British Journal of Mathematics & Computer Science
7(4): 280-292, 2015, Article no.BJMCS.2015.123
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Social Learning under Uncertainty Based on Dempster-Shafer Approach for Minimizing True Error of Machine Learning

Hegazy Zaher[1*], Mohamed Abdullah[2] and Naglaa Raga Said[2]

[1]Department of Mathematical Statistics, Institute of Statistical Studies and Research (ISSR), Cairo University, Egypt.
[2]Department of Operations Research, (ISSR), Cairo University, Egypt.

Article Information

DOI: 10.9734/BJMCS/2015/15619
Editor(s):
(1) Longhua Zhao, Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, USA.
Reviewers:
(1) Anonymous, KSA.
(2) Ting Wang, State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China.
(3) Anonymous, China.
(4) Anonymous, USA.
(5) Anonymous, Canada.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=935&id=6&aid=8169

Original Research Article

Received: 08 December 2014
Accepted: 31 January 2015
Published: 18 February 2015

Abstract

Minimizing true error of the classification process under uncertainty is one of the difficult issues in the field of machine learning. Researchers do not address this topic until this time despite its importance in practical life. This paper can be considered as a development of the concept of social learning presented the intellectual leap in the machine learning area as given before by the authors. Novelty in this paper is to present a new approach that can deal with the conditions of uncertainty resulting from multiple sources. This paper also presents a new method of social learning based on benefits offered by the Dempster-Shafer theory (DST) of evidence. The paper provides experimental results on six benchmarks. The results attained from the comparison using six benchmarking problems illustrate a superior performance of the proposed method compared with the best results attained in the literature of machine learning domain till now.

Keywords: Uncertainty; social machine learning; dempster-shafer theory; true risk; Vapnik – Chervonenkis theory.

**Corresponding author: Hgsabry@yahoo.com;*

1 Introduction

The Russian school provided a major development in the field of machine learning. This school is based on a deep and strong background in mathematics. This ideology reflected by the adage of Vapnik" nothing more practical than a good theory". Vapnik [1] creates the Russian school in the field of machine learning. The work of Vapnik itself depends on the structure of machine learning, not the data itself as believed by others. Pioneers of the Russian school in this field are Vapnik and Chervonenkis who introduce a new theory called Vapnik and Chervonenkis theory (VC-dimension). VC theory has three dimensions: conceptual, mathematical, and constructive learning. Firstly, the conceptual dimension is presented and developed only by Vapnik [1]. This dimension is concerned with basic characteristics of inference from finite samples based on the idea of Empirical Risk Minimization (ERM). Secondly; the mathematical dimension is presented and demonstrated by Vapnik and Chervonenkis [2].

They prove the following inequality that describes an upper bound of classification error based on the structure of machine learning for unseen data.

$$\varepsilon(f) \leq \varepsilon_n(f) + \sqrt{\frac{h(\log(\frac{2n}{h}) + 1) - \log(\frac{\eta}{4})}{n}} \quad (1)\ [2]$$

Where $\varepsilon(f), \varepsilon_n(f)$ denotes the true error and empirical error. The probability that this bound is crashed equals η, machine learning capacity (VC dimension) of classifiers is h. VC confidence is the second term (the square root) of the inequality (1) [2]. It is worth mentioning that these two dimensions were well known in the seventies of the last century. The main dilemma of the mathematical dimension is the inability to find a method clear and specific on how to measure the VC - dimension of various machine learning. Inequality (1) reflected the general shape but failed to explain how to find the VC - dimension to each machine learning. So, it could not answer many questions about the practical challenges in applying the theory. This deficit has weakened many of the importance of this theory in practice and application.

