Цаконас А.Д., Дуніас Г.Д., Штовба С.Д. Прогнозування результатів футбольних матчів за допомогою машини опорних векторів // Вісник Житомирського інженерно-технологічного інститута.- 2003.- №1. С.181–186.

This article is publised in Ukrainian in "Herald of Zhytomyr Engeneering-Technological Institute" – 2003.- №1. P. 181–186.

UDC 681.3

A. Tsakonas, Ph.D Student, University of the Aegean, Chios, Greece
G. Dounias, Ph.D., Lecturer, University of the Aegean, Chios, Greece
S. Shtovba, Ph.D., Assistant Professor, Vinnitsa State Technical University, Ukraine

FORECASTING FOOTBALL MATCH OUTCOMES WITH SUPPORT VECTOR MACHINES

Анотація. В статі пропонується метод прогнозування резльтатів футбольних ігор, що грунтується на такій технології софт-комп'ютингу як автоматичне навчання на базі машині розділяючої гіперплощини. Розроблена в статті модель прогнозування враховує такі показники команд: різниця кількості вибувших провідних гравців; різниця ігрових динамік команд; різниця класу команд; фактор свого полю; результати особистих зустрічей команд. Тестування показує, що запропонова модель забезпечує добру збіжність прогнозованих та дійсних результатів футбольних матчів, що дозволяє рекомендувати машину розділяючої гіперплощини як перспективний підхід для прогнозування результатів різних спортивних чемпіонатів.

1. Introduction

The prediction of sport game results corresponds to an interesting real-world application of modern decision making and forecasting, while it could also be considered as a good benchmark problem for testing diverse techniques of extrapolation and prediction under difficult conditions of limited available statistics and uncertainties of influence factors. By referring to terms and methodologies such as “intelligent techniques”, “soft computing”, or “computational learning” [1] we mean in fact, a large variety of new powerful techniques for intelligent data analysis, which provide a suitable way for handling complexity, uncertainty and fuzziness of real-world problems. The aim of the present paper is to demonstrate an example of how to predict football game winners by applying such a specific modern intelligent technique, namely Support Vector Machines (SVM). Data representing the Ukrainian football championship during the 10 last years are used for the creation and testing of the intelligent prognostic models applied within this paper.

2. The problem statement

The task of creating football winner prediction models could be reduced to that of finding out functional mapping of the form:

$$\mathbf{x} = (x_1, x_2, ..., x_n) \rightarrow y \in \{d_1, d_2, d_3\}, \qquad (1)$$

where $\mathbf{x}$ - denotes a vector of features (i.e. influence factors), such as team level, climate conditions, playing place, results of past games etc.;

y - denotes the football game result for assessment of one of the terms: d_1 - «host team's win», d_2 - «draw» and d_3 - «guest team's win».

For the need of SVM application the problem is re-stated as follows:

$$\mathbf{x} = (x_1, x_2, ..., x_n) \rightarrow y \in \{-1, 1\}, \qquad (2)$$

where x - denotes the same vector as previously;

y - denotes the football game result for assessment of one of the terms: -1 is equal to result «host team will not win» and 1 is equal to result «guest team will not win».

3. Feature selection

The features carrying the major influence on the game prediction results, always correspond to a subjective choice for every different decision-maker, nevertheless there are some common aspects taken into account from all decision makers. According to [2] these features taken finally into account, are the followings:

x_1 - difference of infirmity factors (as number of traumatised and disqualified players of host team minus the same players of guest team);

x_2 - difference of dynamics profile (as score of host team for five last games minus score of guest team for the five last games);

x_3 - difference of ranks (host team's rank minus guest team's rank, in the current championship);

x_4 - host factor (as $HP/HG - GP/GG$, where HP denotes the total home points of the host team in the current championship; HG is the number of played home games by the host team; GP is the total guest points of the guest team in the current championship; HG is the number of played guest games by the guest team);

x_5 - personal score (as goal difference for all the games of the teams involved, within 10 years).

Note, that the above features do not consist confidential information, but it is easy for the decision maker to know the feature values before the game.

