Journal of Computational Information Systems 8: 21 (2012) 9105–9112
Available at http://www.Jofcis.com

# An Algorithm for Acupuncture Clinical Assessment Based on Multi-view KNN

Zhaohui LIANG [1], Gang ZHANG [2], Guozheng LI [3], Wenbin FU [1,*]

[1] *The 2nd Affiliated Hospital, Guangzhou University of Chinese Medicine, Guangzhou 510120, China*

[2] *School of Automation, Guangdong University of Technology, Guangzhou 510006, China*

[3] *Department of Control Science and Engineering, Tongji University, Shanghai 201804, China*

**Abstract**

Patient-reported outcome (PRO) technique is frequently used in clinical assessment of acupuncture, especially in the evaluation of subjective symptoms in chronic diseases. Since different PRO tools stress various perspectives of patients status, the natural inconsistency of the outcomes brings contradiction to summarize an overall outcome. In order to minimize such contradiction, and to determine the personalized optimized therapy protocol for individual patient, we propose a machine learning framework as complement for doctors profession and experience. The framework makes use of historic therapy records for evaluation. We apply K-Nearest-Neighbor algorithm to several totally separated attributes subset, to determine the most similar cases for current patient. The proposed algorithm is immune to inconsistency and yields a constructive result. An experiment was conducted based on the clinical data of a multi-center RCT shows that the proposed algorithm is effective to clinical outcome assessment.

*Keywords*: K Nearest Neighbor; Multi-view; Patient-reported Outcome; Outcome Inconsistency

## 1 Introduction

Acupuncture is nowadays widely accepted and practiced in China and western countries. It was reported that 43% of primary care repliers in the US used acupuncture in practice [1], but high quality evidence originated from randomized control trials (RCT) with rigid design, big sample size, multiple factor controlled is in need [2, 3] to confirm its efficacy. During the latest decade, new researches based on the past experience with better or innovative design are emerging and providing new and high quality evidence for the efficacy of acupuncture. [4] In practice, acupuncture is mainly administered to chronic diseases and disorders such as stroke complications, headache, menstrual problems, asthma and pain [2]. As a result, the efficacy of acupuncture cannot be assessed by some objective parameters like death rate, blood pressure, blood sugar, etc. Alternatively, subjectively assessing, e.g. the application of patient-reported

*Corresponding author.
*Email address:* fuwenbin@139.com (Wenbin FU).


November 1, 2012

outcome (PRO), self-filling questionnaires and scales is common in the clinical evaluation for acupuncture efficacy.

In the study on neck pain, the effect of acupuncture was measured by three subjective rating questionnaires, i.e. Northwick Pack Neck Pain Questionnaire (NPQ) for neck pain intensity and loss of quality of life (QoL) and physical functions [5], McGill Pain Questionnaire (MPQ) for general pain intensity [6], and Medical Outcomes Study Short Form 36 (SF-36) for health related quality of life (HRQOL) [7].

This work focuses on that the above PRO tools (i.e. NPQ, MPQ and SF-36) actually reflect different domains of patients' condition. The data of a multi-center randomized controlled trail (RCT) was applied to our experiment. The efficacy was evaluated by the outcomes evaluated in different check points before, in the middle, after the treatment, and during the follow-up after the treatment.

The goal is to evaluate score changes of the concerned measure outcomes by basic information and the scores before treatment. Let $P$ be a set of patients, $p_i \in P$ and $p_i = (A, Q)$ where $A$ is a column vector describing basic information of the patient, and $Q$ is real vector recording scores before treatment. Let $Y$ be a set of series measure outcomes. We aim at finding a function $f : P \to Y$ of minimal loss on the historic data set. Since there is not explicit form of $f$, we achieve the target by a local learning method, i.e. obtains evaluation result directly from historic data. We propose a multi-view K-nearest-neighbors algorithm to construct proper reference set for evaluation. Intuitively, the original records are split into two disjoint views, among which the view of patient information provides natural categories of patients, while the view of pretreatment scores provides the feature of neck pain of an individual patient. Intersection of nearest records from both views yields a robust reference set for evaluation, which is the intuition of this paper.

The rest of the paper is organized as follows. Related work (Section 2) reviews some important work related to our work. Multi-view KNN (Section 3) presents the main algorithm. Evaluation (Section 4) reports experiment result and finally Conclusion (Section 5) concludes this paper.