Thirdly, the constructive dimension which is extended to the mathematical dimension. This dimension is concerned with taking a suitable complexity model to fit the data to attain an acceptable bound of the error through determining the VC - dimension for different types of machine learning. One of the Pioneers in the constructive dimension is Girosi [3] who presents a new technique for estimating VC dimension bounds based on the theory of statistical learning. In the same direction Kon and Raphael [4] enhance the paper of Girosi [3] for finding the L1-norm of an upper bound function. Later, Kon and Raphael [5] develop the Girosi's approach [3] for existing approximate bound of the error based on Hilbert spaces. This method is called Reproducing Kernel Hilbert Spaces. Key et al. [6] who pay their attention for determining the bound of the error using Bayesian decision–theoretic trend. Also, Miller [7] presents a new mechanism for selecting the variables in the model of regression that minimizes the approximate bound of the error. In the same direction, Teytaud and Lallich [8] display the mechanism of using an appropriate VC-dimension to be a restricted bound on the accepted risk from the database. This method is named "association rules". This area attracts many researchers who contribute various methods for estimating the upper bound based on the theory of statistical learning. They enhance the accuracy of the bound by presenting a new concept called Rademacher's complexity but this method allows the curse of dimensionality [9]. Vapnik and Chervonenkis [10] present the new concept of Structural Risk Minimization (SRM) which displays a trade-off between the complexity of the bias function and the accuracy of the approximation. Also, Onshuus, and Usvyatsov [11] propose a method to give the uniform bound on VC. In the same trend, the uniform bound on VC technique has more attention through the work of Shelah et al. [12].

One of the recent applications is presented by Matteo Riondato and Fabio [13] for estimating the sample complexity bound which is based on empirical VC-Dimension in the field of data mining. The constructive dimension presents a variety work for estimating VC-dimension for different types of machine learning as Support vector machines, Decision tree, gene expression programming….. etc. However, the previous constructive work has the following Weaknesses:

1. Absence of the work that is based on social learning.
2. Majority of the work is devoted to developing a neural network area.
3. Assuming all machine learning problems under certain conditions away from realism, but there is no academic work explains how to deal under conditions of uncertainty.
4. Determining the optimal structural based on a single criterion (accuracy level).This trend can be considered as monomeric outlook does not take into account other dimensions such as CPU time.
5. Current techniques cannot perform minimizing both terms of Structural Risk Minimization (SRM) simultaneously.

These weaknesses are considered the main motivator for many researchers to tackle these challenges. Recently, Zaher et al. [14] introduce a new trend called social machine learning. They tackle the first three challenges through proposing new approach is called Tropical Collective Machine Learning (TCML). This approach is the first academic research that introduces the concept of social machine learning in this field.

This paper presents a novel approach for tackling the social learning under uncertainty conditions. This approach is based on Dempster-Shafer theory for dealing with subjective probability of applying on six machine learning techniques. The main merit of the proposed approach is the ability to attain an empirical approximately true error by determining suitable model complexity under uncertainty. The following sections of this paper are structured as follows; Section two presents a brief overview of Dempster-Shafer theory used as a base for the experimental work. Section three displays the proposed approach and gives the used algorithm. Section four shows the data sets used simulation Procedure; performance measures, and display experiments and results. Section five displays the discussion and Section six conclusions of the paper.

2 Dempster-Shafer Evidence Theories

The theory is considered as the development and generalizations of the Bayesian theory provided by Dempster [15] then the improvements are added to this theory by Shafer [16]. The core of this theory is the evidence theory which based on the fundamental concept named Basic Belief Assignment (BBA) that has a clear weakness that is inability to introduce a model of probability in a classical way [17]. This theory is composed of three fundamental functions. The first function is named the basic belief assignment (BBA). The second function is named Belief Function (Bel). Finally, the third function is named Plausibility (Pl). The BBA, denoted by m (X) determines a mapping function convert the power set in the interval between zero and one. The BBA is presented by the equations:

$$m(x) \rightarrow [0,1] \tag{2}$$

$$m(\phi) = 0 \tag{3}$$

$$\sum_{\forall A \in X} m(A) = 1 \tag{4}$$ [18]

This theory presents the concept of the ignorance, also assigns certain value for it. In real life, data fusion is required where multiple sources of information are combined to inform the best judgment of the situation. The main merit from information accumulation is the ability to simply and summarizes the collected data to gain the benefits of multiple sources of information. There are many types of link sources of information, which is called "Combination rules". The real benefit generated from multiple sources is presenting different measures for the same frame of discernment. The fundamental assumption of the Dempster -Shafer theory that sources used must be independent. Dempster introduce his combination rule (for two sources) as follows:

$$m^{1.2}(c) = \frac{\sum_{A \cap B = C} m^1(A) m^2(B)}{\sum_{A \cap B \neq \phi} m^1(A) m^2(B)} = \frac{\sum_{A \cap B = C} m^1(A) m^2(B)}{1 - \sum_{A \cap B = \phi} m^1(A) m^2(B)}$$

