A Novel Text Classifier Based on Quantum Computation

Ding Liu, Xiaofang Yang, Minghu Jiang
Laboratory of Computational Linguistics, School of Humanities,
Tsinghua University, Beijing , China
`Dingliu_thu@126.com` `xfyang.thu@gmail.com`
`jiang.mh@mail.tsinghua.edu.cn`

Abstract

In this article, we propose a novel classifier based on quantum computation theory. Different from existing methods, we consider the classification as an evolutionary process of a physical system and build the classifier by using the basic quantum mechanics equation. The performance of the experiments on two datasets indicates feasibility and potentiality of the quantum classifier.

1 Introduction

Taking modern natural science into account, the quantum mechanics theory (QM) is one of the most famous and profound theory which brings a world-shaking revolution for physics. Since QM was born, it has been considered as a significant part of theoretic physics and has shown its power in explaining experimental results. Furthermore, some scientists believe that QM is the final principle of physics even the whole natural science. Thus, more and more researchers have expanded the study of QM in other fields of science, and it has affected almost every aspect of natural science and technology deeply, such as quantum computation.

The principle of quantum computation has also affected a lot of scientific researches in computer science, specifically in computational modeling, cryptography theory as well as information theory. Some researchers have employed the principle and technology of quantum computation to improve the studies on Machine Learning (ML) (Aïmeur et al., 2006; Aïmeur et al., 2007; Chen et al., 2008; Gambs, 2008; Horn and Gottlieb, 2001; Nasios and Bors, 2007), a field which studies theories and constructions of systems that can learn from data, among which classification is a typical task. Thus, we attempted to build a computational model based on quantum computation theory to handle classification tasks in order to prove the feasibility of applying the QM model to machine learning.

In this article, we present a method that considers the classifier as a physical system amenable to QM and treat the entire process of classification as the evolutionary process of a closed quantum system. According to QM, the evolution of quantum system can be described by a unitary operator. Therefore, the primary problem of building a quantum classifier (QC) is to find the correct or optimal unitary operator. We applied classical optimization algorithms to deal with the problem, and the experimental results have confirmed our theory.

The outline of this paper is as follows. First, the basic principle and structure of QC is introduced in section 2. Then, two different experiments are described in section 3. Finally, section 4 concludes with a discussion.

2 Basic principle of quantum classifier

As we mentioned in the introduction, the major principle of quantum classifier (QC) is to consider the classifier as a physical system and the whole process of classification as the evolutionary process of a closed quantum system. Thus, the evolution of the quantum system can be described by a unitary operator (unitary matrix), and the remaining job is to find the correct or optimal unitary operator.

2.1 Architecture of quantum classifier

The architecture and the whole procedure of data processing of QC are illustrated in Figure 1. As is shown, the key aspect of QC is the optimization part where we employ the optimization algorithm to find an optimal unitary operator U'.
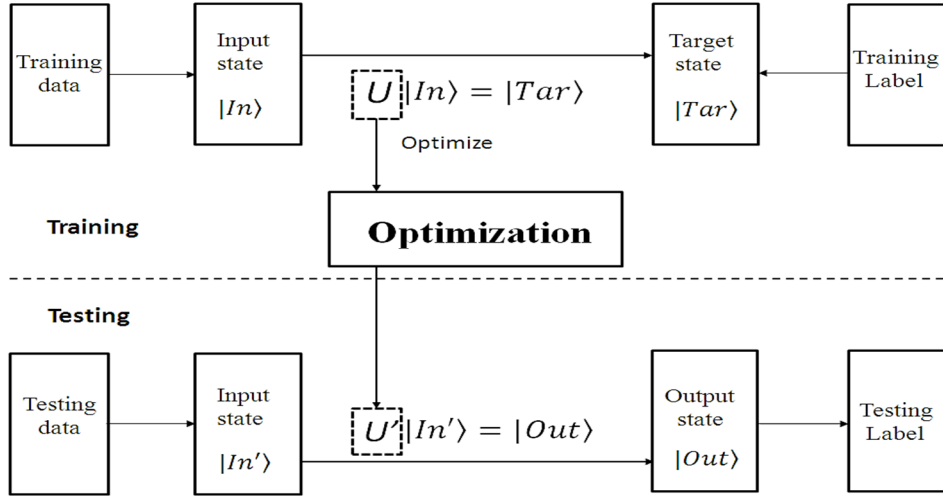
Figure 1. Architecture of quantum classifier

The detailed information about each phase of the process will be explained thoroughly in the following sections.

