A Multi-Class Classifier of Modified Convolution Neural Network by Dynamic Hyperplane of Support Vector Machine

Nur Suhailayani Suhaimi[1] . Zalinda Othman[2*] . Mohd Ridzwan Yaakub[3]

[1] Faculty of Computer and Mathematical Sciences, University of MARA Technology, 73000 Merlimau, Melaka, Malaysia.
[2,3] Center for Artificial Intelligence Technology, Faculty of Information Science & Technology, The National University of Malaysia, 43600 UKM, Bangi, Selangor, Malaysia.

Abstract
In this paper, we focused on the problem of evaluating multi-class classification accuracy and simulation of multiple classifier performance metrics. Multi-class classifiers for sentiment analysis involved many challenges, whereas previous research narrowed to the binary classification model since it provides higher accuracy when dealing with text data. Thus, we take inspiration from the non-linear Support Vector Machine to modify the algorithm by embedding dynamic hyperplanes representing multiple class labels. Then we analyzed the performance of multi-class classifiers using macro-accuracy, micro-accuracy and several other metrics to justify the significance of our algorithm enhancement. Furthermore, we hybridized Enhanced Convolution Neural Network (ECNN) with Dynamic Support Vector Machine (DSVM) to demonstrate the effectiveness and efficiency of the classifier towards multi-class text data. We performed experiments on three hybrid classifiers, which are ECNN with Binary SVM (ECNN-BSVM), and ECNN with linear Multi-Class SVM (ECNN-MCSVM) and our proposed algorithm (ECNN-DSVM). Comparative experiments of hybrid algorithms yielded 85.12 % for single metric accuracy; 86.95 % for multiple metrics on average. As for our modified algorithm of the ECNN-DSVM classifier, we reached 98.29 % micro-accuracy results with an f-score value of 98 % at most. For the future direction of this research, we are aiming for hyperplane optimization analysis.
Keywords:
multi-class classification; enhanced convolution neural network; dynamic support vector machine.

1. Introduction

Social media provides a tremendous source of text data for expressing human honest reviews and opinions. Text data from social media can be used in sentiment analysis, where it tells how people genuinely feel. Rather than a simple count of mentions or comments in an image uploaded, sentiment analysis considers emotions. However, the emotions are not simply binary; yes or no; true or false; black or white. Most reviews and opinions can be categorized into multiple labels such as like, dislike and neutral (three or more class labels). Some opinions or reviews can be categorized in a degree of membership, such as very recommended, slightly recommended, less recommended, and not recommended. The challenges with multi-class classification are basically due to the imbalance da taset, noisy data from free-text user input, and binary classification model usage for the current analysis. These challenges lead to lower performance of classification in multi-class datasets. Machine learning has attracted many researchers to using social media data for sentiment analysis and prediction (Salter-Townshend & Murphy, 2013; Yaakub, Latiffi & Zaabar, 2019; Ahmad, Bakar, & Yaakub, 2020). We are using real-world data from Social Media, consisting of user reviews and opinions. Six datasets were gathered from multiple sources; reviews from iPhone SE users, reviews from top-20 branded item buyers in the US (2018), reviews from e-clothing customers, recommendations for HBO movies, recommendations for Netflix shows and recommendations for Disney Plus shows.

2. Related Works

This section briefly explains related research to this paper analysis.

2.1 Multi-class datasets issues

Performance issues are the primary concern when the aim is to evaluate and compare multi-class classifiers or machine learning techniques. Many metrics seem reliable for testing the ability of a multi-class classifier (Grandini, Baghli and Visani, 2020). However, the issues of a single metric alone are sufficient to measure when involving dynamic and unstructured data with multiple class labels always

Manuscript received November 5, 2023
Manuscript revised November 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.11.3

being questioned. Zhao, Barber, Taylor and Milan (2019) claimed that difficult to capture confidential information in text data with multiple classes due to its variety. Classifying text data seems to be supervised learning, and new vocab and text patterns in multi-class could lead to semi-supervised learning. Chaitra and Kumar (2018) reviewed multiple classifiers for the multi-class problem. They summarized that different approaches are significant in using specific metrics to measure certain findings that the researcher focused on. When dealing with some labelled data, and most of them are unlabeled, then a mixture of supervised and unsupervised techniques can be used. Another issue of multi-class datasets is misclassification costs which may differ in the different classes. It is always tricky for the user to provide precise values for such misclassification costs (Benitez-Pena, Blanquero, Carrizosa and Ramirez-Cobo, 2018). In contrast, identifying acceptable misclassification rate values may be much easier. Another issue is the imbalanced datasets. Barman and Chowdhury (2018) claimed that the imbalanced dataset decreased the classifier's performance in supervised learning. Hence, in 2019, Cappozzo, Greselin and Murphy implemented semi-supervised learning for imbalanced datasets and re-implemented them by Pei, Lin, Yang and Zhong (2020). While Li, Zhao, Sun, Gan, Yuan and Tong (2020) implement parameter-free extreme learning for imbalanced classification issues. Grandin et al. (2020) review the possibility of multiple metrics for an imbalanced dataset. The most common techniques applied to supervised multi-class classification are based on natural extensions of the tools valid for the binary case, such as support vector machines (Rahtgamage & Duleep, 2018; Rathgamage & Iacob, 2019), decision trees, k-nearest neighbours, naïve Bayes (Al-aidaroos, Bakar, & Othman, 2010) or machine learning (Ahmad, Bakar & Yaakub, 2020). However, the issues of performance when involving multi-class are still under-explored.

