International Journal of Computer & Information Technologies (IJOCIT)
www.ijocit.org & www.ijocit.ir
ISSN=2345-3877

Evaluation of Cancer Classification Using Combined Algorithms with Support Vector Machines

Mahnaz Rafie[1] , Ali Broumandnia[2]
Department of Computer Engineering[1,2], Islamic Azad University[1,2]
Ramhormoz Branch[1], South Tehran Branch[2]
Ramhormoz[1], Tehran[2]
Iran[1,2]

Mahnaz.rafie@gmail.com , Broumandnia@{azad.ac.ir, gmail.com}

Abstract: Support vector machine (SVM) is a supervised learning method, which has considerable applications. It shows excellent performance in many pattern recognition applications. Also, combining SVMs with other theories has been proposed as a new direction to improve classification performance. Thus, in this paper some important aspects to reach the best performance in combined algorithms with SVM for cancer classification are explained. Since delay and accuracy are the important parameters to improve the performance in SVMs, some of the methods with these parameters are compared to use the best algorithms in the future works. Finally some directions for researches are provided.

***Keywords*:** Classification, Kernel functions, Machine learning, Support vector machine, Pattern recognition

1. Introduction

Machine Learning is the study of methods for programming computers to learn. Computers are applied to a wide range of tasks, and for most of these it is relatively easy for programmers to design and implement the necessary software. However, there are many tasks for which this is difficult or impossible. The learning classifications consist of four general groups:

First, there are problems for which there exist no human experts. For example, in modern automated manufacturing facilities, there is a need to predict machine failures before they occur by analyzing sensor readings. Because the machines are new, there are no human experts who can be interviewed by a programmer to provide the knowledge necessary to build a computer system. A machine learning system can study recorded data and subsequent machine failures and learn prediction rules.

Second, there are problems where human experts exist, but where they are unable to explain their expertise. This is the case in many perceptual tasks, such as speech recognition, hand-writing recognition, and natural language understanding. Virtually all humans' exhibit expert-level abilities on these tasks, but none of them can describe the detailed steps that they follow as they perform them. Fortunately, humans can provide machines with examples of the inputs and correct outputs for these tasks, so machine learning algorithms can learn to map the inputs to the outputs.

Third, there are problems where phenomena are changing rapidly. In finance, for example, people would like to predict the future behavior of the stock market, of consumer purchases, or of exchange rates. These behaviours change frequently, so that even if a programmer could construct a good predictive computer program, it would need to be rewritten frequently. A learning program can relieve the programmer of this burden by constantly modifying and tuning a set of learned prediction rules.

Fourth, there are applications that need to be customized for each computer user separately. Consider, for example, a program to filter unwanted electronic mail messages. Different users will need different filters. It is unreasonable to expect each user to program his or her own rules, and it is infeasible to provide every user with a software engineer to keep the rules up-to-date. A machine learning system can learn which mail messages the user rejects and maintain the filtering rules automatically [1,2].

It should be noted that some common classification methods in machine learning are fisher's linear discriminate analysis, weighted voting, Naïve Bayes, neural networks, decision tree, cluster, nearest neighbor, support vector machines, boosting [3]. This paper is about Support Vector Machine (SVM), which introduce a learning method. Some of the advantages of SVMs are: training is relatively easy, good generalization in theory and practice, work well with few training instances, find globally best model, it scales relatively well to high dimensional data [4].

The structure of this paper is organized as follows: in Section 2 support vector machines, kernel functions, and Kernel Selection are described then combined algorithms in SVMs and evaluation of the accuracy of them in Cancer Classification are presented in Section 3. Finally, the conclusion is explained in Section 4.

2. Support Vector Machine

Support Vector Machine (SVM) is evolved as an active area of research, which is presented through statistical learning theory. To develop the SVM classifiers, consider a training set $\{x_i, y_i\}_{i=1}^{N}$, where x_i denotes the input feature vector and y_i the target output. The target output $y_i = 1$ constitutes the positive group while, $y_i = -1$ the negative group [4].