## 2 Related Work

Previous study on PRO outcomes indicated for evaluation local learning methods work well with properly defined metrics. Zhaohui Liang et al. [9] proposed a similarity based learning framework to evaluate improvement of concerned measurement outcomes by constructing kernel based similarity between current therapy sample and an outstanding records set from historic data. The outstanding historic records are determined by a multi-metric sorting mechanism with changing ratios of several concerned measurement outcomes. They obtained an acceptable accuracy upon a mid-scale clinical dataset. However, the potential importance of outstanding records to current sample is to be further revealed.

Zhang Gang et al. [10] proposed a local learning method in evaluating almost the same task as that in [9], while they adopt pure local learning method. To find historic data records located in a region concentric of a given sample, they adopt k-nearest-neighbour (KNN) algorithm with Euclidean distance to find local records, with which a standard decision tree learner is trained to evaluate only the given sample. However, their work treated patient information and the concerned measurement outcomes equally in Euclidean distance function, leading to improper determination of local records. We argue that bias and scaling of different groups of properties in a record should be considered separately.

Motivated by previous work, we propose a multi-view k-nearest-neighbour evaluation method, which separately treats properties from patient information and concerned measurement outcomes before combining them by intersection. Our method avoids weighting of different groups of properties, which is of high computational cost.

# 3 Multi-view KNN

Let $D$ be a set of historic data records, $D = (A_1, A_2, \cdots, A_r, T)$, $A_k \subseteq A$ $(k = 1, 2, \cdots, r)$, where $A$ is the universal of all concerned attributes and $T$ is the concerned target. In our study, we assume that $\forall i, j$, $A_i \cap A_j = \phi$. We call $A_k(k = 1, 2, \cdots, r)$ a view of data records. The goal of our study is to evaluate $T$ for a given test sample by local records in a reference set. The locality is determined by combination of individual locality of each concerned view. Algorithm 1 illustrates the main steps of multi-view KNN.

**Algorithm 1 Multi-View KNN**

Input: multi-view training data set $D_{train} = (A_1, A_2, \ldots, A_r, T_{train})$
test data set $D_{test} = (A_1, A_2, \ldots, A_r)$
number of nearest neighbors $k$

Output: target set $T_{test}$ of test data set

1: for i = 1 to r
2: $Ref_i = knnsearch\ (D_{train}(A_1), D_{test}, k)$
3: end
4: $Reference = Ref_1$
5: for i = 2 to r
6: $Reference = Reference \cap Ref_i$
7: end
8: if *Reference* is null then exit
9: $T_{test}$ = Majority Voting for Field $T$ in *Reference*
10: return $T_{test}$

Algorithm 1 calculates reference sets through different views of training data set. Line 2 adopts procedure knnsearch to find $k$ nearest neighbors for each test sample in $D_{test}$ according to different views of $D_{train}$. In line 6 an intersection is performed in all reference sets. Note that this would lead to an empty reference set if $k$ is set to a very small value and the concerned views are too much. In our study, we empirically select a proper $k$ to obtain a reference set of acceptable size. In line 9, a majority voting procedure is applied to get a predict value through reference set. With such setting, Algorithm1 supports only single value output, but also structure output the same as target values of reference set.

# 4 Evaluation

## 4.1 Data set

The data set of this study originated from a multi-center RCT to assess the efficacy of acupuncture for cervical spondylosis funded by the Ministry of Science and Technology. Totally there are records in the dataset. We remove 112 records for the lack of MPQ values. The data set is attributed as patients' demographical information, outcome of the PRO questionnaires, i.e. NPQ, MPQ, and SF-36.

## 4.2 Therapeutic effect evaluation

To evaluate the therapeutic effect of the acupuncture treatment, we test the proposed algorithm on a two-view subjective score evaluation problem. The whole data set is randomly divided into two disjoint parts for either training or test with ratio $r$. The first view is patient basic information and the second is scores before treatment. The proposed algorithm evaluates scores after treatment and we apply $L_1$ loss to calculate difference between the algorithm output and actual scores as Eq. (2).

$$Loss(Y, Y^*) = \frac{1}{d} \sum_{i=1}^{d} |y_i - y_i^*| \tag{1}$$

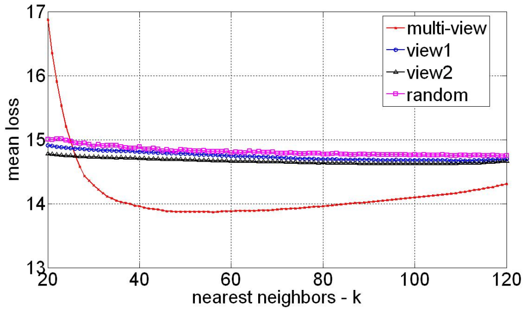
where $d$ is dimensionality of $Y$ and $Y^*$, $y_i \in Y$ and $y_i^* \in Y^*$ are ingredients of concerned target $Y$, and $Y^*$ is true value of concerned target. When there are more than one concerned targets, we record their mean losses calculated by Eq. (2).

