Machine learning methods for improving acupuncture data consistency: a review

G Zhang*, X Xie, Y Huang

Abstract

Introduction

The widespread accepted and practice of acupuncture in Asia and western countries is mainly for its good efficacy to patients with chronic diseases and disorders such as stroke complications, headache, menstrual problems, asthma and so on. It has been reported that 43% of primary care repliers in the USA use acupuncture in clinical practice, but greater evidence can be traced back to the use of randomized control trails to ensure its efficacy. In fact, several new researches have been proposed to provide new and high-quality evidence for showing the efficacy of acupuncture. For the same purpose, questionnaires based on patient-reported outcome (PRO) technique were also developed and applied in clinical practice and trials, in which the Northwick Park Neck Pain Questionnaire (NPQ) and McGill Pain Questionnaire (MPQ) and short-form 36 (SF-36) quality of life questionnaire were widely recognized and applied.

However, owing to variations in syndrome types and the unequal importance of the measures, as well as the subjectiveness of various patients, there may be substantial inconsistency between these measures, leading to some flaw in evaluation, for either medical or computational model. To compensate or correct them, we need to learn a mapping (linear or non-linear) in a supervised manner to adjust the effect of different PROs on final diagnosis.

Following this idea, the methods based on Canonical Correlation Analysis (CCA), Kernel Canonical Correlation Analysis (KCCA) and metric learning have been proposed and evaluated. CCA is a well-known statistical analysis method first proposed by Hotelling in 1936, aiming at finding a linear transformation matrix to keep the correlation between two groups of random variables but of low rank. Inspired by the success of CCA and kernel methods, a new analysis framework named KCCA has been proposed.

Figure 1 shows the changing ratios of NPQ and MPQ scores of three patients evaluated across four therapy stages of neck pain caused by cervical spondylosis. The three sub-graphs indicate that the changing ratios of NPQ and MPQ may be conflicting in some cases. Zhang et al. suggested that the inconsistency originates from the difference of effects in practice and individual subjectiveness, instead of the design of PROs. With the development of machine learning methods and their applications in Traditional Chinese Medicine, some studies have reported how to eliminate the inconsistency between PROs within the machine learning framework.

Conclusion

For future study, non-metric similarity between patient-reported outcomes and the corresponding learning models is an important issue. Meanwhile, mathematical definitions of therapeutic effectiveness of acupuncture are also valuable to be further studied.

* Corresponding author
Email: ipx@gdut.edu.cn
School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Figure 1: NPQ and MPQ score changing ratios in four stages.

Licensee OA Publishing London 2014. Creative Commons Attribution License (CC-BY)

With the introduction of kernels, CCA framework obtained the ability of analysing non-linear relationship between input variables in a linear optimization framework. Nowadays, methods based on KCCA have been widely accepted and applied to many practical tasks\textsuperscript{12-14}. Particularly, for data samples lying in high-dimension space, KCCA can serve as an effective method for dimensionality reduction and feature combination\textsuperscript{15-18}.

KCCA has been reported as an effective method for attaining data consistency in the evaluation of therapeutic effectiveness of acupuncture. In 2011, Zhang et al.\textsuperscript{9} proposed to apply non-linear transformation induced by a kernel function for different clinical PROs and combine them into a consistency feature space, in which the transformation function is learnt from historical data with expert knowledge by machine learning algorithm. The transformation matrix was obtained by solving an optimization problem with the goal of maximizing correlation between input PROs. In 2009, Zhao et al.\textsuperscript{15} proposed to introduce prior knowledge into the traditional KCCA algorithm by applying a fuzzy set-based approach in feature representation. Then Liang et al.\textsuperscript{16} extended the KCCA method and proposed a multi-variable version by introducing a hybrid metric and block-wise optimization method to reduce the natural inconsistency of the concerning PROs. However, the time complexity of KCCA-based methods is very high because there are matrix inversion operations and a quad-optimization with matrix constraints. Hence it works only with small-scale data sets.

Therefore, a supervised learning metric algorithm was introduced by Zhang et al.\textsuperscript{20} to learn a weighted combination of the related PROs with least inconsistency. The authors used a learnt weight vector to combine NPQ and MPQ, which are supervised by the SF-36 scores. Indeed, several researchers have reported to overcome the non-metric property of data sets by the metric learning method. For instance, the distance metric learning method with locality constraints of Lu et al.\textsuperscript{21} is closely related to the consistent metric learning method proposed by Zhang et al.\textsuperscript{20}. Moreover, a metric decomposition method that Laub and Müller\textsuperscript{12} introduced also aimed at tackling the non-metric problem.