4. Support Vector Machines

SVMs [3] correspond to a relatively new computational intelligence technique, related to the machine learning concept. SVMs are used in pattern recognition as well as in regression estimation and linear operator inversion. SVMs have interesting attributes, different than other computational intelligence techniques, such as neural networks, as SVMs are always able to find a global minimum and they have a simple geometric interpretation. SVMs are also capable of handling large number of data or attributes and their learning is comparable in terms of speed with that of neural networks. More specifically, in order to estimate a classification function such as:

$$f : \mathbf{x} \to \{\pm 1\}, \tag{3}$$

the most important is to select an estimate f from a well restricted so-called *capacity* of the learning machine. Small capacities may not be sufficient to approximate complex functions, while large capacities may fail to generalize, which is the effect of what is called “overfitting”.

In contrast to the neural networks' approach, where the “early stopping” method is used to avoid overfitting, in SVMs overfitting is limited according to the statistical theory of learning from small samples [3]. The simpler decision functions are the linear functions. In the case of SVM, the implementation of linear functions corresponds to finding a large margin separating between two classes. This margin is the minimum distance of the training data points to the separation surface. The procedure to find the maximum margin separation is a convex quadratic problem [4]. An additional parameter enables the SVM to misclassify some outlying training data in order to get larger margin between the rest training data, without however affecting the optimization be the quadratic problem. If we transform the input data into a feature space F using a map such as:

$$\phi : \mathbf{x} \to F, \tag{4}$$

then, a linear learning machine is extended to a non-linear one.

In SVMs the latter procedure is applied implicitly. What we have to supply, is a dot product of pairs of data points $\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \in F$ in feature space. Thus, to compute these dot products, we supply the so-called *kernel* functions that define the feature space via:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j). \tag{5}$$

We don't need to know ϕ, because the mapping is performed implicitly. SVMs can also learn which of the features implied by the kernel are distinctive for the two classes. The selection of the appropriate kernel function may boost the learning process.

4. The SVM algorithm

As assumed in section 3, we are given training set $S = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_N, y_N)\}$, where each point $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{in})$ belongs to $\mathbf{R}^n$, and $y_i \in \{-1, 1\}$ is a label that identifies the class of point $\mathbf{x}_i$. The goal is to determine a function

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b, \tag{6}$$

where $\mathbf{w} = (w_1, w_2, ..., w_n)$ and b are the parameters of shattering hyperplane;

$\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), ..., \phi_m(\mathbf{x}))$ corresponds to a mapping from $\mathbf{R}^n$ into a feature space $\mathbf{R}^m$. This is the standard Kernel Hilbert Space mapping used for kernel learning machines [5].

According to Statistical Learning Theory [3] in order to obtain a function with controllable generalization capability, we need to control the Vapnik Chervonenkis dimension of the function through structural risk minimization. SVMs are a practical implementation of this idea. The formulation of SVM leads to the following quadratic programming problem [5]:

Problem **P1**:

Minimize $\frac{1}{2}\mathbf{w}\cdot\mathbf{w}+C\cdot\sum_{i=1}^{N}\xi_i$,

subject to $y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i)+b)\geq 1-\xi_i,\ \xi_i\geq 0,\ i=1,2,...,N$,

where C is a positive penalty coefficient for a misclassification.

The solution $\mathbf{w}^*$ of this problem is given by the equation:

$$\mathbf{w}^*=\sum_{i=1}^{N}a_i^* y_i\phi(\mathbf{x}_i)\,, \qquad (7)$$

where $\mathbf{a}^*=(a_1^*,a_2^*,...,a_N^*)$ is the solution of following *Dual Problem*:

Problem **P2**:

Maximize $-\frac{1}{2}\mathbf{a}^T\mathbf{D}\mathbf{a}+\sum_{i=1}^{N}a_i$,

subject to $\sum_{i=1}^{N}y_i a_i=0;\ \ 0\leq a_i\leq C,\ i=1,2,...,N$,

where $\mathbf{D}$ is a $N\times N$ matrix such that:

$$D_{ij}=y_i y_j\phi(\mathbf{x}_i)\phi(\mathbf{x}_j)\,. \qquad (8)$$

By combining equations (6) and (7) the solution of Problem **P1** is given by:

$$\sum_{i=1}^{N}y_i a_i^*\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x})+b^*\,. \qquad (9)$$

The points for which $a_i^*>0$ are called Support Vectors (SVs). They are the points that are either misclassified by the computer separating function or are closer than a minimum distance - the margin of the solution - from the separating surface [5]. In many applications they form a small subset of the training points.