2.2 SVM Hyperplane

As Benitez-Pena et al. (2018) mentioned, SVM creates a boundary called a hyperplane and tries to search for the maximum margin hyperplane (Horn, Demircioglu, Bischl, Glasmachers and Weihs, 2018), which breaks the space to create the best homogenous partitions on two different classes. The support vectors are the points from each class nearest to the maximum margin hyperplane, which is a crucial feature of SVMs (Pei, Lin, Yang and Zhong, 2021). In addition, one can find some techniques for multi-class classification that take advantage of the SVM methods for binary classification. The most popular multi-class SVM-based approaches are One-Versus-All (OVA) and One-Versus-One (OVO). However, this will fix the hyperplane into its maximum two regions, which is helpful for binary classification. Performing multi-class datasets into binary class classifiers leads to misclassifying concepts. Many researchers implement programming formulations to reduce the problem of misclassifying datasets into their destined categories. Blanco, Japon and Puerto (2019) successfully mixed-integer with linear and non-linear programming formulation to reduce dimensionality and fix variable strategies in multi-class classification problems. In 2015, Zhu and Zhong used a support vector machine with minimum class variance to consider class distribution before constructing the hyperplane. Then Ven-Den-Burg and Groenen (2016) improvised SVM by generalising the classification boundaries from the Azure ML approach (Metthew & Sohini, 2015) to reduce the classification errors in multi-class. A review by Prasanna and Don (2018) involving SVM for multi-class classification stated several areas to be experimented with to increase the performance classifiers by modifying the SVM according to the aim of the research. In recent times, there have been several improvements in the design of traditional SVM, such as Lagrangian Support Vector Machine (LSVM), Proximal Support Vector Machine (P-SVM), and Least Square Support Vector Machine (LS-SVM). These models generate two *parallel* hyperplanes for classification according to their distance from these hyperplanes. Mangasarian et al. proposed the Generalised Eigenvalue Proximal Support Vector Machine (GEPSVM), which solves a pair of generalised Eigen-value problems to generate two non-parallel (Ju, Tian, Liu and Qi, 2015) proximal hyperplanes for the two classes. Based on the idea of GEPSVM, Jayadeva, Kemchandani & Chandra (2007) proposed a Twin Support Vector Machine (TWSVM), which solves a pair of Quadratic Programming Problems (QPPs) for generating the two non-parallel hyperplanes. Saigal and Khanna (2020) presented a quantitative analysis of the established SVM-based classifiers on the behaviour of Least-squares Support Vector Machines, Twin Support Vector Machines and

Least-squares Twin Support Vector Machines (LS-TWSVM) classifiers for multi-class. Blanco et al. (2019) optimise the hyperplane arrangements for multi-class problems based on the Difference in Convex Function in Yang and Wang (2013).

2.3 Hybridized Classifiers

Hybridizing method brings together the best aspect from one method to another. Combining the strength of two approaches may overcome the limitation of one. Ansari, Sattar and Babu (2015) hybridized Fuzzy Inference and Neural networks to discover patterns from web log data. Ghareb, Bakar and Hamdan (2015) hybridized feature selection based and genetic algorithm for text classification. Lausser, Schmid, Schirra, Wilhelm and Kestler (2016) injected rank-based into their neural network classifier for expression data. Berrendero and Carcamo (2018) use Linear Components and Quadratic classifiers for multi-class analysis. Zhao, Barber, Taylor and Milan (2019) forecast interval in streaming data by combining bagging and boosting. In 2020, Kuo, Yalley, Kao and Huang improved the ensembles method of Bagging and Boosting for Diabetic multi-class datasets, while Sueno, Gerardo and Medina (2020) combined SVM and Naïve Bayes for multi-class document classification. However, not all combinations will be suitable for any specific objective, even using similar datasets in a similar problem domain.

3. ECNN Algorithm

CNN is very reliable and robust in extracting text from image data. Feature selection in CNN is widely used to extract essential features from images. However, CNN is lacking text pre-processing feature. Thus, we enhanced CNN by injecting three gates of filters to pre-process the text input before undergoing the classification process.