Another method for improving the subjective data consistency of acupuncture is metric learning, which has been proposed by Zhang et al.\textsuperscript{20}. Different from the idea of finding an optimal mapping, the metric learning-based methods put the data samples in a Mahalanobis space induced by a training data set, in which the inconsistency between data samples is minimized. The difference between these two methods is the awareness of the distance or dissimilarity between data samples. This review discusses machine learning methods for improving consistency of acupuncture data.

**Discussion**

The authors have referenced some of their own studies in this review. These studies were conducted in accordance with the Declaration of Helsinki (1964) and the protocols of these studies were approved by the relevant ethics committees related to the institution in which they were performed. All human subjects in these referenced studies gave informed consent to participate.

In this section, we present a theoretical review of KCCA-based methods for improving subjective data consistency of acupuncture use, as well as the supervised metric learning framework.

**Kernel Canonical Correlation Analysis**

Intuitively speaking, the essence of KCCA-based methods is to find a nonlinear transformation in kernel space to maximize the correlation between the concerning PROs. Zhang et al.\textsuperscript{9} introduced KCCA into the problem related with subjective data analysis of acupuncture use. In their work, three PROs for neck pain evaluation, i.e. NPQ, MPQ and SF-36, are considered. Their method aimed at finding an optimal mapping matrix to reduce the inconsistency between these three PROs for each patient. Because the original KCCA method was designed for two groups of random variables, and could not satisfy practical requirements because there are at least three measurements related to a patient, a modified method that can simultaneously optimize more than two PROs was proposed in 2012.

Let $X=(x_1, x_2, ..., x_n)$ and $Y=(y_1, y_2, ..., y_n)$ be two groups of random variables. The goal of CCA is to find two transformation matrices $M_X$ and $M_Y$ to maximize the covariance between $M_X X$ and $M_Y Y$. Since matrix $M$ provides only linear projection, it may not express non-linear relationship. Kernel function can be introduced to perform non-linear mapping. Suppose $K$ is a valid kernel function that induces a mapping from input space to feature space. The goal of KCCA is to find optimal $M_X$ and $M_Y$, as shown in Eq. (1).

$$\max_{M_X, M_Y} \frac{\sum_{i,j} K(x_i, x_j) M_{ij} K(y_j, y_j)}{\sum_{i,j} K^2(x_i, x_i) M_{ij} K^2(y_j, y_j)}$$  

(1)

For Eq. (1), rescaling of $M_X$ and $M_Y$ would not affect the optimization problem, so they can be rescaled to obtain $M_X K^2 M_X = I$ and $M_Y K^2 M_Y = I$. According to the result of Cazzanti and Gupta\textsuperscript{23}, if the dimensionality of feature space is higher than the number of data samples in the input space, the optimal solution of $M_X$ and $M_Y$ of Eq. (1) lies in span \{$\phi(x_1), ..., \phi(x_n)$\} and span \{$\phi(y_1), ..., \phi(y_n)$\} respectively, i.e. $M_X = \sum_i \alpha_i \phi(x_i)$ and $M_Y = \sum_j \beta_j \phi(y_j)$.

By using norm form of $\alpha$ and $\beta$ as regularization terms to control the
The optimization problem of KCCA can be expressed as following:

\[
\max_{\alpha, \beta} \langle K_{x}, K_{y} \rangle \beta - \eta_1 \| \alpha \|^2 - \eta_2 \| \beta \|^2
\]

\[
\text{s.t. } K_{x} = \sum_{i} \phi(x_{i}) \phi(x_{i}),
\]

\[
K_{y} = \sum_{i} \phi(y_{i}) \phi(y_{i})
\]

Following the idea of Yu et al., the problem of KCCA can be solved as an Eigen decomposition problem with a close-form solution as Eqs. (4) and (5) shows:

\[
w = (a^T, b^T)^T
\]

\[
B^{-1}Aw = pw
\]

\[
A = \begin{pmatrix}
0 & K_{x}K_{x}^T \\
K_{x}K_{x}^T & 0
\end{pmatrix}
\]

\[
B = \begin{pmatrix}
K_{x}K_{x}^T & 0 \\
0 & K_{y}K_{y}^T
\end{pmatrix}
\]

Note that Eq. (5) is a normal Eigen decomposition of matrix \(B^{-1}A\). \(p\) is an Eigen value and \(w\) is the Eigen vector.