For certain choices of the mapping $\phi(\mathbf{x})$ we can express the dot product in the feature space defined by the ϕ's as $\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)=K(\mathbf{x}_i,\mathbf{x}_j)$, where K is called the kernel of the Reproducing Kernel Hilbert Space defined by the ϕ's [5].

We may observe that the spatial complexity of Problem **P2** is N^2, independent from the dimensionality of the feature space. This observation allows us to extend the method in feature spaces of infinite dimension [3]. In practice however, because of memory and speed requirements, Problem **P2** presents limitations on the size of the training set [6].

5. Results

Although our problem is actually a multi-class classification (predict the winner with three possible outcomes: home, host, draw) little research or none has been done in the one-step multi-class [7]. Thus we solve this classification problem as a common regression problem, where the SVM algorithm has to minimize the mean square error. Then, in order to get the predicted outcome, the following rules are applied to the de-normalized forecasted values:

- if *forecasted_value*>=0 consider positive or zero score result – "guest team will not win";
- if *forecasted_value*<0 consider negative or zero score result – "host team will not win".

While SVM classification must be applied between two classes, we select to ignore the draw case as a special case (a no winner case) keeping the sign of the output indicating the predicted class. The algorithm was given as input a set of 105 training data records and the SVM was tested on 70 test data records [2]. All data were normalized in [-1, 1] range. The software applied was the mySVM [8].

We selected as kernel function the *dot* function (simple multiplication) as we had no evidence for the appropriateness of other, more complex functions. We also set the capacity parameter of the SVM equal to $C=1000$. This parameter has to be positive, its value is then divided by the number of examples that are used for training. The other important parameter (see *section 4*) is the insensitivity known as *epsilon*, which is a constant that the prediction can deviate from the functional value without being penalized. In the algorithm it sets both a positive (*epsilon+*) and a negative insensitivity (*epsilon-*). Here we set *epsilon*=0.01.

The algorithm statistics in detail are presented in *Table 1*, and are explained in the paragraph that follows. *Support Vectors* is the number of support vectors produced. *Bounded SVs* are the number of support vectors at the upper bound, then the minimum and the maximum values of the alphas are shown. $|\mathbf{w}|$ is the 2-norm of the hyperplane vector and *VCdim* is an estimator of the Vapnik Chervonenkis dimension computed from the last two values. $(w_1,...,w_5)$ is the hyperplane vector for the attributes and b is the additional constant of the hyperplane. The following results were obtained after 1377 iterations:

- Train Set Mean square error - 0.052297589;
- Test Set Mean square error - 0.053676842.

By applying the classification rules described in the previous paragraph we received the following results: **Correct Prediction on Test Set** is 43 out of 70 examples (accuracy 61.4%).

Table 1 - Support vector learning output statistics

Parameter	Value
Support Vectors	97
Bounded SVs	90
min SV	-9.7087379
max SV	9.7087379
$\|\mathbf{w}\|$	0.12128035
VCdim	<= 1.3774434
w_1	0.2527201
w_2	-0.010411425
w_3	0.28175218
w_4	0.18387293
w_5	0.099184523
b	0.06384628

In order to compare our model with other approaches, in *Table 2* we considered results obtained by other computational intelligent approaches, in previous work [2]. Those results were obtained for a prediction including the draw result of the matches, thus their quotation is here indicatory. Also, results for the fuzzy model and the neural network include the classification score on an 175-element set (training and testing sets). These results can help however to draw general conclusions on the effectiveness of the method in this data set.

Table 2 – Comparison of the SVM model with other approaches

Model	Correct classification
Fuzzy model	64 % (both sets)
Neural network	64 % (both sets)
Genetic programming model	64.28 % (test set)
Support Vector Machines	61.4% (test set)

6. Conclusions - Further Research

This paper briefly demonstrates the application of modern statistical or entropy-based approaches, such as Support Vector Machines. The latter, relatively new computational intelligence approach, was implemented in a common (for SVM theory) “±1” outcome basis, with positive values corresponding to a “guest team will not win” outcome and negative values to a “host team will not win” outcome. These prime results presented in the paper, are indicative of the usability of the SVMs, denoting the competitiveness of this approach among other intelligent approaches for data driven forecasting and decision making. Further research in this domain, may involve hybrid computational intelligent schemes (see a detailed review in [9], for details), while those approaches have been proved in many cases capable of capturing nearly stochastic or chaotic processes offering a high classification and prediction rate.