3.1 Filter Gate 1

We remove duplicate data from text data from online users frequently coming in repetitive sentences to show extreme emotion in their reviews. This filter is required to avoid data duplication and influence the statistical value of data classification accuracy. We apply Fourier Transform to get the character frequency spectrum.

block = array ($x_1, x_2, x_3, \ldots\ldots x_n$)
FT = fft (block)

$$\boldsymbol{\omega}_n \leftarrow e^{2xi/n} \quad (1)$$

we arrange the words according to its frequency by using eq. (1). Thus, the block array are as follows:

$x^{[0]} \leftarrow (x_0, x_2, \ldots.. x_{n-2})$

$x^{[1]} \leftarrow (x_1, x_3, \ldots.. x_{n-1})$
$y^{[0]} \leftarrow$ RECURSIVE-FFT ($x^{[0]}$)
$y^{[1]} \leftarrow$ RECURSIVE-FFT ($x^{[1]}$)

$$y_k \leftarrow y_k^{[0]} + \boldsymbol{\omega} y_k^{[1]} \quad (2)$$

$$y_{k+(n/2)} \leftarrow y_k^{[0]} + \boldsymbol{\omega} y_k^{[1]} \quad (3)$$

$$\boldsymbol{\omega} \leftarrow \boldsymbol{\omega}\, \boldsymbol{\omega}_n \quad (4)$$

Eq. (2) arranged the words unit as extracted and next unit followed by eq. (3). The similarity of words being calculated to determine its repetitive value as a score in eq. (4). If similarity ≥ threshold, the repetitive word in a sentence detected. Thus, we remove the word.

3.2 Filter Gate 2

Filtering words that are not in the vocabulary is the following process in our algorithm, ensuring that the function of text classification in sentiment is accurate and precise. The online platform has the advantage of allowing users to type whatever is on their mind and leads to the production of different words that are not in the vocabulary dictionary. We apply porter's algorithm to stem, remove plurals, and recode singular and index penultimate letters. Then we analyze the word with vocabulary stored in GloVe.

if x = <c>vcvc<v>; we remove endings
if consonant > 1; we remove endings

Then we remove punctuation and non-alphabet before comparing with GloVe to determine the out of vocabulary word.

for n != 'null';

m = load.GloVe;
if oov.GloVe(x) = 'yes'
then remove x;
else GloVe.update(x);
return x;

This filtering gate does not directly remove words that are not in the vocabulary. The text inputs are checked with the GloVe dictionary to ensure the existence of the word that carries meaning in the whole sentence or simply describes the expression only. Words that merely describe the expression and do not interfere with the meaning of the whole sentence will be removed.

3.3 Filter Gate 3

In the spelling correction filtering algorithm as in Filter Gate 3, the spelling code is imported from SymSpell to check spelling in the online dictionary storage. SymSpell is an algorithm to find all words within the maximum edit distance from an extensive list of Twitter sentence order in a short time. It can be used for spelling correction and fuzzy string search. This algorithm imports Python port SymSpell version 6.7, a Symmetric Delete spelling correction algorithm that provides higher speed and lower memory usage. The entire sentence that enters this filtering door is given special weight to bring value to the sentiment of the sentence. It scans the text and extracts the words contained in it. It then compares each word with a list of known correctly spelt words. Some suggested words were pulled from data storage in place of words confirmed as misspellings. The suggested word with the highest similarity value will replace the confirmed word as a misspelling. SymSpell is a constantly updated online storage thus it becomes a vocabulary library platform that contains the latest text and correct terms.

4. Dynamic Hyperplane of SVM

Conventional SVM addresses binary class problems. For example; Class has two elements, which are, Class = {negative, positive}. The SVM aims at separating both classes by utilizing a linear classifier, $\omega^{T}x + \beta = 0$, where ω is the *score vector*. We will assume throughout this paper that Class = {negative, neutral, positive} and Class = {not recommended, slightly recommended, recommended, very good, and very recommended} for the reduction of multiclass problems in this case.

4.1 Multi-class region of SVM classifier

As mentioned, multi-class classification aims to classify objects in the correct class with higher accuracy. However, ignoring imbalance (either in the class size or in the misclassification cost structure) may have unwanted consequences on the classification performance, see Benitez-Pena et al. (2018); Barman and Chowdhury (2018); Chaitra and Kumar (2018), Grandini et al. (2020). For instance, social media databases usually offer free text reviews and more categories of reviews to express emotion. For example, the political sentiments on Twitter may have genuine feelings, sarcastic expressions, or hoaxes that lead to multi-labels (Awwalu et al., 2019), and the number of classes is no longer binary. If a standard SVM is used for classifying the dataset, then the estimated rates (average values according to a 10-fold cross-validation approach) are depicted in Table 1. Even though both rates are high, it might be of interest to increase the accuracy of predicting sentiment value, perhaps at the expense of deteriorating the classification rates in the other class. This problem of accuracy metric alone will be addressed in this paper.

