Quantum-inspired Interactive Networks for Conversational Sentiment Analysis

Abstract

Conversational sentiment analysis is an emerging, yet challenging Artificial Intelligence (AI) subtask. It aims to discover the affective state of each person in a conversation. There exists a wealth of interaction information that affects the sentiment of speakers, which is crucial but largely ignored by existing sentiment analysis works. To fill this gap, we tackle the challenge of modeling intra-utterance and inter-utterance interaction dynamics for conversational sentiment analysis by inspiring from quantum theory, which has exhibited unique advantages in capturing inter-feature correlations for text modeling. We propose the quantum-inspired interactive networks (QIN), which leverages LSTM network and the mathematical formalism of quantum theory (QT), to learn both such dynamics. Specifically, a density matrix-based CNN (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 dataset. The experimental results demonstrate the effectiveness of the QIN model.

1 Introduction

As one of the core research topics in Natural Language Processing (NLP), 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 booming of social network services produces a huge volume of textual records of communications between humans, carrying a rich source of information including sentiments, comments or opinions, which is often evolving 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.

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 words 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] showing the presence of these two patterns in a conversation. From Fig. 1, we could notice that the evolution of *Jen* and *Ross*'s affective states is influenced by both intra- and inter-utterance interactions.

Existing works in conversational sentiment analysis mainly leverage intra-utterance interactions, e.g., learning relations between words, extracting textual features, etc., to judge sentiment, while the inter-utterance interactions are largely ignored. For instance, Ojamaa et al. [2015] used a lexicon-based technology to extract the speaker's attitude from conversational texts. However, they neglected the interaction information and used only 23 dialogue files, which were not suitable for machine learning based assessments. Bhaskar et al. [2015] proposed to combine both acoustic and textual features for emotion classification of audio conversations. Although they enhanced the efficiency of emotion classification, they did not consider interactions among speakers. Huijzer et al. [2017] performed affective analysis of emails. They had noticed, but did not model the interaction problem between customer support and a customer.

In recent years, quantum theory (QT) 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, Quantum Language Model (QLM) [Sordoni *et al.*, 2013] models query and document as density matrix on a quantum probabilistic space and compute density matrix-based metrics as the ranking function. NNQLM [Zhang *et al.*, 2018a] builds an end-to-end question answering (QA) network to jointly model a question-answer pair based on their density matrix representation. Zhang et al. [2018b] leverages an improved version of QLM for twitter sentiment analysis. In these tasks, QT-based models are often considered as a generalization of classical approaches with unique advantages in capturing inherent intricacies of features in interactions. This motivates

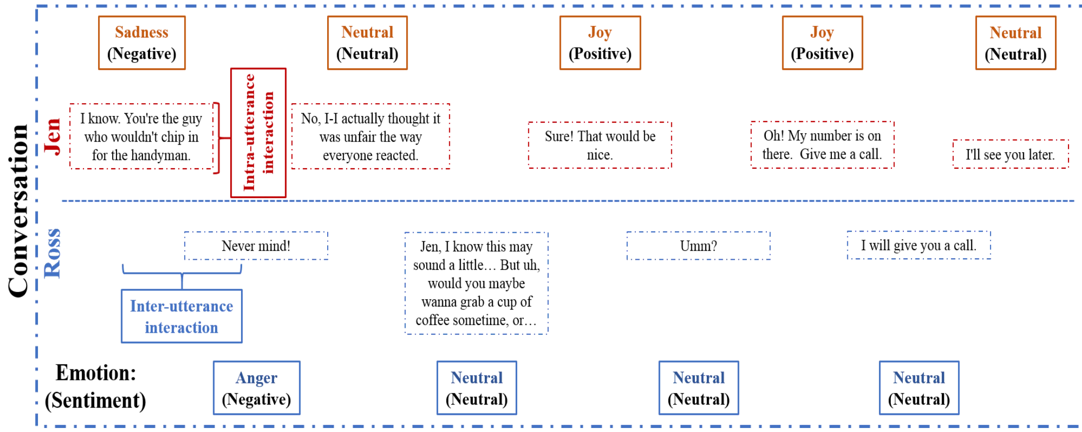


Figure 1: Two interaction dynamics in a conversation. Red and blue colors are used to show the emotion shift of Jen and Ross respectively.

us to explore the use of quantum theory as a theoretical support for capturing the intra- and inter-utterance interaction dynamics, both of which are complicated in nature.