2.2 Encode input state and target state

In quantum mechanics theory, the state of a physical system can be described as a superposition of the so called eigenstates which are orthogonal. Any state, including the eigenstate, can be represented by a complex number vector. We use Dirac's braket notation to formalize the data as equation 1:

$$|\phi\rangle = \sum_n C_n |E_n\rangle \qquad (1)$$

where $|\phi\rangle$ denotes a state and $C_n \in \mathbb{C}$ is a complex number with $C_n = \langle E_n | \phi \rangle$ being the projection of $|\phi\rangle$ on the eigenstate $|E_n\rangle$. According to quantum theory, C_n denotes the probability amplitude. Furthermore, the probability of $|\phi\rangle$ collapsing on $|E_n\rangle$ is $\mathrm{P}(E_n) = \frac{|C_n|^2}{\sum_n |C_n|^2}$.

Based on the hypothesis that QC can be considered as a quantum system, the input data should be transformed to an available format in quantum theory — the complex number vector. According to Euler's formula, a complex number z can be denoted as $z = re^{i\theta}$ with $r \geq 0, \theta \in \mathbb{R}$. Equation 1, thus, can be written as:

$$|\phi\rangle = \sum_n r_n e^{i\theta_n} |E_n\rangle \qquad (2)$$

where r_n and θ_n denote the module and the phase of the complex coefficient respectively.

For different applications, we employ different approaches to determine the value of r_n and θ_n. Specifically, in our experiment, we assigned the term frequency, a feature frequently used in text classification to r_n , and treated the phase θ_n as a constant, since we found the phase makes little contribution to the classification.

For each data sample $sample_k$, we calculate the corresponding input complex number vector by equation 3, which is illustrated in detail in Figure 2.

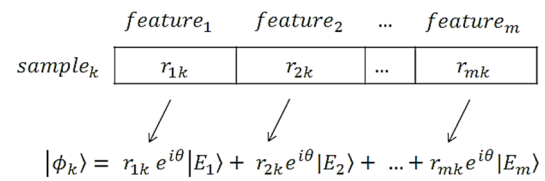$$|\phi_k\rangle = \sum_{j=1}^{m} r_{jk} \cdot e^{i\theta} |E_j\rangle \qquad (3)$$

Figure 2. Process of calculating the input state

Each eigenstate $|E_j\rangle$ denotes the corresponding $feature_j$, resulting in m eigenstates for all the samples.

As is mentioned above, the evolutionary process of a closed physical system can be described by a unitary operator, depicted by a matrix as in equation 4:

$$|\phi'\rangle = U|\phi\rangle \qquad (4)$$

where $|\phi'\rangle$ and $|\phi\rangle$ denote the final state and the initial state respectively. The approach to determine the unitary operator will be discussed in

485

section 2.3. We encode the target state in the similar way. Like the Vector Space Model(VSM), we use a label matrix to represent each class as in Figure 3.

	$Class_1$	$Class_2$	...	$Class_w$
$Label_1$	L_{11}	L_{12}	...	L_{1w}
$Label_2$	L_{21}	L_{22}	...	L_{2w}
...	...	...	...	...
$Label_w$	L_{w1}	L_{w2}	...	L_{ww}

Figure 3. Label matrix

For each input sample $sample_k$, we generate the corresponding target complex number vector according to equation 5:

$$|\varphi_k\rangle = \sum_{j=1}^{n} L_{jk} \cdot e^{i\theta} |E_j\rangle \qquad (5)$$

where each eigenstate $|E_j\rangle$ represents the corresponding $Label_j$, resulting in **w** eigenstates for all the labels. Totally, we need $m + w$ eigenstates, including features and labels.

2.3 Finding the Hamiltonian matrix and the Unitary operator

As is mentioned in the first section, finding a unitary operator to describe the evolutionary process is the vital step in building a QC. As a basic quantum mechanics theory, a unitary operator can be represented by a unitary matrix with the property $U^\dagger = U^{-1}$, and a unitary operator can also be written as equation 6:

$$U = e^{\frac{-iH}{\hbar}t} \qquad (6)$$

where H is the Hamiltonian matrix and $\hbar$ is the reduced Planck constant. Moreover, the Hamiltonian H is a Hermitian matrix with the property $U^\dagger = (U^t)^* = U$. The remaining job, therefore, is to find an optimal Hamiltonian matrix.