In order to cope with the imbalanced dataset for multi-class, either in terms of class size or structure of misclassification costs, different methods have been suggested, see Sueno et al. (2020), Zhao et al. (2019), Pei et al. (2020) and Zu and Zhong (2015). Those methods are based on adding parameters or adapting the classifier construction, among others. For example, Sueno et al. (2020) formulated an improved Naïve Bayes vectorization concerning SVM.

4.2 Dynamic non-linear hyperplane for SVM

This paper's new formulation of the SVM is focused on minimising the overall misclassification rate and the classifier's performance in the multi-class problem. In order to achieve that, novel dynamic non-linear hyperplanes are added to the SVM formulation and hybridised with ECNN to minimise the noisy data problem. The keystone of the new model is its ability to achieve a more profound control over misclassification in contrast to previously existing models of the multi-class classifier. The proposed algorithm will be called Dynamic Support Vector Machine (DSVM), and the resulting classification algorithm will be referred to as the ECNN-DSVM classifier.

The remainder of this paper is structured as follows. In Sect. 4, the two other hybrid algorithms will be justified as a comparative measure with our proposed model. Section 5 aims to illustrate the performance of the new classifier. A description in the depth of the experiments' design and obtained results will be given in Sect 6. The paper ends with some concluding remarks and possible extensions in Sect. 7.

5. Hybrid classifier

We set up experiments with three hybrid classifiers, which are ECNN with Binary SVM (ECNN-BSVM), and ECNN with linear Multi-Class SVM (ECNN-MCSVM) and our proposed algorithm (ECNN-DSVM).

5.1 ECNN-BSVM algorithm

We introduce the standard BSVM with ECNN as text pre-processing algorithm.

$$T = \{(x_1, y_1), \dots, (x_p, y_p)\} \in (R^n \times y)^p \quad (5)$$

where; $x_i \in R^n$

$y_i \in y = \{+1, -1\}$

$i = 1, \dots, p$

standard FGTP+MCNN applied convex quadratic programming problem.

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{p} \xi_i \quad (6)$$

thus $y_i\big((w \cdot x_i) + b \geq 1 - \xi_i\big), i = 1, \dots, p$

$\xi_i \geq 0, i = 1, \dots, p$

where $\xi = (\xi_i, \dots, \xi_p)^T$ and $C > 0$ used as penalty parameter

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_i \alpha_j \, y_i y_j K(x_i, x_j) - \sum_{i=1}^{p} \alpha_i \quad (7)$$

thus $\sum_{i=1}^{p} y_i \alpha_i = 0,$

$0 \leq \alpha_i \leq C, \ , i = 1, \dots, p$

where $K(x, x')$ = Kernel function = convex QPP

5.2 ECNN-MCSVM algorithm

Conventional multi-class support vector machine used binary classification model with training set.

$$T = \{(x_1, +1), \dots, (x_m, +1), (x_{m+1}, -1), \dots, (x_{m+n}, -1)\} \quad (8)$$

where $x_i \in R^n, \ \ i = 1, \dots, m + n$

Let

$S = (x_1, \dots, x_m)^T \in R^{m \times n}$

$T = (x_{m+1}, \dots, x_{m+n})^T \in R^{m \times n}$

$p = m + n$

For multi-class labels; 3 classes and 5 classes we used two convex as follows:

$$\min_{w_+, b_+, \xi_-} \frac{1}{2} C_3(\|w_+\|^2 + b_+^2) + \frac{1}{2}\eta_+^T \eta_+ + C_1 e_-^T \xi_- \quad (9)$$

thus $Sw_+ + e_+ b_+ = \eta_+$

$-(Tw_+ + e_- b_+) + \xi_- \geq e_- \ , \quad \xi_- \geq 0$

and

$$\min_{w_-, b_-, \xi_+} \frac{1}{2} C_4(\|w_-\|^2 + b_-^2) + \frac{1}{2}\eta_-^T \eta_- + C_2 e_+^T \xi_+ \quad (10)$$

thus

$Tw_- + e_- b_- = \eta_-$

$(Sw_- + e_+ b_-) + \xi_+ \geq e_+ \ , \quad \xi_+ \geq 0$

where

$C_i,\ \ i = 1,2,3,4$ are penalties

$e_+\ and\ e_-\ are\ vectors$

Dual forms are as follows:

$$\max_{\lambda,\alpha} -\frac{1}{2}(\lambda^T\alpha^T)\hat{Q}(\lambda^T\alpha^T)^T + C_3 e_-^T\alpha \quad (11)$$

thus $0 \le \alpha \le C_1 e_-$

where $\hat{Q} = \begin{pmatrix} SS^T + C_3 I & ST^T \\ TS^T & TT^T \end{pmatrix} + E$

and

$$\max_{\theta,\gamma} -\frac{1}{2}(\theta^T\gamma^T)\tilde{Q}(\theta^T\gamma^T)^T + C_4 e_+^T\gamma \quad (12)$$

thus $0 \le \gamma \le C_2 e_+$

where $\tilde{Q} = \begin{pmatrix} TT^T + C_4 I & TS^T \\ ST^T & SS^T \end{pmatrix} + E$

I = identity matrix of m x m

E = identity matrix of p x p

5.3 ECNN-DSVM algorithm

Our algorithm apply multiple classification problem with the training sets as follows:

$$T = \{(x_1, y_1), \dots, (x_p, y_p)\} \quad (13)$$

where

$x_i \in R^n\ ,\ \ i = 1, \dots, p$

$y_i \in \{1, \dots, K\}$ is the corresponding pattern of x_i

For multi-class classification, we apply K nonparallel hyperplanes;

$(w_k \cdot x) + b_k = 0,\ \ k = 1, \dots, K$

$$S_k \in R^{P_k \times n}\ \ ,\ k = 1, \dots, K \quad (14)$$

$$T_k = [S_1^T, \dots, S_{k-1}^T, S_{k+1}^T, \dots, S_k^T] \quad (15)$$

$$\min_{\hat{\Pi}} \frac{1}{2}\hat{\Pi}^T\hat{\Lambda}\hat{\Pi} + \hat{K}^T\hat{\Pi} \quad (16)$$

thus

$e_{k_1}^T\lambda - e_{k_2}^T\alpha = 0$

$\hat{C}_1 \le \hat{\Pi} \le \hat{C}_2$

where

$$\hat{\Pi} = (\lambda^T, \alpha^T)^T \quad (17)$$

$$\hat{K} = \left(0, -C_1 e_{k_2}^T\right)^T \quad (18)$$

$$\hat{C}_1 = \left(-\infty e_{k_1}^T, 0\right)^T \quad (19)$$

$$\hat{C}_2 = \left(+\infty e_{k_1}^T, C_2 e_{k_2}^T\right)^T \quad (20)$$

$$\hat{\Lambda} = \begin{pmatrix} \hat{Q}_1 & \hat{Q}_2 \\ -Q_2^T & \hat{Q}_3 \end{pmatrix} \quad (21)$$

where

$\hat{Q}_1 = k(T_k, T_k^T) + C_1 I$

$\hat{Q}_2 = k(T_k, S_k^T)$

$\hat{Q}_3 = k(S_k, S_k^T)$

The new hyperplane vector is assigned to class k (k ∈ 1,⋯ ,K), depending on which hyperplane K is given and it is located furthest after applying Dual-Lagrangian. The decision function is defined as a matrix space. Unlike the existing binary classification, this ECNN-DSVM model does not need to consider the additional kernel-generated vector space because only the label classes in the space defined, thus the kernel function can be used directly in this study if for linear equations for binary classification and also easily adapted to non-linear multiclass classification. Partitioning the vector space are the identity matrix and K(·,·) is the kernel function. In conclusion, the corresponding class data is similar to the linear case except that the result in (x · x′) is taken as a non-linear multiclass, instead of a binary linear function such as K(x, x′). For this experiment, the main focus is to allow the hyperplane to be divided into vector spaces of more than 2 classes non-linearly.

6. Result and discussion

In this section, we present our experiments' selected results with all six data sets. We obtained the result for macro-accuracy versus micro-accuracy as guidance to further analyze the algorithm performance. The difference between micro-average and macro-average is the metric for macro-average calculated independently for each class before considering the average of all classes equally. In contrast, a micro-average computes the average metric by aggregating the contributions of all classes. We use five metrics regarding macro and micro accuracy: precision, recall, f-score, sensitivity and specificity.

Table 1 below shows the performance of multi-class classifier towards all six datasets used.