The decision surface of SVM in the form of hyperplane is written as:

$$\mathrm{w}.\emptyset(\mathrm{x}) + \mathrm{b} = 0 \qquad (1)$$

Where *x* is the input vector, *w* the weight vector, and *b* the bias.

The calculation of *w* and *b* is subjected to the constraints:

$$\text{Minimize } \emptyset(\mathrm{w}) = \frac{1}{2}\|\mathrm{w}\|^2 \qquad (2)$$

$$\text{Subject to } \mathrm{y_i}(\mathrm{w}.\emptyset(\mathrm{x}) + \mathrm{b}) \geq 1$$

A new set of slack variables are introduced to express the above optimization problem:

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mahnaz Rafie*
1st November, 2013 Volume 1, Issue 2

Minimize $\phi(w, \xi_i) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$ (3)

Subject to $y_i(w.\phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$

Where *C* denotes the regularization parameter.

The decision function of SVM is defined as:

$$f(x) = \mathrm{sign}\left(\sum_{i=1}^{N} \alpha_i y_i k(x, x_i) + b\right) \quad (4)$$

Where α_i is called Lagrange multipliers and $k(x, x_i)$ a kernel function that the most popular kernel functions are as follow [2, 5, 6]:

A. Polynomials of degree q:

$$K(x^t, x) = (x^T x^t + 1)^q \quad q = 2,3, \ldots \quad (5)$$

A polynomial mapping is a popular method for non-linear modeling where q is selected by the user [2, 5].

B. Radial-basis functions:

$$K(x^t, x) = \exp\left[-\frac{\|x^t - x\|^2}{\sigma^2}\right] \quad (6)$$

This defines a spherical kernel. Where x^t is the center and σ, supplied by the user, defines the radius [2,5].

C. Gaussian Radial Basis Function:

Radial basis functions have received significant attention, most commonly with a Gaussian of the form,

$$K(x^t, x) = \exp\left[-\frac{\|x^t - x\|^2}{2\sigma^2}\right] \quad (7)$$

Classical techniques utilizing radial basis functions employ some method of determining a subset of centers. Typically a method of clustering is first employed to select a subset of centers. An attractive feature of the SVM is that this selection is implicit, with each support vectors contributing one local Gaussian function, centered at that data point. By further considerations it is possible to select the global basis function width, s, using the SRM principle [6].

D. Exponential Radial Basis Function:

A radial basis function of the form,

$$K(x^t, x) = \exp\left[-\frac{\|x^t - x\|}{2\sigma^2}\right] \quad (8)$$

Produces a piecewise linear solution which can be attractive when discontinuities are acceptable [6].

E. Multi-Layer Perceptron:

The long established MLP, with a single hidden layer, also has a valid kernel representation,

$$K(x^t,x) = \tanh(\rho(x^T x^t) + e) \qquad (9)$$

For certain values of the scale, ρ, and offset, e, parameters. Here the SV correspond to the first layer and the Lagrange multipliers to the weights [6].

F. Polynomials of degree q:

More complicated kernels can be obtained by forming summing kernels, since the sum of two positive definite functions is positive definite [6].

$$K(x^t, x) = \sum_i K_i(x^t,x) \qquad (10)$$

In practice, a low degree polynomial kernel or RBF kernel with a reasonable width is a good initial try [7].

3. Combined Algorithms in SVM

Some of the algorithms can be combined with support vector machines such as; support vector clustering in [8], support vector machine classification based on fuzzy clustering for large data sets in [9], genetic algorithm (GA) approach combined with support vector machines (SVMs) for the classification of high dimensional microarray data [10] that support vector machine classification based on fuzzy algorithms such as support vector machine classification based on fuzzy clustering for large data sets in [9] and pairwise fuzzy support vector machines [7] has better performance and accuracy than the others that their results are shown. The performance of pairwise SVMs in% for hiragana data is shown in Table. 1:

Table 1. The Performance of Pairwise SVMs in% for Hiragana Data

Kernel	SVM	FSVM	ADAG Max	ADAG Min	ADAG Ave	NUM
Dot	99.23	99.44	99.53	99.15	99.34	2655
Poly4	99.45	99.62	99.65	99.39	99.51	2521
RBF0.1	99.56	99.70	99.70	99.51	99.60	1