[19]

3 The Proposed Approach Dempster-Shafer Collective Machine Learning (DSCML)

The proposed approach (DSCML) in Fig. 1 is composed of two consecutive stages. The first stage is to apply the technique called Tropical Collective Machine Learning (TCML) to select the best appropriate structure of machine learning that will be used in the next stage [14]. The second stage is based on Dempster-Shafer approach for dealing with uncertainty of subjective probability of collective machine learning to attain minimum true error. TCML technique is an intelligent computational method that calculates the expected structure cost of machine learning. TCML algorithm, given by Zaher et al. [14].

3.1 Data Sets (Benchmarking Problems)

This subsection displays in Table 1 six types of benchmark of classification that be used in the experiments.

Table 1. The data of benchmark of classification that be used in the experiments

Data set	Bench mark name	No. of variables	No. of rows
First	Vovel	10	990
Second	Telugu Vovel	3	871
Third	Wisconsin breast cancer	9	699
Fourth	Heptitates	19	155
Fifth	Cleveland heart disease	13	303
Sixth	Diabetes	8	766

These datasets selected used in the experiments are generated from UCI machine learning database.

http://www.ics.uci.edu/˜mlearn/MLRepository.html (20 July 2013).

The first stage: TCML algorithm, Zaher et al. [14]

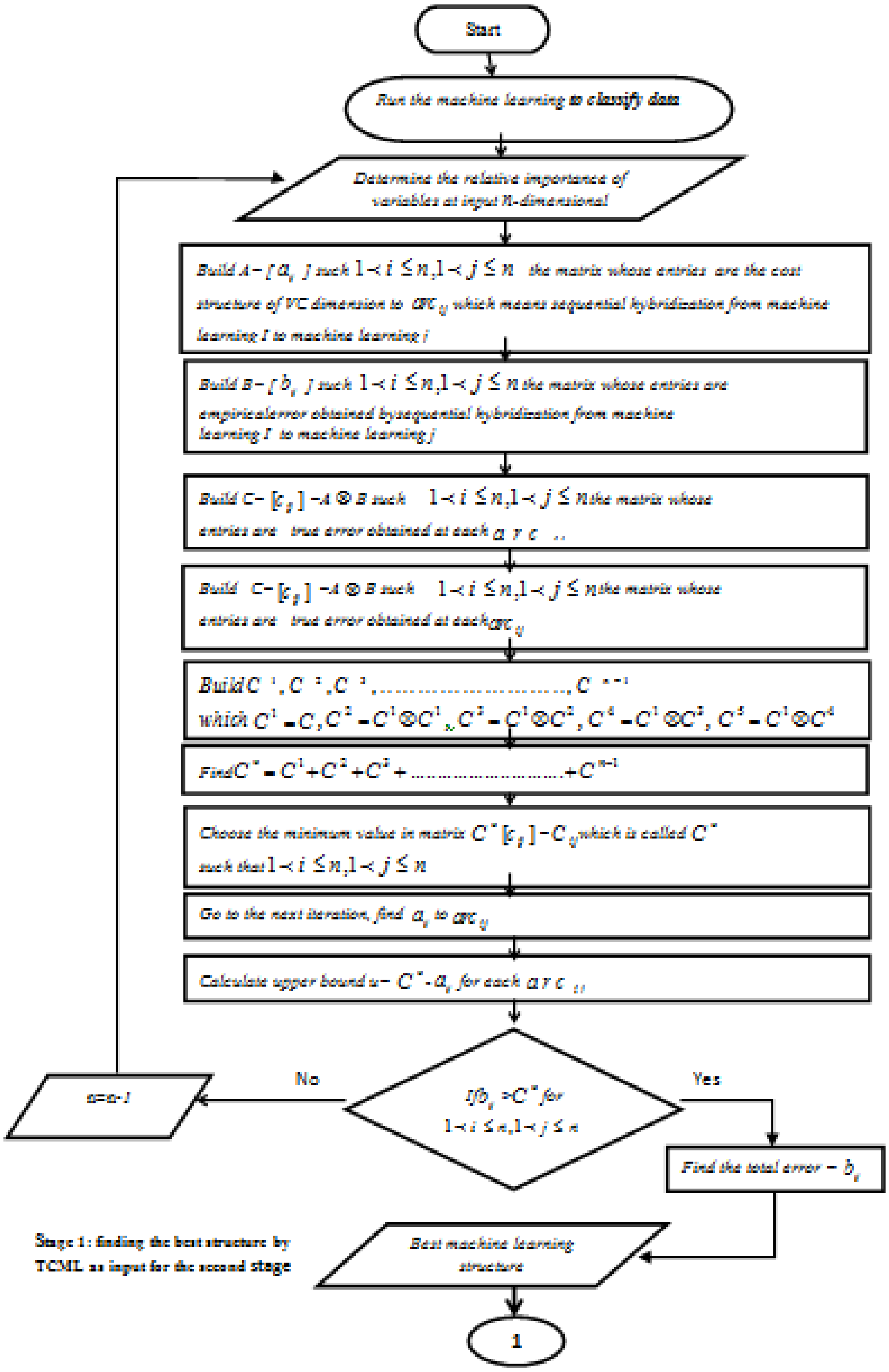
1- Run the machine learning to classify the data.
2- Determine the relative importance of variables at input n-dimensional.
3- Build A= [a_{ij}] such $1 \prec i \leq n, 1 \prec j \leq n$ the matrix whose entries are the cost structure of VC dimension to arc_{ij} which means sequential hybridization from machine learning i to machine learning j .
4 –Build B= [b_{ij}] such $1 \prec i \leq n, 1 \prec j \leq n$ the matrix whose entries are empirical error obtained by sequential hybridization from machine learning i to machine learning j .
5-Build C=$[c_{ij}]$ =A$\otimes$B such $1 \prec i \leq n, 1 \prec j \leq n$ the matrix whose entries are true error obtained at each arc_{ij} .
6-Build $C^1, C^2, C^3, ..., C^{n-1}$
Which $C^1 = C$, $C^2 = C^1 \otimes C^1$, $C^3 = C^1 \otimes C^2$, $C^4 = C^1 \otimes C^3$, $C^5 = C^1 \otimes C^4$.
7-Find $C^* = C^1 + C^2 + C^3 + ... + C^{n-1}$
8-Choose the minimum value in matrix $C^*[c_{ij}]$ = C_{ij} which is called C^*
Such that $1 \prec i \leq n, 1 \prec j \leq n$.
9-Go to the next iteration, find a_{ij} to arc_{ij} .
10- Calculate upper bound u= C^* - a_{ij} for each arc_{ij} .
11- If b_{ij} > C^* for $1 \prec i \leq n, 1 \prec j \leq n$ stop, Else n=n-1 go to step 1
12- Find the total error = b_{ij}