Because NPQ, MPQ, and SF-36 may have bias especially when the therapy is delivered in various pragmatic conditions and the receivers are diagnosed with different subtypes of CS. From the idea of Zhang et al., formal definition of the problem is given before applying KCCA to solve it. NPQ measures nine aspects of pain, including pain intensity, sleep, symptoms duration, objects housework, social activities and driving. As the assessing tool for non-specific pain, MPQ is also recorded in the same way as NPQ. During the therapeutic procedure, there are several check points, and the scores of both PROs are recorded. To evaluate the therapeutic effectiveness, Zhang et al. have defined the following method:

Definition 1 (Subjective Therapy Effect Score, STES):

Given a subjective measure vector \(p\), the sum of evaluation change ratios of all check points is the Subjective Therapy Effect Score:

\[
\delta_{\text{STES}} = \sum_{i=1}^{n} \frac{(p_{i+1} - p_{i})}{p_{i}}
\]

Note that a small \(\delta_{\text{STES}}\) means a good therapy effect. The optimization problem of KCCA is divided into two sub-problems. Radial basic function is a widely used kernel function:

\[
K_{\text{RBF}}(x, y) = e^{-\frac{\|x-y\|^2}{\sigma^2}}
\]

where \(\| \cdot \|\) is 2-norm in Euclidean space and \(\sigma\) the width parameter of radial basis function (RBF) kernel.

All properties of NPQ and MPQ represented in vectors are normalized into range [0, 1] and the width parameter \(\sigma\) in RBF kernel is removed at the same time. The optimization problem can be solved according to Eqs. (4) to (7). The improvement of STES can be used for the therapeutic effectiveness evaluation after applying the learned mapping matrix to the original PROs.

To process more than two PROs simultaneously, compared with the traditional KCCA, Liang et al. proposed a modified KCCA-combined pair-wise correlation into an optimization problem. According to their work, in case of three PROs, i.e. NPQ, MPQ and SF-36, the optimization problem is:

\[
\max_{\phi_{\text{NPQ}}, \phi_{\text{MPQ}}, \phi_{\text{SF-36}}} C_{1} + C_{2} + C_{3}
\]

\[
C_{2} = \text{corr}(\phi_{\text{NPQ}}(\text{NPQ}), \phi_{\text{MPQ}}(\text{MPQ}))
\]

\[
C_{3} = \text{corr}(\phi_{\text{SF-36}}(\text{SF-36}), \phi_{\text{MPQ}}(\text{MPQ}))
\]

\[
C_{1} = \text{corr}(\phi_{\text{NPQ}}(\text{NPQ}), \phi_{\text{SF-36}}(\text{SF-36}))
\]

Liang et al. developed a step-wise optimization strategy. Formally speaking, as there are three cross-related target functions in the optimization problem, the algorithm optimizes two measurements each time. Each step is a traditional pair-wise KCCA problem. The procedure stops until the improvement of target function is smaller than a preset threshold.

Supervised metric learning

Generally, the key point of supervised metric learning proposed in Zhang et al. is to learn a parameterized metric that can obtain a consistent score after combining NPQ and MPQ scores of each check point together. SF-36 is used as a supervisor to guide the learning procedure. The idea of this method is to put all data samples into a learned space where maximal consistency between them can be achieved. Figure 2 sketches the main idea of this method.