International Journal of Computer & Information Technologies (IJOCIT)
www.ijocit.org
ISSN = 2345-3877

The comparison between SVM based on fuzzy clustering and normal SVM with 10^5 data is shown in (table (2)) and (table (3)):

Table 2. Normal SVM

Normal SVM				
%	#	V_s	t	Acc
0.25%	2725	7	3554	99.80%
25%	27250	-	-	-
50%	54500	-	-	-
75%	109000	-	-	-

Table 3. SVM Based on Fuzzy Clustering

SVM based on fuzzy clustering				
%	V_s	T	Acc	k
0.25%	6	36.235	99.74%	20
25%	4	221.297	99.88%	50
50%	4	922.172	99.99%	100
75%	4	2436.372	99.99%	150

H. Evaluation of Accuracy of Standard SVM and Combined Algorithms with SVM in Cancer Classification

The accuracy of the most Combined Algorithms with SVM, datasets, and the year of published them in 2002 until now is shown in Table 4.

Table 4. Evaluation of Accuracy of SVM and Combined Algorithms with SVM in Cancer Classification

	Title	Dataset	% Accuracy	Year
1	Study of support vector machine and serum surface-enhanced Raman spectroscopy for noninvasive esophageal cancer detection [11]	esophageal	85.2	2013
2	Support vector machine classifier for estrogen receptor positive and negative early-onset breast cancer [12]	Breast	93	2013
3	Diagnosis of breast cancer with an innovative adaptive Support Vector Machine [13]	Breast	94.29	2012
4	Prediction of breast cancer in mammagram image using support vector machine and fuzzy C-means [14]	Breast	100	2012
5	Epithelial–mesenchymal transition biomarkers and support vector machine guided model in preoperatively predicting regional lymph node metastasis for rectal cancer [15]	Rectal	72.3	2012
6	Feature selection for lung cancer detection using SVM based recursive feature elimination method [16]	Lung	87.5	2012
7	An Efficient Breast Cancer Screening System Based on Adaptive Support Vector Machines with Fuzzy C-Means Clustering [17]	Breast	99.87	2011
8	Prognostic classifier for stage II gastric cancer based by support vector machine [18]	Gastric	86.3	2011
9	Application of support vector machine in cancer diagnosis [19]	Colorectal, Gastric,	45.7-97.5	2011

		Lung		
10	Breast cancer classification by using support vector machines with reduced dimension [20]	Breast	94.40	2011
11	A Novel SVM based CSSFFS Feature Selection Algorithm for Detecting Breast Cancer [21]	Breast	98.2425	2011
12	Cancer Classification using Support Vector Machines and Relevance Vector Machine based on Analysis of Variance Features [22]	Lymphom a	96.20-100	2011
13	Optimize Support Vector Machine Classifier based on Evolutionary Algorithm for Breast Cancer Diagnosis [23]	Breast	100	2010
14	A Novel Gene-Based Cancer Diagnosis with Wavelets and Support Vector Machines [24]	SMD	98	2010
15	Support vector machines combined with feature selection for breast cancer diagnosis [25]	Breast	99.51	2009
16	Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines [26]	Ovarian	90	2009
17	Detection of Lung Cancer with Breath Biomarkers Based on SVM Regression [27]	Lung	more than 95.5	2009
18	Feature Selection for Cancer Classification Based on Support Vector Machine [28]	Ovarian	100	2009
19	Augmenting Detection of Prostate Cancer in Transrectal Ultrasound Images Using SVM and RF Time Series [29]	Prostate	95	2009
20	Support vector machines combined with feature selection for breast cancer diagnosis [30]	Breast	99.51	2008
21	Classification of FTIR Gastric Cancer Data Using Wavelets and SVM [31]	Gastric	100	2007
22	SVM Approach to Breast Cancer Classification [32]	Breast	100	2007
23	Multi-class Classification of Cancer Stages from Free-text Histology Reports using Support Vector Machines [33]	Lung	64, 82	2007
24	Gene Selection for Cancer Classification using Wilcoxon Rank Sum Test and Support Vector Machine [34]	Breast, leukemia	100	2006
25	Granular SVM-RFE gene selection algorithm for reliable prostate cancer classification on microarray expression data [35]	prostate	100	2005
26	Cancer molecular classification based on support vector machines [36]	leukemia	100	2004