The second stage of proposed algorithm DSCML

1-Run the machine learning with optimal structure achieved by the TCML model to classify data.
2- Determine the new weight of the relative importance for every variable.
3-Construct the weight matrix A= [a_{ij}] such $1 \prec i \leq n, 1 \prec j \leq n$ the matrix. This matrix presents the relative importance of each variable for the first source (first machine learning model).
4 –Construct the matrix b= [b_{ij}] such $1 \prec i \leq n, 1 \prec j \leq n$ the matrix. This matrix presents the relative importance of each variable for the first source (first machine learning model).
5- Construct the matrix C. This matrix is formulated by combined belief function achieved by the two belief function.

$$m^{1.2}(c) = \frac{\sum_{A \cap B = C} m^1(A) m^2(B)}{\sum_{A \cap B \neq \phi} m^1(A) m^2(B)} = \frac{\sum_{A \cap B = C} m^1(A) m^2(B)}{1 - \sum_{A \cap B = \phi} m^1(A) m^2(B)}$$

6-Estimate the final true error achieved at this dimension.
7-If the true error satisfies the stopping criteria, move to step 9, else move to step 8.
8-Eliminate the dimension with the lowest relative weight and replace n by n-1 then Go to step 1.
9- Output the final structure of the model.


Start
Run the machine learning to classify data
Determine the relative importance of variables at input n-dimensional
Build A = [a_ij] such 1 ≺ i ≤ n, 1 ≺ j ≤ n the matrix whose entries are the cost structure of VC dimension to arc_ij which means sequential hybridization from machine learning I to machine learning j
Build B = [b_ij] such 1 ≺ i ≤ n, 1 ≺ j ≤ n the matrix whose entries are empiricalerror obtained bysequential hybridization from machine learning I to machine learning j
Build C = [c_ij] = A ⊗ B such 1 ≺ i ≤ n, 1 ≺ j ≤ n the matrix whose entries are true error obtained at each a r c
Build C = [c_ij] = A ⊗ B such 1 ≺ i ≤ n, 1 ≺ j ≤ n the matrix whose entries are true error obtained at each arc_ij
Build C^1, C^2, C^3, ..., C^(n-1)
which C^1 = C, C^2 = C^1⊗C^1, C^3 = C^1⊗C^2, C^4 = C^1⊗C^3, C^5 = C^1⊗C^4
Find C* = C^1+C^2+C^3+ ....+C^(n-1)
Choose the minimum value in matrix C*[c_ij] = C_ij which is called C*
such that 1 ≺ i ≤ n, 1 ≺ j ≤ n
Go to the next iteration, find a_ij to arc_ij
Calculate upper bound u = C* - a_ij for each a r c_ij
If b_ij > C* for 1 ≺ i ≤ n, 1 ≺ j ≤ n
No
Yes
u=u-1
Find the total error = b_ij
Best machine learning structure
Stage 1: finding the best structure by TCML as input for the second stage
1