Since H is a Hermitian matrix, we only need to determine $(m + w)^2$ free real parameters, provided that the dimension of H is $(m+w)$. Thus, the problem of determining H can be regarded as a classical optimization problem, which can be resolved by various optimization algorithms (Chen and Kudlek, 2001). An error function is defined as equation 7:

$$err(H) = \frac{1}{\sum_{(\phi_t,\phi_i)\in T}|\langle\phi_t^k|\phi_o^k\rangle|} \qquad (7)$$

where T is a set of training pairs with ϕ_t, ϕ_i, and ϕ_o denoting the target, input, and output state respectively, and ϕ_o is determined by ϕ_i as equation 8:

$$|\phi_o\rangle = e^{\frac{-iH}{\hbar}t}|\phi_i\rangle \qquad (8)$$

In the optimization phase, we employed several optimization algorithm, including BFGS, Generic Algorithm, and a multi-objective optimization algorithm SQP (sequential quadratic programming) to optimize the error function. In our experiment, the SQP method performed best outperformed the others.

3 Experiment

We tested the performance of QC on two different datasets. In section 3.1, the Reuters-21578 dataset was used to train a binary QC. We compared the performance of QC with several classical classification methods, including Support Vector Machine (SVM) and K-nearest neighbor (KNN). In section 3.2, we evaluated the performance on multi-class classification using an oral conversation datasets and analyzed the results.

3.1 Reuters-21578

The Reuters dataset we tested contains 3,964 texts belonging to "earnings" category and 8,938 texts belonging to "others" categories. In this classification task, we selected the features by calculating the χ^2 score of each term from the "earnings" category (Manning and Schütze, 2002).

For the convenience of counting, we adopted 3,900 "earnings" documents and 8,900 "others" documents and divided them into two groups: the training pool and the testing sets. Since we focused on the performance of QC trained by small-scale training sets in our experiment, we each selected 1,000 samples from the "earnings" and the "others" category as our training pool and took the rest of the samples (2,900 "earnings" and 7,900 "others" documents) as our testing sets. We randomly selected training samples from the training pool ten times to train QC, SVM, and KNN classifier respectively and then verified the three trained classifiers on the testing sets, the results of which are illustrated in Figure 4. We noted that the QC performed better than both KNN and SVM on small-scale training sets, when the number of training samples is less than 50.
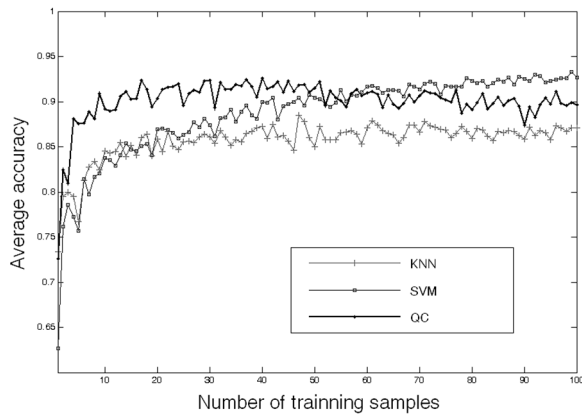
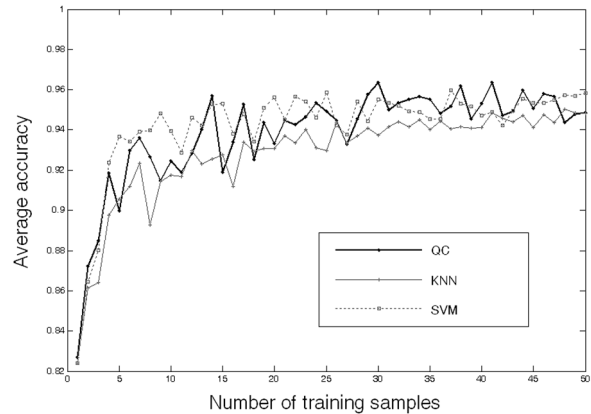Figure 4. Classification accuracy for Reuters-21578 datasets

Figure 5. Classification accuracy for oral conversation datasets

Generally speaking, the QC trained by a large training set may not always has an ideal performance. Whereas some single training sample pair led to a favorable result when we used only one sample from each category to train the QC. Actually, some single samples could lead to an accuracy of more than 90%, while some others may produce an accuracy lower than 30%. Therefore, the most significant factor for QC is the quality of the training samples rather than the quantity.