Table 1. Macro-accuracy versus Micro-accuracy

Dataset		Accuracy	Recall	Precision	F-Score	Sensitivity	Specificity
Iphone SE	Macro	**73.00**	0.72	0.70	0.71	0.72	**0.85**
	Micro	81.99	0.91	0.91	0.91	0.91	**0.93**
Branded item	Macro	71.50	**0.77**	**0.69**	**0.73**	**0.77**	0.79
	Micro	88.00	0.81	0.78	0.79	0.81	0.89
e-clothing	Macro	70.50	0.68	0.71	0.69	0.68	0.75
	Micro	79.42	0.75	0.85	0.80	0.75	**0.88**
HBO	Macro	**63.50**	**0.53**	0.72	**0.61**	**0.53**	**0.68**
	Micro	83.50	0.67	0.98	0.81	0.67	0.85
Netflix	Macro	74.50	0.72	0.70	0.71	0.72	0.84
	Micro	**94.50**	**0.93**	**0.98**	**0.95**	**0.93**	**0.95**
DisneyPlus	Macro	72.50	0.69	**0.76**	0.72	0.69	0.73
	Micro	72.50	0.69	0.76	0.72	0.69	0.73

By considering both macro and micro-average of accuracy, precision, f-score, sensitivity and specificity, we obtained performance results above 50 % for all metrics. Among all six datasets, reviews from Netflix viewers yielded the highest micro-average accuracy value of 94.50 %, while the highest macro-average accuracy is 73 % from the iPhone SE user review dataset.

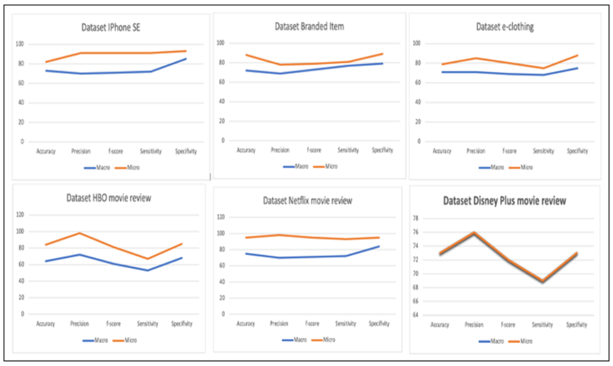

Fig. 1 Macro versus Micro-average Performance in Multi-class Classification

The difference between macro and micro averaging is cleared with macro considering the value of each class equally while micro takes the value of each sample equally. Micro-Average F1-Score and Micro-Average of Accuracy consider all the units together. A poor performance in small classes is not crucial since the number of units belonging to those classes is smaller than the dataset size. As for the dataset DisneyPlus movie review, we have an equal number of samples for each class, then macro and micro resulted in the same accuracy rate. As shown in the classification results for five other datasets where the imbalance datasets occurred, the value of micro-accuracy on average outperforms the value of macro-accuracy. We consider the number of other metrics which contributes the most to the multi-class accuracy result as supported value since a single metric may result in biased and drifting goals. Thus, the size of samples in a dataset does influence the accuracy measure for multi-class classification.

As for the following experiments, we consider the micro-average metrics for all classifiers. We compare algorithm performance to evaluate our proposed ECNN-DSVM with two other algorithms, as shown in Table 2.

Table 2. Performance Comparison by Algorithms

Dataset	Algorithm	Micro-accuracy	f-score	sensitivity	specificity	MSE
	1	68.33	0.61	0.78	0.69	$1e^{-01}$
Iphone SE	2	71.48	0.68	0.91	0.91	$1e^{-01}$
	3	77.43	0.79	0.94	0.93	$1e^{-01}$
Branded items	1	64.15	0.60	0.78	0.84	$1e^{-02}$
	2	65.32	0.64	0.81	0.89	$1e^{-02}$
	3	73.67	0.58	0.84	0.92	$1e^{-02}$
	1	60.77	0.81	0.71	0.88	$1e^{-04}$
e-clothing	2	65.15	0.87	0.75	0.88	$1e^{-04}$
	3	87.58	0.60	0.86	0.90	$1e^{-04}$
	1	60.58	0.58	0.67	0.81	$1e^{-04}$
HBO	2	83.50	0.81	0.67	0.85	$1e^{-04}$
	3	85.50	0.87	0.75	0.89	$1e^{-04}$
	1	60.99	0.60	0.82	0.88	$1e^{-06}$
Netflix	2	94.50	0.95	0.93	0.95	$1e^{-06}$
	3	98.29	0.98	0.88	0.95	$1e^{-06}$
	1	71.34	0.69	0.66	0.79	$1e^{-04}$
Disney Plus	2	72.50	0.72	0.69	0.73	$1e^{-04}$
	3	88.25	0.92	0.74	0.81	$1e^{-04}$

Figure 2 depicts the performance of all three algorithms regarding micro-accuracy, f-score value, sensitivity and specificity.

