10.12871/clicit2014222

Emotion Recognition with a Kernel Quantum Classifier

Fabio Tamburini
FICLIT - University of Bologna, Italy
fabio.tamburini@unibo.it

Abstract

English. This paper presents the application of a Kernel Quantum Classifier, a new general-purpose classifier based on quantum probability theory, in the domain of emotion recognition. It participates to the EVALITA 2014 Emotion Recognition Challenge exhibiting relatively good results and ranking at the first place in the challenge.

Italiano. *Questo contributo presenta l'applicazione di un classificatore quantistico basato su kernel, un nuovo classificatore basato sulla teoria della probabilità quantistica, nel dominio del riconoscimento delle emozioni. Ha partecipato alla campagna di valutazione sul riconoscimento delle emozioni nell'ambito di EVALITA 2014 ottenendo buoni risultati e classificandosi al primo posto.*

1 Introduction

Quantum Mechanics Theory (QMT) is one of the most successful theory in modern science. Despite its ability to properly describe most natural phenomena in the physics realm, the attempts to prove its effectiveness in other domains remain quite limited.

This paper presents the application of a Kernel Quantum Classifier, a new general-purpose classifier based on quantum probability theory, in the domain of emotion recognition.

With regard to this specific evaluation challenge, we did not develop any particular technique tailored to emotion recognition, but we applied a "brute force" approach to this problem as described, for example, in (Schuller *et al.*, 2009). A very large set of general acoustic features has been automatically extracted from speech waveforms and the emotion detection task has been put totally in charge of the classifier.

In section 2 we will describe the proposed classifier, in section 3 the evaluation results will be analysed comparing them with the results obtained using a state-of-the-art classifier applied to the same task and in section 4 we will draw some provisional conclusions.

2 System description

2.1 Quantum Probability Theory

A *quantum state* denotes an unobservable distribution which gives rise to various observable physical quantities (Yeang, 2010). Mathematically it is a vector in a complex Hilbert space. It can be written in Dirac notation as $|\psi\rangle = \sum_1^n \lambda_j |e_j\rangle$where λ_j are complex numbers and the $|e_j\rangle$ are the basis of the Hilbert space ($|.\rangle$ is a column vector, or a *ket*, while $\langle .|$ is a row vector, or a *bra*). Using this notation the inner product between two state vectors can be expressed as $\langle\psi|\phi\rangle$ and the outer product as $|\psi\rangle\langle\phi|$.

$|\psi\rangle$ is not directly observable but can be probed through measurements. The probability of observing the elementary event $|e_j\rangle$ is $|\langle e_j|\psi\rangle|^2 = |\lambda_j|^2$ and the probability of $|\psi\rangle$ collapsing on $|e_j\rangle$ is $P(e_j) = |\lambda_j|^2 / \sum_1^n |\lambda_i|^2$ (note that $\sum_1^n |\lambda_i|^2 = \||\psi\rangle\|^2$ where $\|\cdot\|$ is the vector norm). General events are subspaces of the Hilbert space.

A matrix can be defined as a *unitary operator* if and only if $UU^\dagger = I = U^\dagger U$, where † indicates the Hermitian conjugate. In quantum probability theory unitary operators can be used to evolve a quantum system or to change the state/space basis: $|\psi'\rangle = U|\psi\rangle$.

Quantum probability theory (see (Vedral, 2007) for a complete introduction) extends standard kolmogorovian probability theory and it is in principle adaptable to any discipline.

2.2 Kernel Quantum Classifier

(Liu *et al.*, 2013) presented a quantum classifier based on the early work of (Chen, 2002). Given an Hilbert space of dimension $n = n_i + n_o$, where n_i is the number of input features and n_o is the number of output classes, they use a unitary operator U to project the input state contained in the subspace spanned by the first n_i basis vectors into an output state contained in the subspace spanned by the last n_o basis vectors: $|\psi^o\rangle = U\,|\psi^i\rangle$. Input, $|\psi^i\rangle$, and output, $|\psi^o\rangle$, states are real vectors, the former having only the first n_i components different from 0 (assigned to the problem input features of every instance) and the latter only the last n_o components. From $|\psi^o\rangle$ they compute the probability of each class as
$P(c_j) = |\psi^o_{ni+j}|^2 / \sum_1^{no} |\psi^o_{ni+i}|^2$ for $j = 1..n_o$.