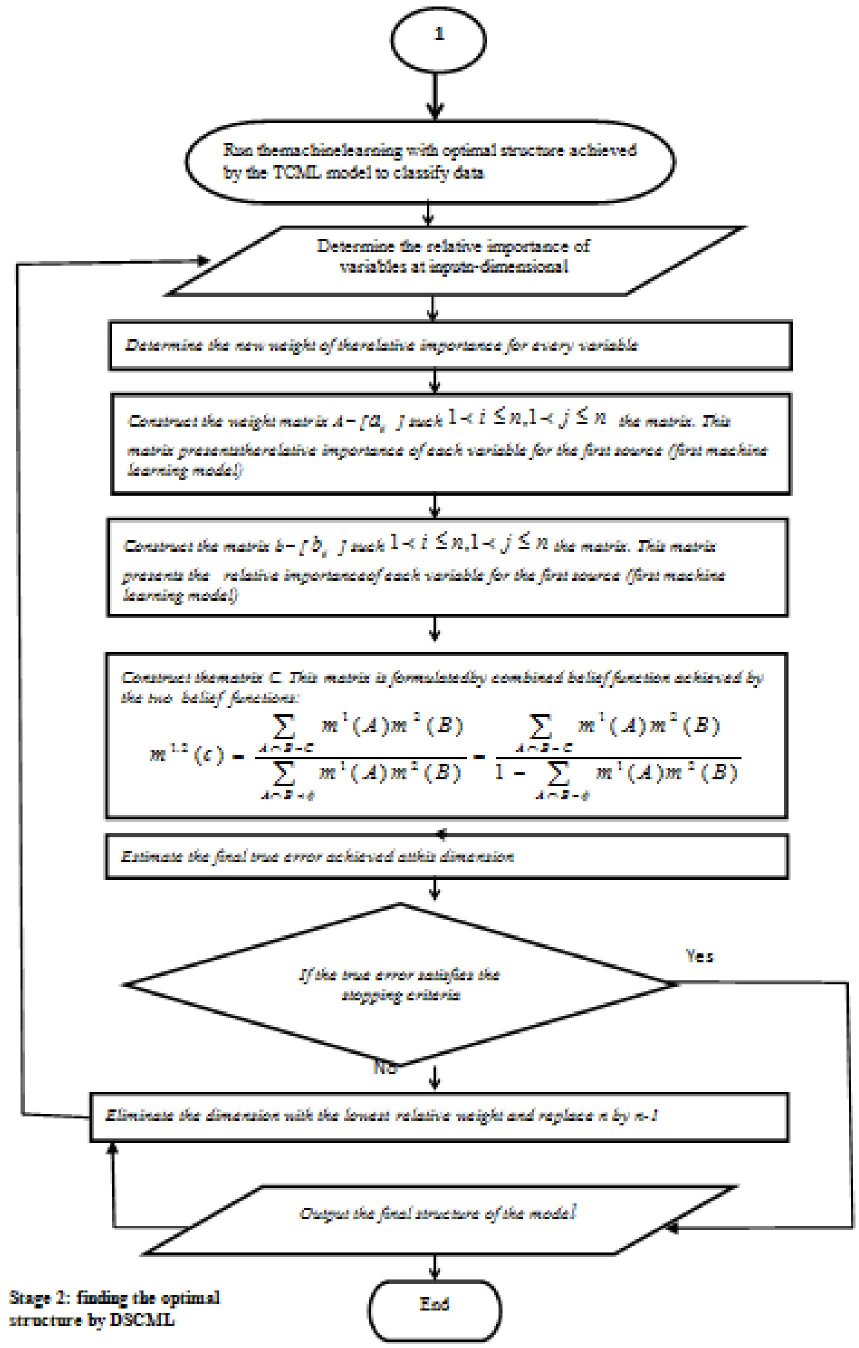


Fig. 1. Flow chart of DSCML (two stages)

3.2 Performance Measures

The accuracy level is used to measure the accuracy of the proposed approach.

CPU time is a second measure to indicate the consumed time of the algorithm.

3.3 Simulation Procedure

The simulation procedure is used in the experiments based on dividing the whole data set, fifty percentages of them used for training and thirty percentages of them used for validating, twenty percentages of them used for testing. These experiments produced: 20 times repeated; the fold cross validation is 10.

4 The Experiments &Results

This section is divided into two subsections. The first of them is assigned for computing the VC–dimension of various techniques of machine learning. The second subsection is assigned for finding the optimal number of dimensions chosen for minimizing the true risk.

4.1 Computing the VC-Dimension

Based on the general inequality (1), the second term of the inequality called (structure cost of machine learning) depends on three variables: his VC- machine learning complexity, n is the sample size and η is the accuracy level required by the machine learning. It is logically through the process of minimizing the controlled variable is h. It is worth mentioning to notice the existence of a positive relationship between the data number of dimensions and VC-dimension. The appropriate structure for various machine learning techniques is shown in the Table 2 using the proposed approach with confidence level 0.9999.

4.2 A Comparison between the Best Results Achieved in the Literature of Machine Learning Field and TCML

This subsection presents a comparison between the TCML technique and the best results achieved in the literature of the machine learning field.

http://duch- links.wikispaces.com/Classification results

Table 3 shows the superior performance of TCML compared with other techniques this dominant performance due to the strength points of the technique:

1- Applying the concept of social learning that maximizes benefits generated from the integration of the information.
2- Specifying the suitable model complexity accurately which leads to minimizing the first term of Vapnik formula.