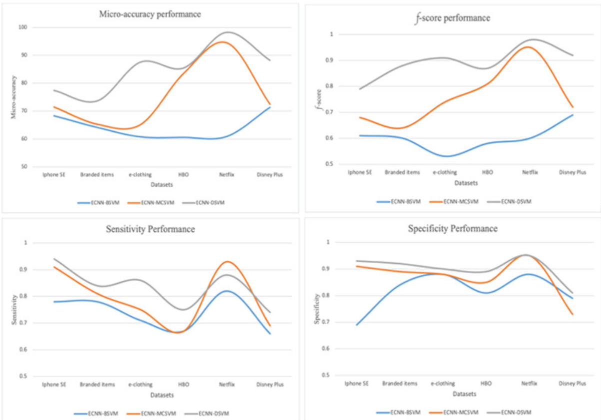

Fig. 2. Comparative Performance of Proposed ECNN-DSVM

Figure 3 depicts the performance of all three algorithms by considering a single metric (micro-accuracy alone) and multiple metrics (average of micro-accuracy, f-score value, sensitivity and specificity).

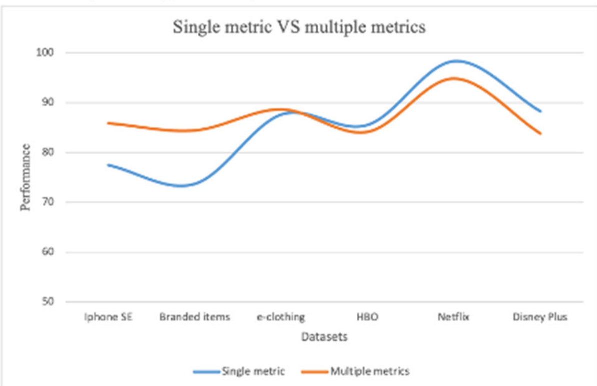

Fig. 3. Single metric versus multiple metrics

Based on the performance of classifiers measured in single and multi-metrics, as shown in Figure3, a significant gap was found in three datasets (iPhone SE reviews, Branded Item revies and Netflix movie recommendations). This shows a better performance of multiple metrics in measuring the classifier with an average of 18 % better than a single metric; even a single metric provides higher performance in Netflix recommendation and Disney Plus recommendation with less than a 5 % average performance gap.

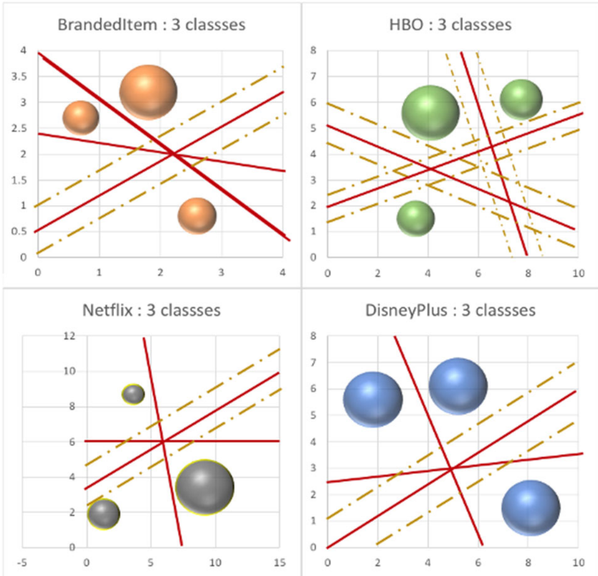

Fig. 4. Dynamic Hyperplane generated for 3 classes dataset

The region produced by the hyperplane of SVM on multi-class datasets for three class labels is shown in Figure 4. The dynamic hyperplane successfully separated the classes with a significant margin gap. The distance of intra-classes among all three classes is, on average, 53% separable from hyperplane minimum and maximum distances. However, as shown in Figure 5, several misclassifications occurred.

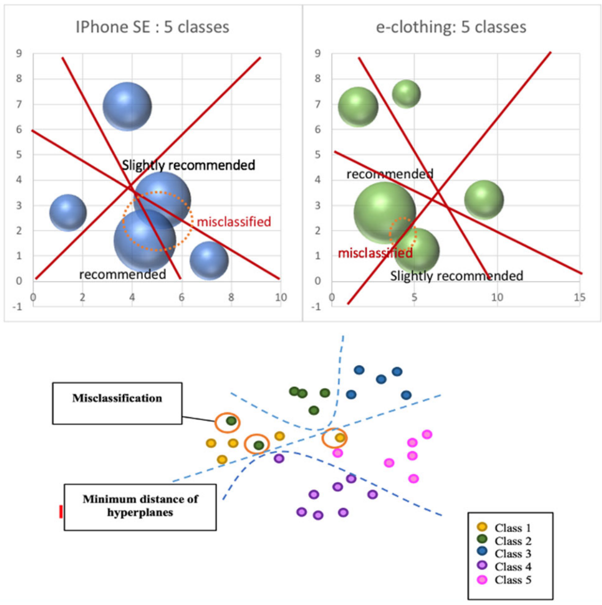

Fig. 5 Dynamic Hyperplane generated for 5 classes dataset

Dynamic hyperplane performed into five class labels datasets, and the value of false-positive results in class 'slightly recommended' and 'recommended' appears to be overlapped. This false-positive results in high specificity rates of 0.93 and 0.88 for the iPhone-SE reviews dataset and e-clothing reviews dataset, respectively (see Table 2). Although the two classes' similarities in text for classes' slightly recommended' and 'recommended' are very minimal and almost redundant, the other classes (not recommended, very good and strongly recommended) yield significant distances among intra-classes with more than 70% separable hyperplane margin from minimum to maximum hyperplane convex.

