

支援向量迴歸建立血壓預測模型改善血壓量測的不準確性

An Kao, Jui-Chung Hung*, and Chih-Peng Huang

Department of Computer Science

University of Taipei

Taipei 100, Taiwan

juichung@gmail.com

摘要

本篇論文是改善血壓量測時因外部干擾所產生的不準確數據，一般量測者測量血壓時常有的干擾包括：外部環境雜訊、儀器本身及白袍症(White Coat Hypertension)等，白袍症是量測者面對醫護人員因為緊張感而導致血壓異常升高的現象。量測者會因為看到不準確的數據而恐慌，甚至造成醫護人員誤判病情。本篇利用支援向量迴歸(Support Vector Regression, SVR)在非線性資料上能夠估測出其趨勢的特性，預測血壓趨勢的合理性，忽略異於趨勢的干擾因子(Interference)，進而減少錯誤數據。在模擬環境中，以實際血壓數據加入干擾因子，使用SVR估測，從實驗成果顯示，SVR並不會因為某些偏差值而影響趨勢。我們的實驗結果顯示運用SVR在降低干擾因子上有很好的效果。

關鍵詞：支援向量迴歸、干擾因子、白袍症

1 緒論

現今人們越來越重視健康管理，其中一種方法為血壓量測，如今已有多套的血壓監測系統 [1]，紀錄每次的血壓資料，依數據來自控管生活飲食習慣。依據美國心臟協會(American Heart Association, AHA)與馬約診所(Mayo Clinic)[2][3]的數據指出，血壓分成六個類別，各類別的收縮壓(Systolic)與舒張壓(Diastolic)如下：低血壓(<90 mmHg, <60 mmHg); 正常血壓(90-119 mmHg, 60-79 mmHg); 高血壓前期(120-139 mmHg, 80-89 mmHg); 高血壓第一階段(140-159 mmHg, 90-99 mmHg); 高血壓第二階段(160-179 mmHg, 100-109 mmHg)和高血壓危機(≥ 180 mmHg, ≥ 110 mmHg)。量測血壓的地點大致上分成兩種，家裡(Home Blood Pressure, HBP)與診所(Clinic Blood Pressure, CBP)，依據研究調查[4]，HBP所量測出的數據比

CBP更準確且有可靠性，在家中以平穩的狀態進行血壓量測，不會有太多的干擾，而在醫院或診所，有可能發生白袍症(White Coat Hypertension)狀況。根據一個長期研究指出[5]，有12.5%的人會有白袍症，而在這群人中的10.8%會有家裡所測出的高血壓數據，所以在家中所測量的數據會比較準確。血壓測量時的干擾因子包含了受測者的身心狀況(緊張、亢奮等)、量測中的身體動作(手部移動、講話等)、白袍症、血壓機的偵測錯誤和外部環境等都會造成不正確的數據，因此必須適當地預測出血壓數據的趨勢，來判斷該次測量的數據是否正確且合理。

支援向量機(Support Vector Machine, SVM)是由學者Vapnik在1995年與AT&T Bell實驗室團隊提出的方法[6]，是一種統計學習理論(Statistic Learning Theory)為基礎的學習演算法，採取結構風險最小化原則(Structural Risk Minimization, SRM)。正確且合理的數據趨勢需用到SVM的延伸：支援向量迴歸(Support Vector Regression, SVR)，此方法是由學者Vapnik、Colwich與Smola在1997年提出[7]，SVR加入了不敏感損失函數 ϵ 、懲罰因子 C 、鬆弛變數 ξ_i 、 ξ_i^* ，使得模組更具有強健性，常用於非線性的迴歸估測、時間序列預測等問題，例如預測股票指數[8]。在Alex J. Smola and Bernhard Schölkopf論文中[9]有提出詳細的SVR介紹，由於SVR展現了很好的成果，所以本論文採用了此方法，應用在血壓數據的預測，降低干擾因子，防止錯誤的判斷上有不錯的成果。

接下來的論文中，分成4個部分，首先會先介紹本篇論文應用SVR建立模型，第二部分為研究方法，第三部分是實驗模擬成果，最後是本論文的結論。

2 利用支援向量迴歸建立血壓模型

由於SVR屬於監督式學習，本論文使用50筆真實血壓資料，其中的前30筆當作訓練集合，後20筆當作測試資料，訓練集合如下：

$$\{(x_1, \hat{x}_1), (x_2, \hat{x}_2), \dots, (x_i, \hat{x}_i) \dots (x_N, \hat{x}_N)\}$$

$$i = 1, 2, \dots, N, N = 30, \hat{x}_i = x_{i+1}, x_i, \hat{x}_i \in R, x_i \text{ 是}$$