In this experiment setting, we evaluate the scores of check points after the pre-treatment questionnaires. The training data ratio $r$ is set to 15%, 30% and 45% respectively. To show the effect of the proposed algorithm at different sizes of training set, at each ratio training samples are randomly selected 10 times and the mean loss is recorded. Fig. 1 shows the comparison result to three baseline methods.

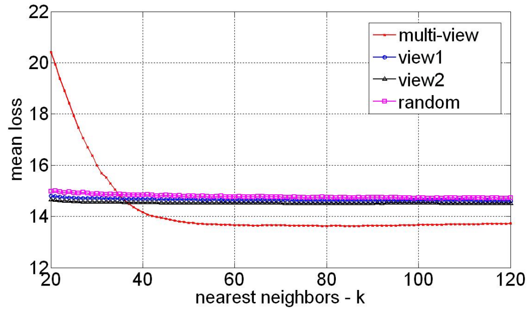
Fig. 1 shows the significant improvement of the proposed algorithm compared three baseline methods. Method $view1$ and $view2$ stand for determining the reference set merely by searching for the nearest neighbors through patient information and the scores of NPQ, MPQ and SF-36 before treatment, respectively. And random stands for the method randomly selected $k$ samples from training set as reference samples. In three different training ratio, method random is of largest loss among all compared methods. Methods $view1$ and $view2$ get close loss and $view2$ is a little better than $view1$. This is mainly because $view2$ considers only subjective scores, which is statistically superior. Method multi-view stands for the proposed algorithm of this paper, which has great loss improvement compared to three baseline methods.

## 4.3 Clinical outcome evaluation based on $OPROO$

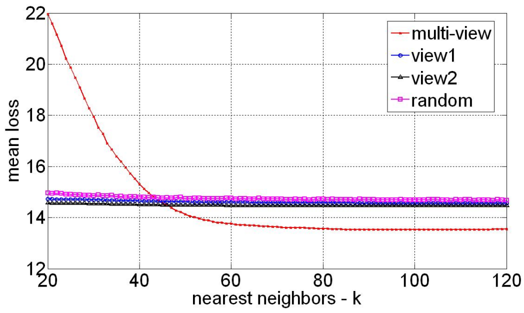
To further show the effectiveness of the proposed algorithm, we apply it to a previous proposed evaluation method that effectively evaluates the overall clinical outcome based on $OPROO$ [10]. Since the definition in [10] is not normalized, we extend the above definition to a normalization version as Definition 1.




(a) $r = 15\%$




(b) $r = 30\%$




(c) $r = 45\%$

Fig. 1: Comparison between multi-view KNN and baseline methods in different training ratio

**Definition 1 (Normalized** $OPROO$**)** *Suppose there are $k$ check points of a subjective measure $S$, we have the Overall PRO Outcome or $OPROO$ as:*

$$OPROO_S = e^{-\sum_{i=1}^{k-1} \alpha (S_{i+1} - S_i)/S_i} \tag{2}$$

Liang and Zhang [10] proposed a local learning algorithm to evaluate therapy effectiveness based on $OPROO$. We will show that multi-view KNN algorithm is also suitable for effectiveness evaluation based on $OPROO$. Fig.2 sheds light on the connection between $OPROO$ and the proposed algorithm.

Following [9]'s experiment setting, we conduct an experiment to show that the proposed algorithm is also suitable for $OPROO$ based therapy effectiveness evaluation. For the whole data set, $OPROO$ for NPQ, MPQ and SF-36 of each sample is calculated and we apply a multi-metric sorting algorithm to get a consistent ranking. Then the top 65% samples according to the ranking are labeled as effective and the rest samples are labeled as ineffective. After that we get a binary classification data set with labels indicating whether the improvement of $OPROO$ reach a pre-defined significant level. The 65% for effective samples is from clinical experience.