References

1. *Zadeh L.* Applied Soft Computing –Foreword // Applied Soft Computing, 2000, Vol.1, P.1-2.
2. *Tsakonas A., Dounias G., Shtovba S., Vivdyuk V.* Soft Computing-Based Result Prediction of Football Games // Proceedings of the First International Conference on Inductive Modeling, Lviv, 2002, Vol. 3, P. 15-21.

3. *Vapnik V.N.* Statistical learning theory.- Wiley-Interscience, 1998.- 736 p.
4. *Vapnik V.N.* The nature of statistical learning theory.- Springer-Verlag, 1999.- 304 p.
5. *Cortes C., Vapnik V.* Support Vector Networks // Machine Learning, 1995, Vol. 20, P. 1-25.
6. *Evgeniou T., Pontil M.,* Support Vector Machines with Clustering for Training with Very Large Datasets.- Springer: Lecture Notes in Computer Science, Vol. 2308, 2002.- P. 346-354.
7. *Boser B., Guyon I., Vapnik V.P.* A training algorithm for optimal margin classifiers // Computational Learning Theory, 1992, Vol.5, P. 144-52.
8. *Ruping S.*, mySVM-Manual. Technical Report.- University of Dortmund, Computer Science Department, 2000.
9. *Tsakonas A., Dounias G.* Hybrid Computational Intelligence Schemes in Complex Domains: An Extended Review.- Springer: Lecture Notes in Computer Science, Vol. 2308, 2002, P. 494-512.

Tsakonas Athanasios Demetrios, MSc, PhD Student, University of the Aegean, Chios, Greece. Scientific interests: Computational Intelligence, Decision Making, Wavelets, Chaos Theory. Tel.: (30937) 891399. E-mail: tsakonas@stt.aegean.gr .

Dounias Giorgos D., PhD, Lecturer, University of the Aegean, Chios, Greece. Scientific interests: Computational Intelligence, Decision Making, Wavelets, Medical Applications of Artificial Intelligenc. Tel.: (30271)-94408. E-mail: g.dounias@aegen.gr .

Shtovba Serhiy Dmytrovych, PhD, Assistant Professor, Vinnitsa State Technical University, Vinnitsa, Ukraine. Scientific interests: Fuzzy Logic, Genetic Algorithms, Decision Making, Reliability, Quality Control. Tel.: (0432)-440430. E-mail: serg@faksu.vstu.vinnica.ua .

Стаття надійшла до редакції “___” __________ 2002р.

А.Д. Цаконас, Г.Д. Дуниас, С.Д. Штовба

Прогнозирование результатов футбольных матчей с помощью машины разделяющей гиперплоскости.

В статье предложен метод прогнозирование результатов футбольных матчей, основанный на такой технологии софт-компьютинга как автоматическое обучение на основе машины разделяющей гиперплоскости. Разработанная в статье модель прогнозирование учитывает следующие показатели команд: разница потерь ведущих игроков; разница игровых динамик команд; разница классов команд; фактор своего поля; результаты личных встреч команд. Тестирование показывает, что предложенная модель обеспечивает хорошую согласованность спрогнозированных и действительных результатов футбольных матчей, что позволяет рекомендовать машину разделяющей гиперплоскости как перспективный подход для прогнозирования результатов различных спортивных чемпионатов.

A. Tsakonas, G. Dounias, S. Shtovba

Forecasting football match outcomes with support vector machines.

A soft computing method for result prediction of football games based on machine learning techniques such as support vector machines is proposed in the article. The model is taking into account the following features of football teams: difference of infirmity factors; difference of dynamics profile; difference of ranks; host factor; personal score of the teams. Testing shows that the proposed model achieves a satisfactory estimation of the actual game outcomes. The current work concludes with the recommendation of support vector machines technique as a powerful approach, for the creation of result prediction models of diverse sport championships.