*本篇論文是由國家科學委員會計畫支援，計畫編號：NSC 102-2221-E-845-001

輸入值， \hat{x}_i 是與 x_i 相對應的目標輸出值，總共有 N 筆訓練資料。血壓資料以線性函式表示成：

$$f(x) = wx_i + b \quad (1)$$

其中 w 是函式 $f(x)$ 的法向量， b 為截距。SVR加入了懲罰因子 C (C 為常數, $C > 0$)、鬆弛變數 ξ_i 、 ξ_i^* 和不敏感損失函數 ε ，我們的目標是要找出最小的 w 。

其中懲罰因子 C 影響了 ξ_i ， ξ_i^* 的值， C 值越大，錯誤的懲罰越大，反之越小。不敏感損失函數 ε 使得SVR具有強健性，如果目標值與預測值的差小於 ε ，則在容忍範圍內，假設 ε 值太小，則不具有強健的特性，反之當 ε 太大，表示容忍程度過大，則無法估測數據的趨勢。

圖一呈現了軟性邊界的線性SVR，有著 ε 半徑的管狀圖與資料間有密切的關係。小的圓圈代表資料，管內的圓圈表示0誤差，在邊界上或是之外的圓圈，我們稱之為支援向量(Support Vector, SV)，也就是能代表資料趨勢的向量。

利用Lagrange求極值的方法解目標函式：

$$L = \frac{1}{2}w^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) - \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^N \alpha_i (\varepsilon + \xi_i - \hat{x}_i + wx_i + b) - \sum_{i=1}^N \alpha_i^* (\varepsilon + \xi_i^* + \hat{x}_i - wx_i - b) \quad (2)$$

其中的 η_i ， η_i^* ， α_i ， α_i^* 為Lagrange乘數 ($\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$)。對Lagrange方程式做偏微分計算：

$$\frac{\partial L}{\partial b} = \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \quad (3)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i = 0 \quad (4)$$

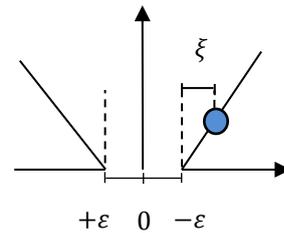
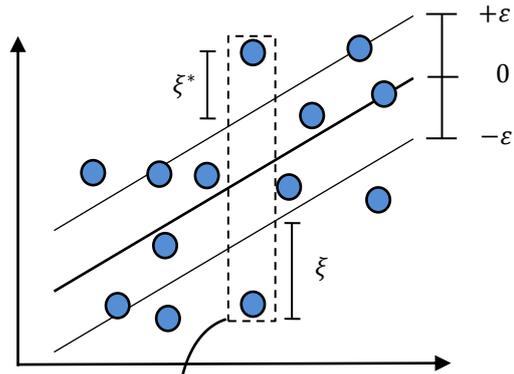
$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \eta_i^{(*)} - \alpha_i^{(*)} = 0 \quad (5)$$

由偏微分後的(4), (5)可以得知：

$$w = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i \quad (6)$$

$$\eta_i^{(*)} = C - \alpha_i^{(*)} \quad (7)$$

將(6), (7)代回Lagrange方程式(2)，處理對偶問題(Dual Problem)與二次規劃問題(Quadratic Programs Problem)：



圖一、SVR示意圖

Maximize

$$-\frac{1}{2} \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) x_i x_j - \sum_{i=1}^{N_S} \varepsilon (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N_S} \hat{x}_i (\alpha_i - \alpha_i^*) \quad (8)$$

$$\text{Subject to } \begin{cases} \alpha_i, \alpha_i^* \in [0, C] \\ \sum_{i=1}^{N_S} (\alpha_i - \alpha_i^*) = 0 \\ i = 1 \dots N_S \quad (N_S = \text{SV 個數}) \end{cases} \quad (9)$$

又因為得知等式(6)，將其帶回函式(1)，可以將線性方程式改寫為：

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i x + b \quad (10)$$

利用Karush-Kuhn-Tucker (KKT)條件去解 b 的值，Lagrange乘數與其相對應的限制式相乘為0。已知 α_i 與 α_i^* 不會同時為0，當 $\alpha_i = 0$ ，則 $\xi_i = 0$ ，相對應的輸入向量點在邊界上；當 $0 < \alpha_i < C$ ，則 $\xi_i = 0$ ，輸入向量在邊線上；當 $\alpha_i < C$ ，則 $\xi_i > 0$ ，輸入向量在邊界外，我們的目標也就是尋找 $\alpha_i^{(*)}$ 不為0的相對應輸入向量，也就是在邊界上或是之外的點，我們稱之為支援向量。