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mahnaz Rafie*
1st November, 2013 Volume 1, Issue 2

27	Gene Selection for Cancer Classification using Support Vector Machines [37]	Colon, Leukemia	98,100	2002

4. Conclusion

Support vector machines (SVMs) are a system for training linear learning machines. They find optimal linear separator and pick the hyperplane that maximises the margin. If the models are nonlinear, kernel functions transform them to higher dimensional space; in the transformed space there is more chance that the classes will be linearly separable [2].

In this paper we show that learning algorithms can be combined with support vector machines such as; support vector clustering in [8], support vector machine classification based on fuzzy clustering for large data sets in [9], genetic algorithm (GA) approach combined with support vector machines (SVMs) for the classification of high dimensional microarray data in [10] that support vector machine classification based on fuzzy clustering has better performance and accuracy than the others [9], also we can use these algorithms to other fields and compare their performance and choose the best one.

References

[1] R. Greiner, B. Silver, S. Becker, and M. Grüninger, A Review of Machine Learning at AAAI-87, In Proceedings of the Journal of Machine Learning, Springer Netherlands, Vol. 3, No. 1, PP. 79-92, 1998.

[2] E. Alpaydin, Introduction to Machine Learning, the MIT Press Cambridge, Massachusetts London, England, 2004.

[3] Y.L.J. Han, Cancer Classification Using Gene Expression Data, Department of Computer Science University of Illinoisat Urbana-Champaign Urbana, USA, 2004.

[4] C. Cortes, V. Vapnik, Support-Vector Networks, Machine Learning, Vol. 20, PP. 273-297, 1995.

[5] B. Scholkopf, and A.J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.

[6] S.R. Gunn, Support Vector Machines for classification and Regression, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, university of south ampton, 1998.

[7] B.E. Boser, I.M. Guyon, and V.N. Vapnik, A training algorithm for optimal margin classifiers, In Proceedings of the fifth annual

workshop on computational learning theory 5, Pittsburgh, PP. 144-152, 1992.

[8] A.B. Hur, D. Horn, H. Siegelmann, and V. Vapnik, Support Vector Clustering, In Proceedings of the Journal of Machine Learning Research, PP. 125-137, 2001.

[9] J. Cervantes, X. Li, and W. Yu, Support Vector Machine Classification Based on Fuzzy Clustering for Large Data Sets, Springer-Verlag, Berlin, Heidelberg, PP. 572–582, 2006.

[10] E.B. Huerta, B. Duval, and J.K. Hao, A Hybrid GA/SVM Approach for Gene Selection and Classification of Microarray Data, Springer-Verlag, Berlin, Heidelberg, PP. 34–44, 2006.

[11] Shao-Xin SX Li, and et al., Study of support vector machine and serum surface-enhanced Raman spectroscopy for noninvasive esophageal cancer detection. In Proceedings of the Journal of Biomedical Optics, Vol.18, No.2, 2013.

[12] R. Upstill-Goddard, D. Eccles, S. Ennis, S. Rafiq, W. Tapper, F. William, J. Fliege, and A. Collins, Support vector machine classifier for estrogen receptor positive and negative early-onset breast cancer, In Proceedings of the PLoS ONE, Vol. 8, No. 7, 2013.

[13] E. Karacan, and E. Kilic, Diagnosis of breast cancer with an innovative adaptive Support Vector Machine, In Proceedings of 20th Signal Processing and Communications Applications Conference, PP. 1-4, 2012, IEEE Computer Society.

[14] G. S. Banu, A. Fareeth, and N. Hundewale, Prediction of breast cancer in mammagram image using support vector machine and fuzzy C-means, In Proceedings of the International Conference on Biomedical and Health Informatics (BHI), PP. 573 – 576, 2012, IEEE Computer Society.