3- The technique TCML can simultaneously minimize both terms of Vapnik formula [14].

4.3 A Comparison between DSCML and TCML

Subsection presents a comparison between the **TCML** technique and **DSCML** supported by T-TEST for the accuracy and the CPU time. DTREG software package is used to perform the calculations.

Table 4 shows the superior performance of DSCML compared with TCML based on the accuracy level and supported by the T-TEST at 1% significance level. DSCML is clearly more accurate in the six benchmarks. In the first benchmark the improvement is achieved by DSCML 1.29% better than TCML with the P-value 0.001 that ensures the dominance of the DSCML compared with TCML. The same comparison shows the improvement achieved in the other five benchmarks.

The solution is composed of two stages. The first stage is implemented by TCML to find the best appropriate structure of machine learning. The second stage can be implemented by TCML or DCSML. Therefore, the first stage is common in the two methods. Table 5 shows the superior performance of DSCML CPU time compared with TCML and supported by the T-TEST at significant level 1%. Column 3 shows the CPU time estimated in seconds for the finished first stage by TCML which is common in the two methods. This superior speeding of DSCML compared with TCML of finding the optimal solution in is clear in the second stage illustrated in columns 1, 2. It is clear that DSCML dominates the TCML in the total CPU time illustrated in columns 4, 5.

5 Discussion

This paper demonstrates the following facts:

- There is a potential gain by using the DSCML approach to quantifying the ignorance in certain value.
- DSCML has a superior and robust performance compared with TCML due to the ability for considering the ignorance/uncertainty in the machines learning.
- The proposed approach can merge multiple machines learning simultaneously.
- DSCML determines the optimal structure with minimum VC-dimension cost better than TCML.
- DSCML is significantly takes less CPU time than TCML to find the optimal solution.
- DSCML achieves the minimum over fitting compared with TCML.