非線性函數藉由映射至高維度空間，以線性函數的方式處理計算，減少時間複雜度。本論文使用多項式核函數運用在SVR上。下列有四個常用的

核函數(Kernel Function)：

核函數	方程式 $K(x_i, x)$
線性核函數 (Linear kernel)	$x_i^T x$
多項式核函數(Polynomial Kernel)	$(x_i^T x + 1)^d$
輻射基底核函數 (Radial Base Function)	$exp\left(\frac{\ x_i - x\ ^2}{2\sigma^2}\right)$
Sigmoid 核函數	$tanh(\alpha x_i^T x + c)$

表一、常用核函數

3 研究方法

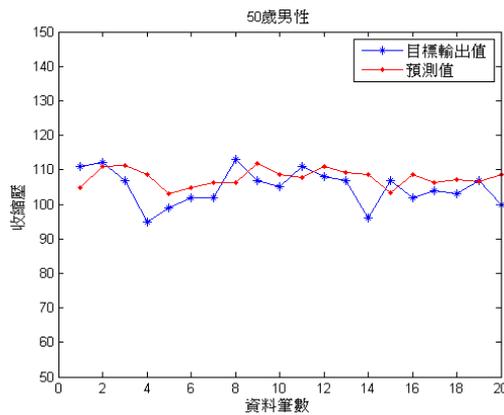
本篇論文的有兩位受測者，各是50的男性與45歲的女性，女性受測者有輕微的低血壓症狀(收縮壓與舒張壓各為，<90 mmHg, <60 mmHg)，而男性並無任何心血管疾病(收縮壓與舒張壓各為，90-119 mmHg, 60-79 mmHg)。圖二是所使用的血壓計設備。每天在晚上睡覺前以坐姿測量，且測量前會休息五至十分鐘，目的是為了得到在身心狀況都在穩定狀態下的數據，而數據都是量測三次所平均的結果，數據收集時間為期兩個月，最後各取其中的50筆收縮壓數據當作實驗資料。前30筆為訓練資料，後20筆為測試資料，在測試資料中，男性的第10、16筆中加入干擾因子，女性則是在第4、10筆加入干擾因子。

SVR模組所需要的參數設置，經過各種不同的組合後，我們選擇了多項式核函數，其參數設置為： $C = 1$ ， $\epsilon = 0.1$ ， $\gamma = 0.04$ 。SVR的模擬實驗是藉由LIBSVM模組在Matlab上所執行的[10]。

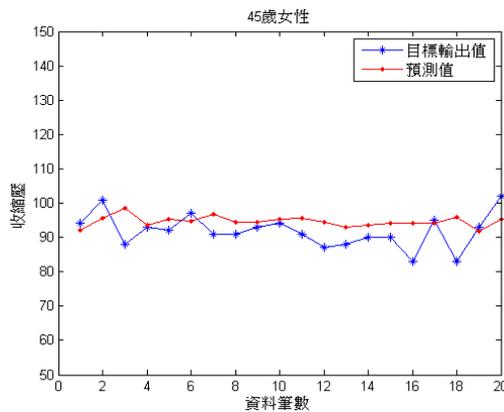
4 實驗成果



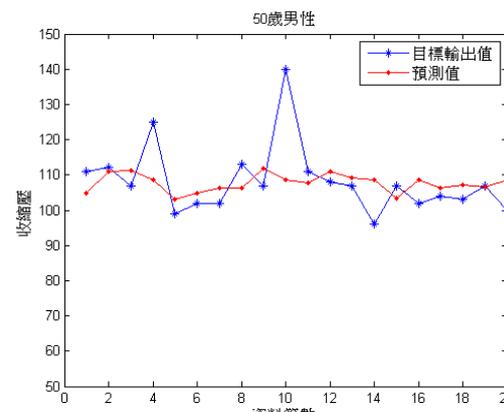
圖二、手臂式電子血壓計WatchBP Home A (Microlife百略醫學BP 3MQ1-1)