The unitary operator U for performing instances classification can be obtained by minimising the loss function

$$err(T) = 1/\sum_{j=1}^{|T|} \langle\psi^o_j|\psi^t_j\rangle\,,$$

where T is the training set and $|\psi^t\rangle$ is the target vector for output probabilities (all zeros except 1 for the target class) for every instance k, using standard optimisation techniques such as Conjugate Gradient (Hestenes, Stiefel, 1952), L-BFGS (Liu, Nocedal, 1989) or ASA (Ingber, 1989).

This classifier exhibits interesting properties managing a classical non-linear problem, the XOR problem, but the simplicity and the low power of this classifier emerge quite clearly when we test it on difficult, though linearly separable, classification problems or on non-linear problems. The classifier is not always able to properly divide the input space into different regions corresponding to the required classes. Moreover, all the decision boundaries have to cross the origin of the feature space, a very limiting constraint for general classification problems, and problems that require strict non-linear decision boundaries cannot be successfully handled by this classifier.

A widely used technique to transform a linear classifier into a non-linear one involves the use of the "kernel trick". A non-linearly separable problem in the input space can be mapped to a higher-dimensional space where the decision borders between classes might be linear. We can do that through the mapping function $\phi : \mathbb{R}^n \to \mathbb{R}^m$, with $m > n$, that maps an input state vector $|\psi^i\rangle$ to a new space. The interesting thing is that in the new space, for some particular mappings, the inner product can be calculated by using *kernel* functions $k(x, y) = \langle\phi(x), \phi(y)\rangle$ without explicitly computing the mapping ϕ of the two original vectors.

We can express the unitary operator performing the classification process as a combination of the training input vectors in the new features space

$$\begin{aligned}
|\psi^o\rangle &= U\ \ |\phi(\psi^i)\rangle \\
|\psi^o\rangle &= \textstyle\sum_{j=1}^{|T|} |\alpha_j\rangle\ \langle\phi(\psi^i_j)|\ \ |\phi(\psi^i)\rangle \\
|\psi^o\rangle &= \textstyle\sum_{j=1}^{|T|} |\alpha_j\rangle\ \langle\phi(\psi^i_j)|\phi(\psi^i)\rangle
\end{aligned}$$

that can be rewritten using the kernel and adding a bias term $|\alpha_0\rangle$as:

$$|\psi^o\rangle = |\alpha_0\rangle + \sum_{j=1}^{|T|} |\alpha_j\rangle\, k(\psi^i_j, \psi^i) \qquad (1)$$

In this new formulation we have to obtain all the $|\alpha_j\rangle$ vectors, $j = 0, ..., |T|$, through an optimisation process similar to the one of the previous case, minimising a standard euclidean loss function

$$\begin{aligned}
err(T) = &\sum_{j=1}^{|T|} \sum_{k=1}^{no} \Big(P_j(c_k) - \psi^t_{j(ni+k)}\Big)^2 \\
&+ \gamma \sum_{j=0}^{|T|} \| |\alpha_j\rangle \|.
\end{aligned}$$

using a numerical optimisation algorithm, L-BFGS in our experiments, where $P(c)$ is the class probability defined above and $\gamma \sum \| |\alpha_j\rangle \|$ is an L_2 regularisation term on model parameters (the real and imaginary parts of $|\alpha_j\rangle$ components).

Once learned a good model from the training set T, represented by the $|\alpha_j\rangle$ vectors, we can use equation (1) and the definition of class probability for classifying new instance vectors.

It is worth noting that the KQC proposed here involves a large number of variables during the optimisation process (namely, $2 * no * (|T| + 1)$) that depends linearly on the number of instances in the training set T. In order to build a classifier applicable to real problems, we have to introduce special techniques to efficiently compute the gradient needed by optimisation methods. We relied on Automatic Differentiation (Griewank, Walther, 2008), avoiding any gradient approximation using finite differences that would require a very large number of error function evaluations. Using such

	Automatic System					
Gold Std.	ang	dis	fea	joy	sad	sur
ang	12	9	1	0	1	7
dis	0	11	3	0	5	2
fea	2	4	5	3	15	1
joy	9	8	1	5	1	6
sad	0	2	0	1	26	1
sur	2	1	1	1	19	6

Table 1: Confusion matrix between the gold standard and the KQC.

	Automatic System					
Gold Std.	ang	dis	fea	joy	sad	sur
ang	16	1	1	2	2	8
dis	6	8	7	0	5	4
fea	3	0	6	4	15	2
joy	10	6	4	7	0	3
sad	0	3	1	1	24	1
sur	2	2	1	1	19	5

Table 2: Confusion matrix between the gold standard and the SVM multiclass classifier proposed in (Joachims *et al.*, 2009).

techniques the training times of KQC are comparable to those of other machine learning methods.

Please, see (Tamburini, in press) for a complete presentation and evaluation of this system.