[15] X. J. Fan, X. B. Wan, Y. Huang, H. M. Cai, X. H. Fu, Z. L. Yang, D. K. Chen, S. X. Song, P. H. Wu, Q. Liu, L. Wang, and J. P. Wang, Epithelial–mesenchymal transition biomarkers and support vector machine guided model in preoperatively predicting regional lymph node metastasis for rectal cancer, British Journal of Cancer 2012; Vol. 106, No. 11, PP. 1735-1741.

[16] K. Kancherla, and S. Mukkamala, Feature selection for lung cancer detection using SVM based recursive feature elimination method, In Proceedings of the 10th European conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, EvoBIO'12, pp. 168-176, Springer-Verlag Berlin, Heidelberg, ISBN: 978-3-642-29065-7, 2012.

[17] J. Palanivel, and N. Kumaravel, An Efficient Breast Cancer Screening System Based on Adaptive Support Vector Machines with Fuzzy C-Means Clustering, In Proceedings of the European Journal of Scientific Research, ISSN: 1450-216, Vol.51, No.1, pp.115-123, 2011.

[18] R. Xu, D. Ruan, M. Oiu, Z. Zeng, H. Zhang, H. Luo, Z. Zhu, B. Sun, and S.Y. Sen, Prognostic classifier for stage II gastric cancer based by support vector machine, 2011 ASCO Annual Meeting, http://www.asco.org/ascov2/Meetings/Abstracts?&vmview=abst_detail_view&confID=102&abstractID=79310.

[19] H. Wang, G. Huang, Application of support vector machine in cancer diagnosis, Springer, December 2011, http://www.ncbi.nlm.nih.gov/pubmed/20842538.

[20] A. Mert, N. Kilic, and A. Akan, Breast cancer classification by using support vector machines with reduced dimension, In Proceedings of the ELMAR, PP. 37 – 40, 2011.

[21] S. Aruna, and S.P. Rajagopalan, A Novel SVM based CSSFFS Feature Selection Algorithm for Detecting Breast Cancer, In Proceedings of the International Journal of Computer Applications, IJCA, Vol. 31, No. 8, PP. 14-20, Published by Foundation of Computer Science, New York, USA, October 2011.

[22] A. Bharathi and A.M. Natarajan, Cancer Classification using Support Vector Machines and Relevance Vector Machine based on Analysis of Variance Features, Journal of Computer Science, Vol. 7, No. 9, PP. 1393-1399, ISSN: 1549-3636, 2011.

[23] R.A.S. Hassan, A.F. Hegazy, A.A. Badr, Optimize Support Vector Machine Classifier based on Evolutionary Algorithm for Breast Cancer Diagnosis, International Journal of Computer Science and Network Security, IJCSNS, VOL.10 No.12, PP. 85-90, December 2010.

[24] A.M. Sarhan, A Novel Gene-Based Cancer Diagnosis with Wavelets and Support Vector Machines, European Journal of Scientific Research, Vol. 46, No. 4, PP. 488-502, November, 2010.

[25] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, journal of Expert Systems with Applications, Vol. 36, No. 2, PP. 3240-3247, 2009.

[26] W. Guan, M. Zhou, C.Y. Hampton, B.B. Benigno, L.D. Walker, A. Gray, J.F. McDonald, and F.M. Fernández, Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines, BMC Bioinformatics, Vol. 10, 2009.

[27] S. Ding, T. Hu, Y. Shen, C. Lin, Y. Huang, Detection of Lung Cancer with Breath Biomarkers Based on SVM Regression, In Proceedings of the Fifth International Conference on Natural Computation, ICNC '09, Vol. 2, pp.131 – 138, 2009, IEEE Computer Society.

[28] W. Luo, L. Wang, J.J. Sun, Feature Selection for Cancer Classification Based on Support Vector Machine, In Proceedings of the WRI Global Congress on Intelligent Systems, 2009, GCIS '09, Vol. 4, PP. 422 – 426, IEEE Computer Society, 2009.