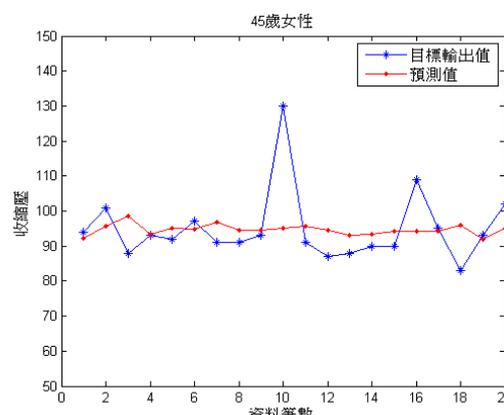
圖三、50歲男性的血壓預測圖



圖四、45歲女性的血壓預測圖



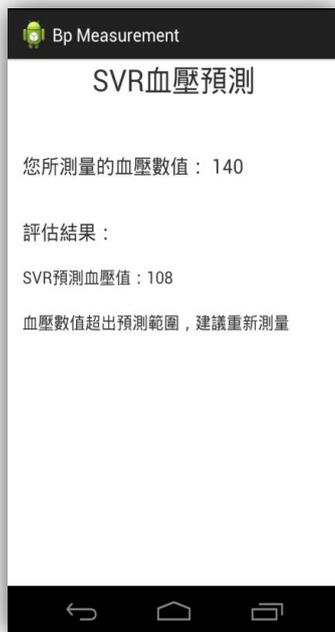
圖五、50歲男性的血壓預測圖(含干擾因子)



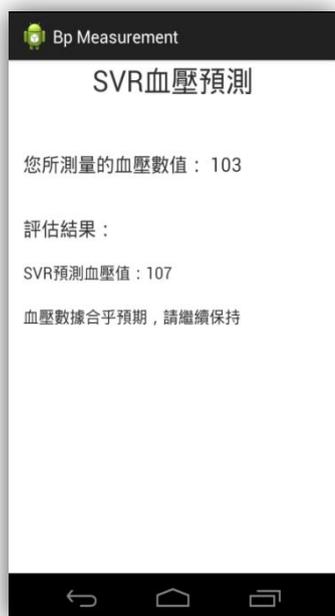
圖六、45歲女性的血壓預測圖(含干擾因子)

根均方差 (Root Mean Squared Error, RMSE)	原始血壓	含干擾因子
50歲男性	5.96	9.38
45歲女性	5.84	9.97

表二、估測結果的根均方差(含干擾因子)



圖七、手機血壓評估系統(有干擾因子)



圖八、手機血壓評估系統(正常血壓)

圖三、圖四為受測者的正常血壓趨勢估測圖，圖五、圖六為加入干擾因子的估測圖，星號(*) 是目標輸出值，小圓點(·) 是預測值。圖七、圖八為手機血壓評估系統，以圖五資料為例，第10筆有干擾因子的加入，SVR預測血壓值為108，而量測值為140，告知量測者超出血壓範圍，建議重新量測，避免造成恐慌；第18筆為正常血壓，請量測者繼續保持。由實驗結果可以得知，SVR能夠正確的預測血壓趨勢且在加入干擾因子的情況下不會受到影響，能夠正常的判斷量測者的血壓是否異常。

表二呈現了原始血壓與有干擾因子情況下的根均方差值，由圖三、圖四也可以觀察出SVR對於趨勢預測有不錯成果。

5 結論

在測量血壓時，不論是在家中或是在醫院診所，都有可能因為干擾因子導致錯誤數據的產生，造成測量者的不安，甚至造成醫護人員對病情的誤判。

本論文提出了以SVR預測非線性數據趨勢的特性，估測出正確的血壓走勢，減少干擾因子的影響，當發生血壓異常時，可以重新測量以確定血壓的正確性，SVR合理的預測趨勢，更可以從中觀察出是否邁入高血壓、低血壓前期或是在正常血壓範圍內。

參考文獻

- [1] 陳永男，雲端血壓監測健康管理系統，臺北市立大學資訊科學系碩士班碩士論文，未出版。
- [2] American Heart Association (AHA), http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp.
- [3] Mayo Clinic, <http://www.mayoclinic.org/diseases-conditions/how-blood-pressure/basics/causes/con-20032298>.
- [4] Toshiyo Tamura, Senior Member, IEEE, Isao Mizukura, Masaki Sekine, Member, IEEE, and Yutaka Kimura. Monitoring and Evaluation of Blood Pressure Changes With a Home Healthcare System. *Journals of IEEE Transaction On Information Technology in Biomedicine*, 2011.
- [5] G. Bobrie, N. Genès, L. Vaur, P. Clerson, B. Vaisse, J.-M. Mallion, and G. Chatellier. Is 'isolated home' hypertension as opposed to "isolated office" hypertension a sign of greater cardiovascular risk? *Arch. Intern. Med*, 2011
- [6] Vladimir N. Vapnik. The Nature of Statistical Learning Theory. *New York: Springer*, 1995.
- [7] Vladimir Vapnik, Steven E. Golowich, Alex Smola. Support Vector Method for Function Approximation, Regression Estimation, and

Signal Processing. *Neural Information Processing Systems Conference, 1997.*

- [8] 黃滄萱, 支援向量迴歸結合真實波動分類預測股票指數, 臺北市立大學資訊科學系碩士班碩士論文, 未出版。
- [9] Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing: Springer, 2003.*
- [10] LIBSVM, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.