In this paper, we try to improve existing works by proposing a quantum-inspired interactive networks (QIN) model that jointly captures intra- and inter-utterance interaction dynamics for conversational sentiment analysis. Specifically, we first extract textual features using the density matrix based CNN (DM-CNN) sub-network, which considers 2-order correlations between word vectors within an utterance. Second, we introduce a strong-weak influence model, inspired by quantum measurement theory, to measure the influence between speakers across utterances. The intra-utterance influence is fed into the model by being incorporated into the output gate of a proposed LSTM unit. With textual vectors as inputs, QIN obtains their hidden states, and feeds them to a *softmax* function to determine the affective state.

We have designed and carried out comparative experiments on the MELD dataset to evaluate the QIN model in comparison with a number of typical sentiment analysis models, including a machine learning based algorithm (which is Support Vector Machine, SVM) and five state-of-the-art neural network approaches: a deep convolutional neural network (CNN) and four long short term memory (LSTM) networks variants. The experimental results show that the proposed QIN model significantly outperforms a wide range of baselines, and achieves the state-of-the-art performance.

2 Quantum Preliminaries

This section gives a brief introduction to basic concepts of quantum theory, and further presents the background knowledge of quantum measurement theory.

2.1 Basic Notations and Concepts in Quantum Theory

In quantum theory, the quantum probability space is naturally encapsulated in an infinite Hilbert space [Bourbaki, 1966] (which is a complete vector space possessing the structure of an inner product), noted as $\mathbb{H}$. In line with previous quantum inspired models [Zhang *et al.*, 2018a; Wang *et al.*, 2018], we restrict our problem to vectors spaces over real numbers in $\mathbb{R}$.

With the Dirac's notation, a state vector, φ, can be expressed as a Ket $|\varphi\rangle$, and its transpose can be expressed as a Bra $\langle\varphi|$. In Hilbert space, any n-dimensional vector can be represented in terms of a set of basis vectors, $|\varphi\rangle = \sum_{i=1}^{n} a_i|e_i\rangle$. Given two state vectors $|\varphi_1\rangle$ and $|\varphi_2\rangle$, the inner product between them is represented as $\langle\varphi_1|\varphi_2\rangle$. Similarly, the Hilbert space representation of the wave function is recovered from the inner product $\varphi(x) = \langle x|\varphi\rangle$.

In quantum probability, an event is defined to be a subspace of Hilbert space, represented by any 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 also 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 [GLEASON, 1957] has proven the existence of a mapping function $\mu(|u\rangle\langle u|) = tr(\rho|u\rangle\langle u|)$ for any $|u\rangle$.

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 quantum 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 position variable of the measuring pointer.

As an example, let S denote the quantum system being measured. Suppose $\hat{O}$ is an observable on the system S. Take $\hat{O} = \frac{\hbar}{2}S$, $\hbar$ is Planck's constant that is the quantum of action. A quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, in which α and β are the

	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
Supposing 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

probability amplitudes, and satisfy $|\alpha|^2 + |\beta|^2 = 1$. $|0\rangle$ and $|1\rangle$ are the eigenstates, 0 and 1 are the eigenvalues of the two eigenstates. Then, the system and the measurement device can be entangled, which is formalized as:

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

where the detail of entanglement process can be referred to [Von Neumann, 2018]. Next, we strongly measure the pointer of the measuring device. Supposing the pointer collapses to the vector $|x_0\rangle$, the system is new in the state:

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

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 $\Delta = \sigma^2$. If the pointer collapses to a value x_0 around 1, it means that the amplitude to select $|1\rangle$ is a little higher than the amplitude to select $|0\rangle$ and vice versa. So the collapse of the pointer biases the system's vector. However, if σ is very big, the bias will be very small and the outcome system's vector will be very similar to the original vector. The detail analysis is shown in Table 1.

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: *Given a multi-turn conversation between speakers written in English, how to regard the interactions between them, and how to determine their emotional changes brought by interactions?*

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 networks (DM-CNN), which takes the 2-order 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 Convolutional Neural Networks

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 2-order semantic dependencies. Motivated by Zhang's work [Zhang *et al.*, 2018a], we develop a density matrix based convolutional neural networks (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.

$$\begin{aligned} \Pi_i &= |w_i\rangle\langle w_i| \\ &= \begin{pmatrix} w_{i1} \\ w_{i2} \\ ... \\ w_{id} \end{pmatrix} \times (w_{i1}, w_{i2}, ..., w_{id}) \end{aligned} \quad (5)$$

After defining projectors Π_i for each textual word, we represent a document with a density matrix, which can be formulated as:

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

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

We have obtained density matrix ρ_u for temporarily representing the utterance, and ρ_u is then as input into a deep CNN architecture to learn more abstract textual features, i.e., $\vec{x} = [\vec{r}_1, \vec{r}_2, ..., \vec{r}_n]$. The CNN consists of two convolutional layers, a full connected layers and one softmax layer. Each convolutional layer is connected to a max pooling layer. The first convolutional layer has eight 5×5 filters. The second convolutional layer has sixteen 3×3 filters. Note that the textual features $\vec{x}$ will be used as inputs of QIN model.