3.2 Oral conversation datasets

Besides the binary QC, we also built a multi-class version and tested its performance on an oral conversation dataset which was collected by the Laboratory of Computational Linguistics of Tsinghua university. The dataset consisted of 1,000 texts and were categorized into 5 classes, each containing 200 texts. We still took the term frequency as the feature, the dimension of which exceeded 1,000. We, therefore, utilized the primary component analysis (PCA) to reduce the high dimension of the features in order to decrease the computational complexity. In this experiment, we chose the top 10 primary components of the outcome of PCA, which contained nearly 60% information of the original data. Again, we focused on the performance of QC trained by small-scale training sets. We selected 100 samples from each class to construct the training pool and took the rest of the data as the testing sets. Same to the experiment in section 3.1, we randomly selected the training samples from the training pool ten times to train QC, SVM, and KNN classifier respectively and verified the models on the testing sets, the results of which are shown in Figure 5.

4 Discussion

We present here our model of text classification and compare it with SVM and KNN on two datasets. We find that it is feasible to build a supervised learning model based on quantum mechanics theory. Previous studies focus on combining quantum method with existing classification models such as neural network (Chen et al., 2008) and kernel function (Nasios and Bors, 2007) aiming to improve existing models to work faster and more efficiently. Our work, however, focuses on developing a novel method which explores the relationship between machine learning model with physical world, in order to investigate these models by physical rule which describe our universe. Moreover, the QC performs well in text classification compared with SVM and KNN and outperforms them on small-scale training sets. Additionally, the time complexity of QC depends on the optimization algorithm and the amounts of features we adopt. Generally speaking, simulating quantum computing on classical computer always requires more computation resources, and we believe that quantum computer will tackle the difficulty in the forthcoming future. Actually, Google and NASA have launched a quantum computing AI lab this year, and we regard the project as an exciting beginning.

Future studies include: We hope to find a more suitable optimization algorithm for QC and a more reasonable physical explanation towards the "quantum nature" of the QC. We hope our attempt will shed some light upon the application of quantum theory into the field of machine learning.

References

Esma Aïmeur, Gilles Brassard, and Sébastien Gambs. 2006. Machine Learning in a Quantum World. *Canadian AI 2006*

Esma Aïmeur, Gilles Brassard and Sébastien Gambs. 2007. Quantum Clustering Algorithms. *Proceedings of the 24 th International Conference on Machine Learning*

Joseph C.H. Chen and Manfred Kudlek. 2001. Duality of Syntex and Semantics – From the View Point of Brain as a Quantum Computer. *Proceedings of Recent Advances in NLP*

Joseph C.H. Chen. 2001. Quantum Computation and Natural Language Processing. University of Hamburg, Germany. Ph.D. thesis

Joseph C.H. Chen. 2001. A Quantum Mechanical Approach to Cognition and Representation. *Consciousness and its Place in Nature,Toward a Science of Consciousness.*

Cheng-Hung Chen, Cheng-Jian Lin and Chin-Teng Lin. 2008. An efficient quantum neuro-fuzzy classifier based on fuzzy entropy and compensatory operation. *Soft Comput,* 12:567–583.

Fumiyo Fukumoto and Yoshimi Suzuki. 2002. Manipulating Large Corpora for Text Classification. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*

Sébastien Gambs. 2008. Quantum classification, arXiv:0809.0444

Lov K. Grover. 1997. Quantum Mechanics Helps in Searching for a Needle in a Haystack. *Physical Re view Letters, 79,325–328*

David Horn and Assaf Gottlieb. 2001. The Method of Quantum Clustering. *Proceedings of Advances in Neural Information Processing Systems .*

Christopher D. Manning and Hinrich Schütze. 2002. Foundations of Statistical Natural Language Processing. *MIT Press. Cambridge, Massachusetts,USA.*

Nikolaos Nasios and Adrian G. Bors. 2007. Kernel-based classification using quantum mechanics. *Pattern Recognition,* 40:875–889

Hartmut Neven and Vasil S. Denchev. 2009. Training a Large Scale Classifier with the Quantum Adiabatic Algorithm. arXiv:0912.0779v1

Michael A. Nielsen and Isasc L. Chuang. 2000. Quantum Computation and Quantum Information, *Cambridge University Press, Cambridge, UK.*

Masahide Sasaki and and Alberto Carlini. 2002. Quantum learning and universal quantum matching machine. *Physical Review,* A 66, 022303

Dan Ventura. 2002. Pattern classification using a quantum system. Proceedings of the Joint Conference on Information Sciences.