Table 2. The optimal structure for machine learning techniques used in the experiments

Technique of machine learning	Formula of calculating VC-dimensions	Controlling variables for appropriate structure by DS-ML	VC dimension	Calculated optimal structure cost
Single Decision Tree (STR)	$[\log_2(n-d+2)]+d$ Olcay Taner Yilidiz [20]	n=20,d=3	5.726	0.062079
Multilayer Perceptron (MLP)	$VC = 2[\frac{K}{2}]n$ Eric B. Baum. et al. [21]	K=21 n=3	63.000	0.026055
Radial Basis Function (RBF)	N h(logh) Sakurai A [22]	N=18 h=4	43.348	0.028685
Linear classification (Linear)	$VC = n+1$ Vapnik [23]	n=16	17.000	0.039216
Support vector machines (SVM)	$h \le \min(R^2 A^2, n)+1$ Marcin Owczarczuk [24]	R=0.01667, A=2000 n=18	19.000	0.037604
Group Method of Data Handling (GMDH)	C_h^{n+h-1} Sakurai A [22]	n=12 h=3	55.000	0.026927

Table 3. The results achieved by the TCML compared with the best results btained in the literature of machine learning field

The type of data set	The best method achieved best results for data set	Accuracy level achieved by the reference	Accuracy level of TCML	CPU time estimated in seconds for TCML
Vovel	CART-DB, 10xCV	90.0 Shang, Breiman	93.10	24.85
Telugu vovel	3-NN, Manhattan	87.8 Kosice	90.80	29.18
Wisconsin breast cancer	FSM	98.3 (RA)	99.18	37.98
Heptitates	CMLP2LN/SSV single rule	76.2 WD/K. Grabczewski,	78.20	31.65
Cleveland heart disease	IncNet+ transformations	90.0 Norbert Jankowski.	92.80	32.38
Diabetes	Logistic discrimination	77.7 Statlog	78.80	16.98

Table 4. The results achieved by the DSCML compared with TCML according to classification accuracy

The type of data set	Accuracy level of TCML	Accuracy level of DSCML	TCML 10 cross validation– error rate	DSCML 10 cross validation– error rate	Improvement percentage	P –value
Vovel	93.10	94.39	0.069	0.056	1.290%	0.001**
Telugu vovel	90.80	91.27	0.092	0.087	0.470%	0.006**
Wisconsin breast cancer	99.18	99.47	0.008	0.005	0.290%	0.008**
Heptitates	78.20	79.90	0.218	0.201	1.700%	0.000**
Cleveland heart disease	92.80	93. 40	0.072	0.066	0.600%	0.005**
Diabetes	78.80	80.80	0.212	0.192	2.000%	0.000**

Table 5. The improvement in CPU time achieved by the DSCML compared by TCML

The type of data set	CPU time estimated in seconds for finished second stage by TCML (1)	CPU time estimated in seconds for finished second stage by DSCML (2)	CPU time estimated in seconds for finished first stage by TCML (3)	Total CPU time estimated in seconds for finished two stages by TCML 4=1+3	Total CPU time estimated in seconds for finished two stages by DSCML 5=2+3	Improvement percentage in the second stage	P –value
Vovel	49.12	24.85	32.06	81.18	56.91	49.50%	0.000**
Telugu vovel	39.76	29.18	24.18	68.94	53.36	26.67%	0.000**
Wisconsin breast cancer	55.19	37.98	37.34	92.53	75.32	31.20%	0.000**
Heptitates	46.34	31.65	30.15	76.49	61.80	32.00%	0.000**
Cleveland heart disease	33.19	32.38	26.34	59.53	58.72	2.50%	0.009**
Diabetes	27.80	16.98	18.45	46.25	44.86	39.00%	0.000**

6 Conclusion

Results that have been obtained prove beyond any doubt the ability of the proposed approach to dealing large-scale classification problems under conditions of uncertainty.