3.3 A Quantum Measurement-inspired Strong-weak Influence Model

Influence is an indirect, invisible way of altering the thought, behaviour or nature of an entity, which is a difficult task to model. When one talks to other people, he is influenced by their style of interaction. In the conversation, the speaker's affective state might or might not change, depending on the intensity of interaction. If one speaker's affective state changes, we argue that him or her is greatly affected by others. We call this **strong interaction**. Similarly, if one speaker's words have very small influence, and offer no changes in another speaker's affective state. We call this **weak interaction**.

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 thus develop a strong-weak influence model.

Specifically, we introduce and extend the dynamical "influence model", which is a generalization of HMM for describing the influence each Markov chain has on the others through constructing the influence matrices [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 $q_t^e \in \{1, 2, ..., S\}$. Each entity emits an observable o_t^e at time t following the emission probability $b_{q_t^e}(o_t^e) = P(o_t^e|q_t^e)$. 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:

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

where $R(r_t)$ is a $C \times C$ matrix ($R(r_t)_{e,c}$ 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 $R(r_t)$. $Infl\left(q_t^e|q_{t-1}^c\right)$ is modeled using a $S \times S$ matrix $M^{c,e}$, namely $Infl\left(q_t^e|q_{t-1}^c\right) = M^{c,e}_{q_{t-1}^c, q_t^e}$, 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 $M^{c,e}$. The matrix $M^{c,e}$ is very similar to the transition matrix, which can be simplified by two $S \times S$ matrices: E^c and F^c. E^c captures the self-state transition i.e., $E^c = M^{c,c}$, and F^c represents adjacent state transition, i.e., $F^c = M^{c,e}, \forall e \neq c$.

However, in the turn-taking conversation, only the first speaker's state of each turn (time), denoted by $q_t^e|_{e=1}$, is influenced by the previous states of all entities, while the remaining speakers' states of each turn, denoted by $q_t^e|_{e\geq 2}$, 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., $q_{t-1}^e, q_{t-1}^{e+1}, ..., q_{t-1}^C$. Then, the conditional probability is divided into two parts:

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

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

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

where $R(r_t)_{JJ}$, $R(r_t)_{JR}$, $R(r_t)_{RJ}$, $R(r_t)_{RR}$ are four elements of the influence matrix $R(r_t)$, 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:

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

where $p(x) = \left(2\sigma^2\pi\right)^{-\frac{1}{4}} e^{-(x-\mu_{avg})^2/2\sigma^2}$, denoting the probability density to get x, and μ_{avg} is set to the average of all expectations.

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., $R(2)$ and $R(1)$, from one speaker over another speaker under different interactive environments[1].

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

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

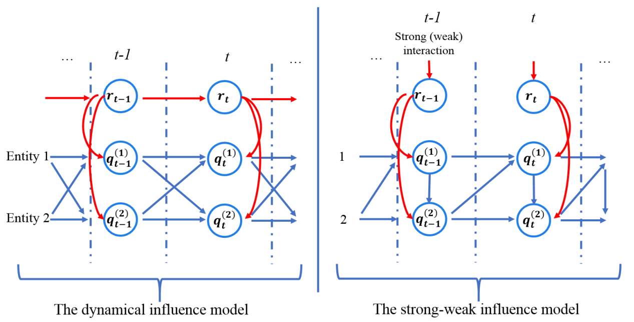


Figure 2: The difference between the dynamical influence model and the strong-weak influence model. The blue lines show the dependence, and the red lines indicate the switching influence.

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, 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 on next speaker from the previous speakers through incorporating the influence scores into the $sigmoid$ function in the QIN, which can be formulated as:

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

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:

$$J = -\frac{1}{N}\sum_i\sum_j y_i^j log\hat{y}_i^j + \lambda_r \|\theta\|^2 \tag{13}$$

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

4 Experiments

4.1 Experimental Settings

Dataset. We conduct experiments on the MELD dataset[2], which contains 13,708 utterances from 1433 dialogues, to validate the effectiveness of QIN model. 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*).

Evaluation metrics. we adopt **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[3] [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.

SVM: We use the bag of words method to generate histograms of word frequencies, and train an SVM classifier to analyze the polarity of each utterance.

CNN: We employ a CNN [Kim, 2014] including two convolutional layer 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: We implement a contextual biLSTM [Poria *et al.*, 2017] to model semantic dependency among the utterances.

