamaa *et al.*, 2015] Birgitta Ojamaa, Päivi Kristiina Jokinen, and Kadri Muischenk. Sentiment analysis on conversational texts. In *Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania*, number 109, pages 233–237. Linköping University Electronic Press, 2015.

[Pan *et al.*, 2012] Wei Pan, Wen Dong, Manuel Cebrian, Taemie Kim, and A Pentland. Modeling dynamical influence in human interaction. *IEEE Signal Processing Magazine*, 29(2):77–86, 2012.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

[Pontiki *et al.*, 2016] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, AL-Smadi Mohammad, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 19–30, 2016.

[Poria *et al.*, 2017] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 873–883, 2017.

[Poria *et al.*, 2018] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*, 2018.

[Sordoni *et al.*, 2013] Alessandro Sordoni, Jian-Yun Nie, and Yoshua Bengio. 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. ACM, 2013.

[Tripathy *et al.*, 2017] Abinash Tripathy, Abhishek Anand, and Santanu Kumar Rath. Document-level sentiment classification using hybrid machine learning approach. *Knowledge and Information Systems*, pages 1–27, 2017.

[Von Neumann, 2018] John Von Neumann. *Mathematical Foundations of Quantum Mechanics: New Edition*. Princeton university press, 2018.

[Wang *et al.*, 2016] Yequan Wang, Minlie Huang, Li Zhao, et al. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, 2016.

[Wang *et al.*, 2018] Panpan Wang, Tianshu Wang, Yuexian Hou, and Dawei Song. Modeling relevance judgement inspired by quantum weak measurement. In *European Conference on Information Retrieval*, pages 424–436. Springer, 2018.

[Yang and Cardie, 2014] Bishan Yang and Claire Cardie. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In *ACL (1)*, pages 325–335, 2014.

[Zhang *et al.*, 2018a] Peng Zhang, Jiabin Niu, Zhan Su, Benyou Wang, Liqun Ma, and Dawei Song. End-to-end quantum-like language models with application to question answering. 2018.

[Zhang *et al.*, 2018b] Yazhou Zhang, Dawei Song, Xiang Li, and Peng Zhang. Unsupervised sentiment analysis of twitter posts using density matrix representation. In *European Conference on Information Retrieval*, pages 316–329. Springer, 2018.

[Zhang *et al.*, 2018c] Yazhou Zhang, Dawei Song, Peng Zhang, Panpan Wang, Jingfei Li, Xiang Li, and Benyou Wang. A quantum-inspired multimodal sentiment analysis framework. *Theoretical Computer Science*, 2018.