This Paper is considered the first paper that introduces the concept of the social machine learning under uncertainty conditions. The results of the comparison between TCML and DSCML prove the superiority of DSCML in terms of accuracy level and CPU time supported by the T-TEST. The comparative results based on well-known six benchmarking problems are done and indicate significant improvement in the speed and accuracy of the solution using DSCML.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Vapnik V, Chervonenkis A. A note on one class of perceptrons. Automation and Remote Control. 1964;(25):821-837.

[2] Vapnik V, Chervonenkis. Ordered risk minimization. Remote Control. 1974;(34):1226-1235.

[3] Girosi F. Approximation error bounds that use VC-bounds. Proceedings of the International Conference on Artificial Neural Networks, Paris. 1995;295-302.

[4] Kon MA, Raphael L, Williams D. Extending Girosi's approximation estimates for functions in Sobolev spaces via Statistical Learning Theory. Journal of Analysis and Applications. 2005;(3):67-90.

[5] Kon MA, Raphael LA. Approximating functions in reproducing kernel Hilbert spaces via statistical learning theory. Wavelets and Splines. G. Chen and M.-J. Lai, Eds., Modern Methods in Mathematics. 2006;271-286.

[6] Key JT, Pericchi LR, Smith AFM. Bayesian Model Choice: What and Why. (With discussion), Bayesian Statistics. 1999;(6):343-370.

[7] Miller A J. Subset selection in regression. (2nd ed.), London. ISBN 1-58488-171-2; 2002.

[8] Teytaud O, Lallich S. Contribution of statistical learning to validation of association rules. Extended version of "Bornesun formes en extraction de règlesd’ association". Proceedings of the International Conference of CAP; 2001.

[9] Mendelson S. A few notes on statistical learning theory, in advanced lectures on machine learning. LNCS 2600, Springer York: Chapman & Hall. ISBN: 978-3-540-00529-2 (Print) 978-3-540-36434-4; 2003.

[10] Vapnik V. Statistical Learning Theory.1st edition, New York: Wiley; 1998.

[11] Onshuus A, Usvyatsov A. On dp-minimality, strong dependence and weight. The Journal of Symbolic Logic. 2011; (76):737-740.

[12] Shelah S. Strongly dependent theories. Israel J. Math; 2008. Accessed 15 July 2014. Available:http://arxiv.org/abs/math/0504197

[13] Riondato M, Vandin F. Finding the true frequent item sets; 2013. Accessed 28 June 2014. cs. LG cs. DB cs.DSstat. ML arXiv: 1301.1218v3. Available:https://scirate.com/search

[14] Zaher H, Abdullah M, Ragaa N. A social learning approach for minimizing true risk of collective machine learning. International Journal of Advanced Research in Computer Science and Software Engineering. 2013;(3):1172-1179.

[15] Dempster AP. Upper and lower probabilities induced by a multi valued mapping. The Annals of Mathematical Statistics. 1967;(38):325–339. Doi:10.1214/aoms/1177698950.

[16] Shafer Glenn A. Mathematical theory of evidence. Princeton University Press, ISBN 0-608-02508-9; 1976. Doi: 10.1090/s0002-9904-1977-14338-3.

[17] Giarratano JC, Riley GD. Expert systems: Principles and programming. 4th ed. Thomson Course Tech, ISBN 0-534-38447-1; 2005.

[18] Rombaut M, Zhu Y. Study on dempster-shafer theory for image segmentation applications. Image and Vision Computing. 2002;(20):15–23.

[19] Tazid A, Dutta P, Boruah H. A new combination rule for conflict problem of dempster-shafer evidence theory. International Journal of Energy, Information and Communications. 2012;(3):35-40.

[20] Olcay TY. On the VC-dimension of the univariate trees. Proceedings of the International Conference on Pattern Recognition Applications and Methods. 2012;193-198.

[21] Baum E, Haussler D. What size nets give valid generalization. Neural Computation. 1989;(1):151-160.

[22] Sakurai A. On the VC-dimension of neural networks with a large number of hidden layers. Proceedings of the International Conference of NOLTA, IEICE. 1993;239-242.

[23] Vapnik V. An overview of statistical learning theory. IEEE Transaction on Neural Networks. 1999;(10):988-999.

[24] Marcin O. Support vector machines with two support vectors. Warszawa, Poland. Department of Applied Econometrics; 2009. Accessed 27 June 2014. Available:http://www.sgh.waw.pl/instytuty/zes/wp/

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=935&id=6&aid=8169