The proposed algorithm outputs $k$ nearest neighbors of the query sample in the training set. We apply majority voting to determine the label of query sample. As mentioned previously, two views are considered and we also implement the methods proposed in [10] for comparison. Fig.3 illustrates the result.

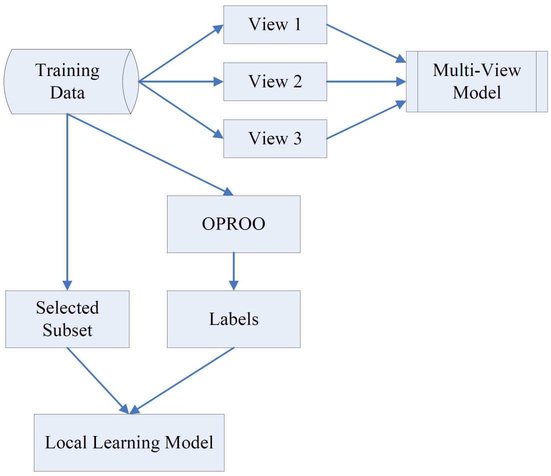



Fig. 2: Connection between multi-view KNN and local learning model

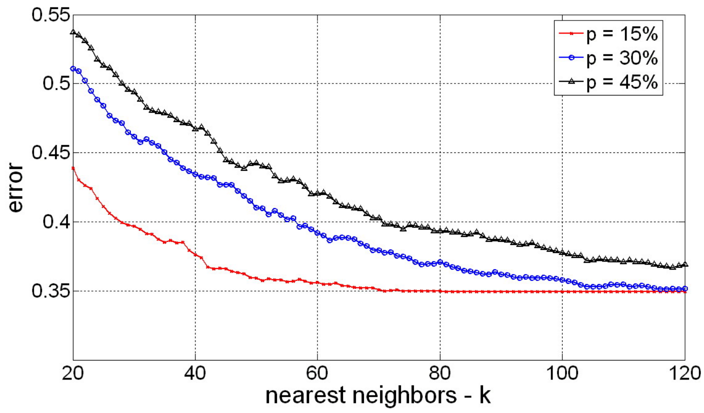



Fig. 3: Multi-view KNN for OPROO based clinical outcome evaluation (with view of patient information)

It is a little strange that small training ratio leads to relative good result, as illustrated in Fig.3. We owe this phenomenon to side effect of majority voting of nearest neighbors. Based on [10]'s analysis, pre-treatment and midterm status has more significant impact than that of patient basic information. Thus we apply a learning model more powerful than majority voting to evaluate the therapy effect with respect to *OPROO*. In doing so, we apply a WEKA implementation decision tree fed with multi-view KNN output as training data set. The difference between majority voting and decision tree is that the latter potentially considers importance of each training data to the target based on their contribution measured by entropy [9]. Fig.4 illustrates the evaluation error of the same experiment setting as previous section, except the majority voting for final evaluation.

From Fig.4 it can be seen that there is significant improvement compared to Fig.3 which took patient information into consideration. It can be concluded that patient information doesn't always work in therapy effectiveness evaluation, since in our algorithm setting an intersection between different views is applied to determine the reference set. Another reason may be laid in that the definition of *OPROO* isn't exactly consistent with underlying principles as patient basic information. We would give a further discussion in the next section. Also note that in both Fig.3 and Fig.4, the evaluation error rates of $k$'s values between 20 and 40 are very high.

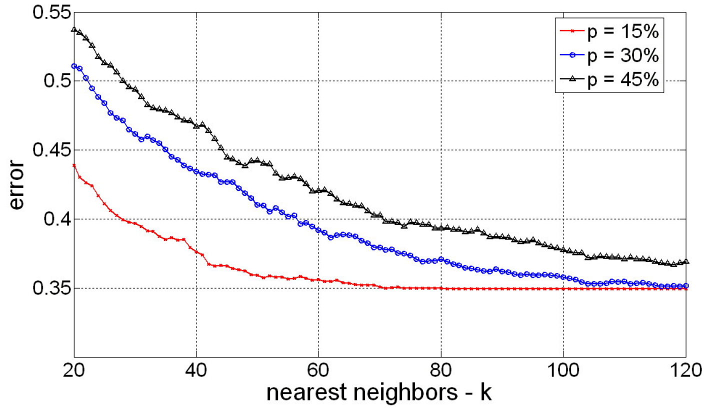



Fig. 4: Multi-view KNN for OPROO based clinical outcome evaluation (without view of patient information)

This is mainly because in this range there are often no reference instances can be found due to intersection between different views.