4.3 Results and Analysis

Table 2 shows the performance comparison of QIN with other baselines. In the case of sentiment classification, almost all deep neural networks outperform SVM by the margin of 1% to 10%, showing that deep neural networks have potentials in automatically generating representations and can bring performance improvement. LSTM, biLSTM and ATAE-LSTM

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

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

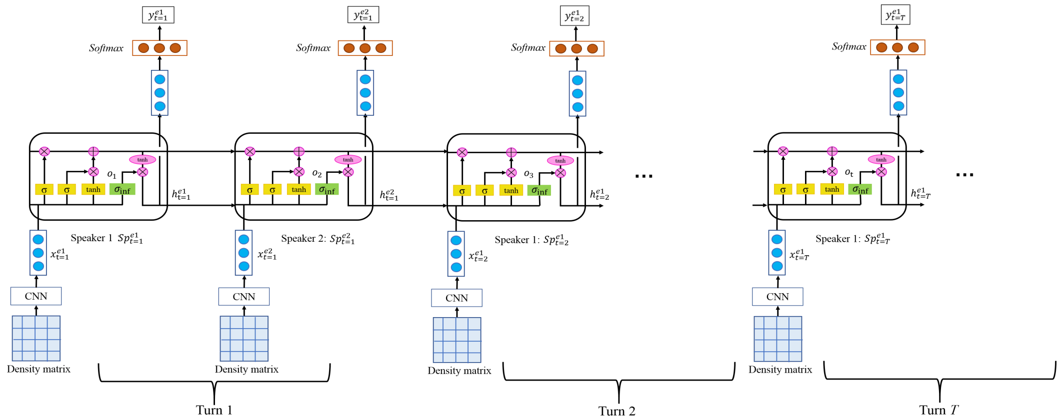


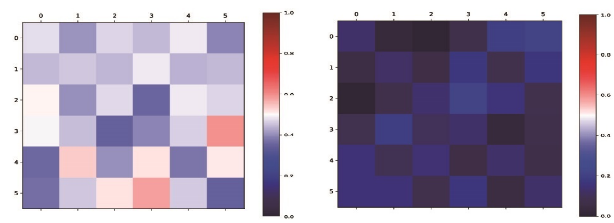
Figure 3: The architecture of quantum-inspired interactive networks.

MELD dataset	Models	Metrics	
		F1	Accuracy
Sentiments (3-class)	SVM	0.623	0.623
	CNN	0.642	0.647
	LSTM	0.626	0.630
	biLSTM	0.611	0.624
	ATAE-LSTM	0.615	0.628
	contextual biLSTM	0.663	0.675
	QIN	**0.681**†	**0.729**†
Emotions (7-class)	SVM	0.566	0.566
	CNN	0.550	0.557
	LSTM	0.541	0.583
	biLSTM	0.537	0.584
	ATAE-LSTM	0.517	0.579
	contextual biLSTM	0.564	0.597
	QIN	**0.637**†	**0.643**†

Table 2: Comparison with baselines. Best performances are in **bold**. The symbol † means statistical improvement.

get worse performance of all neural network baseline methods, because they do not consider the contextual dependencies among utterances. The complete meaning of an utterance may be determined by preceding utterances. Hence, the introduction of attention mechanism does not help improve the performance. CNN stably exceeds LSTM, biLSTM and ATAE-LSTM in the case of sentiment classification while falls behind them in the case of emotion classification. This implies that distinguishing fine-grained emotions might be more dependent on sequence information. 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, all models get poor performance because of the increase in classes. However, QIN still achieves the best performance. Compared with contextual biLSTM, QIN improves the performance about 7.7%. The main reason is that QIN has modelled 2-order semantic dependencies and previous speakers' influence. The results demonstrate the effectiveness and necessity of modelling the interactions in conversational sentiment analysis.



(a) (b)

Figure 4: (a) Strong influence matrix; (b) weak influence matrix. Different colors denote different influences.

5 Visualization of Influence Matrix

Fig. 4 demonstrates a way to present the influence matrices that allow us to observe strong and weak influences. From an overall perspective, we can see that the tones of Figure (a) is pale white color, mixing a hint of light red, while Figure (b) show more black blue zones. This indicates that weak influence matrix has captured weak influences, whose average value is about 0.2. Strong influence matrix capture strong influences, whose average values vary from 0.4 to 0.6.

6 Conclusion

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 demonstrate that our proposed QIN largely outperforms a number of state-of-art sentiment analysis algorithms, and also prove the importance of modeling interactions.

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.

[GLEASON, 1957] ANDREW M. GLEASON. Measures on the Closed Subspaces of a Hilbert Space. *Journal of Mathematics and Mechanics*, 6(6):885–893, 1957.

[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.