[29] M. Moradi, P. Mousavi, A.H. Boag, E.E. Sauerbrei, D.R. Siemens, P. Abolmaesumi, Augmenting Detection of Prostate Cancer in Transrectal Ultrasound Images Using SVM and RF Time Series, In Proceedings of the IEEE Transactions on Biomedical Engineering, Vol. 56, No. 9, PP. 2214 – 2224, 2009.

[30] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, In Proceedings of the Expert Systems with Applications, Vol. 36, No. 2, PP. 3240-3247, 2009.

[31] C.G. Cheng, L.Y. Cheng, R.S. Xu, Classification of FTIR Gastric Cancer Data Using Wavelets and SVM, In Proceedings of the Third International Conference on Natural Computation, 2007, ICNC 2007, Vol. 1, PP. 543 – 547, IEEE Computer Society.

[32] M. Sewak, P. Vaidya, C.C. Chan, Z.H. Duan, SVM Approach to Breast Cancer Classification, Second International Multi-Symposiums on Computer and Computational Sciences, 2007, IMSCCS 2007, PP. 32 – 37, IEEE Computer Society.

[33] A. Nguyen, D. Moore, I. McCowan, M. Courage, Multi-class Classification of Cancer Stages from Free-text Histology Reports using Support Vector Machines, In Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Lyon, France, PP. 5140 – 5143, 2007.

[34] C. Liao, S. Li, Z. Luo, Gene Selection for Cancer Classification using Wilcoxon Rank Sum Test and Support Vector Machine, In Proceedings of the International Conference on Computational Intelligence and Security, Vol. 1, PP. 368 – 373, IEEE Computer Society, 2006.

[35] Y. Tang, Y.Q. Zhang, Z. Huang, X. Hu, Granular SVM-RFE gene selection algorithm for reliable prostate cancer classification on microarray expression data, In Proceedings of the Fifth IEEE Symposium on Bioinformatics and Bioengineering, 2005, BIBE 2005, PP. 290 – 293.

[36] Y. Li, Q. Liu, X. Ruan, Cancer molecular classification based on support vector machines, In Proceedings of the Fifth World Congress on Intelligent Control and Automation, 2004, WCICA 2004, Vol. 6, PP. 5521 – 5524, IEEE Computer Society.

[37] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines, Machine Learning, Vol. 46, No. 1-3, PP. 389-422, 2002.

Authors Profile

Mahnaz Rafie was born in Ahvaz, Iran. She received the B.Sc. degree in computer engineering from Allameh Mohaddese Noori Institute of Higher Education, Iran, in 2006, the M.Sc. degree in Computer Architecture from Azad University of Arak in 2011. She is currently a Ph.D. candidate in Computer Architecture at Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran. Since 2013, she has been with the Department of Computer Engineering, Islamic Azad University, Ramhormoz Branch. Indeed, she is a member of young researchers club since 2010 till now. Her current research interests include Network on Chips, Sensor Networks, Machine Learning and Computer Architectures.

International Journal of Computer & Information Technologies (IJOCIT)
Corresponding Author: *Mahnaz Rafie*
1st November, 2013 Volume 1, Issue 2

Prof. Ali Broumandnia was born in Esfahan, Iran. He received the B.Sc. degree in Computer Hardware Engineering from the Esfahan University of Technology in Esfahan, Iran, in 1992, the M.Sc. degree in Computer Hardware Engineering from the Iran University of Science and Technology in Tehran, Iran, in 1995, and Ph.D. degree in Computer Engineering from the Islamic Azad University-Science and Research Branch in Tehran, Iran, in 2006. From 1993 through 1995 he worked on intelligent transportation control with image processing and designed the Automatic License Plate Recognition for Tehran Control Traffic Company. Since 1996, he has been with the Department of Computer Engineering, Islamic Azad University-South Tehran Branch, where he is currently an Assistance Professor. Since 2003, he has been with Department of Electrical, Computer Engineering and Information Technology, Islamic Azad University-Qazvin Branch, where he is currently a lecturer. He has published over 30 computer books, journal and conference papers. He is interested in Persian/Arabic character recognition and segmentation, Persian/Arabic document segmentation, medical imaging, signal and image processing, and wavelet analysis. He is reviewer of some International journals and conferences.