# 5 Conclusion

KNN plays an important role in local learning. However, traditional KNN relies heavily on distance function defined in input space. Instead of learning a problem specific distance function, we propose to ensemble KNN learners from different attribute subsets in this paper. We believe that in a view (i.e. attributes reflecting some aspect of samples) traditional KNN works well with simple Euclidean distance function. In our task, there are three PRO measurements which can be potentially regarded as different views of therapy effectiveness. Previous work [8, 9] indicated that when a weight vector is attached to these PROs, therapy effectiveness evaluation accuracy has been observed improved. In another word, there should be some bias among these PROs. However, it is time-consuming to work out the optimal weights given several PROs. The proposed method does not require any explicit expressing weights. Also we can integrate weights into our framework by adopting weighted ensemble. A weight vector is learned through a generic algorithm (GA) at the minimization of both model empirical loss and model complexity. The proposed algorithm determined weights of each view instead of each attribute, which would reduce the complexity of the whole problem.

# Acknowledgement

The work is supported by the Eleventh Five-year Project of MST of China (No. 2006BAI12B04-1), the National 973 Program (No. 2010CB530500 and No. 2010CB530503), the Scientific Projects from Guangdong Provincial Administration of Science and Technology (No. 2006B50107006, No. 2008B030301206, and No. 2011B080701036), research grant of Guangdong Provincial Administration of Health (No. C2009025), the Open Research Fund of Zhejiang First-foremost Key Subject-Acupuncture & Moxibustion (No. ZTK2010B04), Research grant of Guangdong Administration of Chinese Medicine (No. 2010144), Creative Foundation of Guangzhou University of

Chinese Medicine (No.10CX24), 2010 Educational Science Research Project of the National 11th Five-Year Program of Guangdong Province (2010tjk465), GuangDong College Student Creative Experiment Project (1184510037) and 2011 Student Outside-Class Technology Project of Guang-Dong University of Technology (No. 153). The clinical data in this paper originates from the trial registered on Chinese Cochrane Center: ChiCTR-TRC-00000184.

# References

[1] K. V. Mardia, J. T. Kent, J. M. Bibby, Multivariate Analysis, Academic Press, London, 1979.

[2] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, R. S. J. Frackowiak, Statistical parametric maps in functional imaging: a general linear approach, Human Brain Mapping 2 (1995), 189-210.

[3] D. G. Hoy, M. Protani, R. De, R. Buchbinder, The epidemiology of neck pain, Best. Pract. Res. Clin. Rheumatol. 24 (2010), 783-792.

[4] W. B. Fu, Z. H. Liang, X. P. Zhu, P. Yu, J. F. Zhang, Analysis on the effect of acupuncture in treating cervical spondylosis with different syndrome types, Chin. J. Integr. Med. 6 (2009), 426-430.

[5] Z. H. Liang, X. P. Zhu, X. B. Yang W. B. Fu, A. P. Lu, Assessment of a traditional acupuncturetherapy for chronic neck pain: a pilot randomized controlledstudy, Complement. Ther. Med. S (2011), S26-S32.

[6] X. Y. Wang, S. H. Yuan, H. Y. Yang, Y. M. Sun, F. P. Cheng, C. L. Zhang, X. C. Huang, Abdominal acupuncture for insomnia in women: a randomized controlled clinical trial, Acupunct. Electrother. Res. 1-2 (2008), 33-41.

[7] Y. Huang, X. M. Liao, X. X. Li, Y. B. Song, Clinical observation on the effects of Bo's abdominal acupuncture in 40 cases of chronic fatigue syndrome, J. Tradit. Chin. Med. 4 (2008), 264-266.

[8] C. R. Meng, L. Fan, W. B. Fu, Y. Li, Clinical research on abdominal acupuncture plus conventional acupuncture for knee osteoarthritis, J. Tradit. Chin. Med. 4 (2009), 249-252.

[9] D. Ruta, A. Garratt, M. Abdalla, K. Buckingham, I. Russell, The SF 36 health survey questionnaire. A valid measure of health status, BMJ 307 (1993), 448-449.

[10] C. L. Lam, E. Y. Tse, B. Gandek, D. Y. Fong, The SF-36 summary scales were valid, reliable, and equivalent in a Chinese population, J. Clin. Epidemiol. 58 (2005), 815-822.