![Figure 2: The framework of metric learning for improving consistency of PROs.](image)

The problem is redefined as follows. Let \(D = (\text{NPQ}, \text{MPQ}, \text{SF-36})\) be a set of records of PRO scores, the goal is to find a vector \(w^*\) such that:

\[
w^* = \arg \min_{w} \sum_{i=1}^{n} C_{1}(w \times \text{NPQ}, \text{MPQ}), \text{SF-36}) + \Omega(w)
\]

In Eq. (14), \(w \times (\text{NPQ}, \text{MPQ})\) stands for weighted combination of the corresponding fields of NPQ and MPQ. \(\Omega(w)\) is a regularization term controlling the trade-off between model complexity and the fitness of the training data. Function \(C_1(A, B)\) mathematically quantifies...
the inconsistency between two input vectors\textsuperscript{29,26}. Following the definition of Liang et al.\textsuperscript{29}, we have:

\[
\text{Con}(A,B) = \sum_i \frac{A_{i,1} - A_i}{A_i} - \frac{B_{i,1} - B_i}{B_i} \quad |A| \quad (15)
\]

Eq. (15) sums up the difference between changing ratios of PRO A and B. If the changing trends of A and B are exactly the same, the value of Con(A, B) would be very small.

For the PRO data sample from acupuncture for neck pain, there are eight subfields in each record of SF-36 that cannot be directly processed. Two methods are used to combine these subfields together\textsuperscript{9,27,28}.

Inspired by the methodology proposed in Zhang and Zhou\textsuperscript{29}, the optimization problem described in Eq. (14) can be considered as a locality preserved metric learning problem. To shed light on this, we first denote the Gram matrix with respect to the combined measure as \( K_i = \text{Con}(w \times (NPQ, MPQ), w \times (NPQ, MPQ)) \). Denote the Gram matrix with respect to NPQ and MPQ as \( K_{\text{NPQ}} \) and \( K_{\text{MPQ}} \). To make the combined measure be compatible with the original PROs, it is a good idea to impose a transformation to fill the gap between the learned combined measure and the original PROs, as suggested in Zhang and Zhou\textsuperscript{29}. By using LogDet divergence for difference evaluation, the optimization problem can be written as following:

\[
w^* = \arg\min L(\beta, K_{\text{NPQ}}) + L(\beta, K_{\text{MPQ}}) + w^* \cdot L(\beta, K_{\text{w}})
\]

Figure 3 shows the relations between three terms in Eq. (16).

The function \( L(\beta, X, Y) \) calculates the LogDet divergence matrices \( X \) and \( Y \). The parameters \( \alpha \) and \( \beta \) control the trade-off between the consistency with SF-36 and the locality of original measures. At last, the optimal \( w \) can be reached through an iterative procedure\textsuperscript{30}. Though for Eq. (16) there is not a close-form solution, it can be solved as a standard optimization problem with a set of linear constraints.

In general, KCCA-based methods and the supervised metric learning framework can effectively reduce the inconsistency between PROs, leading to more confident diagnosis and higher accuracy for computational evaluation models. The machine learning methods learn from a set of training data to obtain optimal mappings or metrics. Hence a historic data set with supervised information is required in these methods. Note that the training procedure does not need to run for each patient. In most cases, the knowledge implied in the PROs is stable, which means we can discover them from a representative data set and use them in similar background.

The effectiveness of consistency improvement for PROs should be evaluated in some models. The definition of therapeutic effectiveness is an important issue closely related to the quality of PROs. Currently some definitions of therapeutic effectiveness of acupuncture for neck pain treatment have been proposed, i.e. STES\textsuperscript{9} and OPRO\textsuperscript{9,27,28}. With these definitions, the effect of the PROs consistency improvement methods can be validated during the optimization procedure in some machine learning models.

Figure 3: Relationship between the Gram matrix of three concerning PROs.

**Conclusion**

PRO data are naturally subjective, and in machine learning perspective, they cannot be perfectly embedded in Euclidean space, which means most of the Euclidean distance-based machine learning methods cannot work. For future study, non-metric similarities between PROs and the corresponding learning models are important issues. Meanwhile, the mathematical definitions of therapeutic effectiveness of acupuncture are also valuable to be further studied.

**Abbreviations list**

CCA, Canonical Correlation Analysis; KCCA, Kernel Canonical Correlation Analysis; MPQ, McGill Pain Questionnaire; NPQ, Northwick Park Neck Pain Questionnaire; OPRO, Overall PRO Outcome; PRO, patient-reported outcome; RBF, radial basis function; STES, Subjective Therapy Effect Score.

**Acknowledgment**

This research was supported by the National Foundation of Natural Science of China (no. 81373883, the Higher Education Research Fund of GDUT (2013Y04), and the College Student Career and Innovation Training Plan Project of Guangdong Prov. (yj201311845015, yj201311845023 and yj201311845031).

**References**