3 EVALITA 2014 ERT results

We applied the KQC to the EVALITA 2014 Emotion Recognition Task without adapting the system in any way and without devising any specific technique for emotion detection. We participated only at the "closed database" subtask that is devoted to evaluate how much information can be extracted from material coming from a single, professional source of information whose explicit task is to portray emotions and obtain models capable of generalizing to unseen subjects.

As we said in the introduction, we applied a "brute force" approach to this problem: we extracted 1582 features from each utterance using the OpenSMILE package (Eyben *et al.*, 2013) and the configuration file contained in the package for extracting the InterSpeech 2010 Paralinguistic Challenge feature set (Schuller *et al.*, 2010).

In this case $ni = 1582$ and $no = 6$; we excluded from the process all the utterances belonging to the "neutral" class following the task guidelines indications. After a training session using all the utterances and classifications in the Development Set provided by the organisation, we tested the trained classifier on the Test Set executing ten different runs. The outputs of the ten classification processes were mixed and the final results submitted for the evaluation contained the most frequent class chosen by the ten runs for each utterance contained in the Test Set.

The official results assigned the first place to this classifier with a classification accuracy of 36.11%. Table 1 outline the confusion matrix between classes.

We performed some other experiments using a different classifier: the standard Support Vector Machine (SVM) multiclass classifier proposed in (Joachims *et al.*, 2009). This widely diffused state-of-the-art classifier exhibit more or less the same performances of the KQC: 36.67% of accuracy in classifying the six emotions considered in the EVALITA 2014 ERT challenge (the best results are obtained by using a linear kernel and $C = 30$). Table 2 shows the confusion matrix for the SVM multiclass classifier.

4 Discussion and Conclusions

Even if a 36.11% of accuracy allowed this system to be the most accurate in the evaluation campaign (out of two participants), such accuracy is very low; it is much better than the random baseline (16.67%), but certainly not enough for real classification problems. Some emotions, anger, disgust and sadness, can be detected with better reliability, but the other emotions, namely fear, joy and surprise, present classification results very unsatisfactory. The experiments conducted with a different but state-of-the-art classifier, namely a SVM multiclass classifier, present more or less the same picture.

The research question posed in the guidelines "to establish how much information can be extracted from material coming from a single, professional source of information whose explicit task is to portray emotions and obtain models capable of generalizing to unseen subjects" cannot be answered, in our opinion, positively. Emotional recordings taken from a single, even professional, speaker, do not seem to provide enough information to generalise the emotion recognition to other speakers.

Despite the design of KQC is a work in progress and the it is not free from problems, it exhibits good classification performances, very similar to a state-of-the-art multiclass classifier.

References

Chen J.C.H. 2002. *Quantum Computation and Natural language Processing*. PhD thesis, University of Hamburg.

Eyben F., Weninger F., Gross F. and Schuller B. 2013. Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. *ACM Multimedia (MM)*, Barcelona, 835–838.

Griewank A. and Walther A. 2008. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. Other Titles in Applied Mathematics 105 (2nd ed.), SIAM.

Hestenes M.R. and Stiefel E. 1952. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49 (6), 409–436.

Ingber L. 1989. Very fast simulated re-annealing. *Mathl. Comput. Modelling*, 12 (8): 967–973.

Joachims T., Finley T. and Yu C-N. 2009. Cutting-Plane Training of Structural SVMs. *Machine Learning Journal*, 77 (1): 27–59.

Liu D.C. and Nocedal J. 1989. On the Limited Memory Method for Large Scale Optimization. *Mathematical Programming B*, 45 (3): 503–528.

Liu D., Yang X and Jiang M. 2013. A Novel Text Classifier Based on Quantum Computation. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, 484–488.

Schuller B., Steidl S., Batliner A. 2009. The INTERSPEECH 2009 Emotion Challenge. *Proceedings of Interspeech 2009*, Brighton, 312–315.

Schuller B., Steidl S., Batliner A., Burkhardt F., Devillers L., Muller C. and Narayanan S. 2010. The INTERSPEECH 2010 Paralinguistic Challenge. *Proceedings of Interspeech 2010*, Makuhari, Japan, 2794–2797.

Tamburini F. in press. Are Quantum Classifiers Promising? *Proceedings of The first Italian Computational Linguistics Conference - CLiC*, Pisa.

Vedral V. 2007. *Introduction to Quantum Information Science*. Oxford University Press, USA.

Yeang C.H. 2010. A probabilistic graphical model of quantum systems. *Proceedings, the 9th International Conference on Machine Learning and Applications (ICMLA)*, Washington DC, 155–162.