7. Conclusions and Future Directions

Among the state-of-the-art results, we focus on the dynamic hyperplane of SVM, considered for multi classes of complicated nonconvex text graphs and real-world applications. We want to prove that the DSVM classifier is available for multi-class problems (more than two labels) that appeared in imbalanced real-world text data (continuous or discrete) and forecasting research. Hence the list of references we offered could be helpful for them in the guidance of solution methods for their problems. On the other hand, the list of SVM classifiers available to solve real-world problems, classified area by area, would allow researchers to find the contribution of most interest to them in their specific domain. The result showed that many class problems in sentiment analysis could provide better performance even with the imbalanced dataset. In addition, the analysis of the related works in considering multiple metrics to measure the performance of the classifier proved the point of view in determining the reliability of an algorithm not rely on classification accuracy solely.

As for future direction, we may further analyse by searching deeply on optimising the distance drawn from one hyperplane to another. We may also consider a non-linear hyperplane to the dataset with a minimal distance of intra-classes with many labels. It is known to the world that a greater distance between hyperplanes would better justify the region's quality.

Acknowledgment

We are very grateful to Universiti Teknologi MARA for supporting this research. We sincerely thank our research lab, Data Mining and Optimization (DMO) of Fakulti Teknologi dan Sains Maklumat (FTSM), Universiti Kebangsaan Malaysia, for the expert knowledge sharing. Thanks to Universiti Kebangsaan Malaysia and the Ministry of Higher Education for providing Fundamental Research Grant Scheme (FRGS) code FRGS/1/2021/ICT06/UKM/02/1 for this research funding.

References

1. Ahmad, I.S., Bakar, A.A. and Yaakub, M.R.: A Survey on Machine Learning Techniques in Movie Revenue Prediction. Springer Nature Computer Science, 1(235), (2020).
2. Al-Aidaroos, K.M., Bakar, A.A. and Othman, Z.: Naïve Bayes Variants in Classification Learning. IEEE Xplore, International Conference on Information Retrieval & Knowledge Management, 276-281 (2010).
3. Ansari, Z.A., Sattar, S.A. and Babu, A.V. A fuzzy neural network based framework to discover user access patterns from web log data. Journal of Advances in Data Analysis and Classification. Springer-Verlag Berlin Heidelberg. (2015).
4. Awwalu, J., Bakar, A.A. and Yaakub, M.R.: Hybrid N-gram Model using Naïve Bayes for Classification of Political Sentiments on Twitter. Springer Nature Neural Computing and Applications, 31, 9207-9220 (2019).
5. Barman, D. and Chowdhury, N. A novel semi supervised approach for text classification. International Journal of Information Technology. DOI: https://doi.org/10.1007/s41870-018-0137-9. (2018).
6. Benitez-Pena, S., Blanquero, R., Carrizosa, E. and Ramirez-Cobo, P. On support vector machines under a multiple-cost scenario. Journal of Advances in Data Analysis and Classification. Springer Nature 2018: Springer-Verlag GmbH Germany. (2018).
7. Berrendero, J. R. and Carcamo, J. Linear Components of Quadratic Classifiers. Journal of Advances in Data

Analysis and Classification. Springer Nature 2018: Springer-Verlag GmbH Germany. (2018).
8. Blanco, V., Japon, A. and Puerto, J. Optimal arrangements of hyperplanes for SVM-based multiclass classification. Journal of Advances in Data Analysis and Classification. Springer Nature 2019: Springer-Verlag GmbH Germany. (2019).
9. Burg, G.J.J.V-D. and Groenen. P.J.F.: GenSVM: A Generalized Multiclass Support Vector Machine. Journal of Machine Learning Research, 17, 1-42, (2016).
10. Cappozzo, A., Greselin, F. and Murphy, T.B. A robust approach to model-based classification based on trimming and constraints; Semi-supervised learning in presence of outliers and label noise. Journal of Advances in Data Analysis and Classification. Springer Nature 2019: Springer-Verlag GmbH Germany. (2019).
11. Chaitra, P.C. and Kumar R.S.: Review of Multi-Class Classification Algorithms. International Journal of Pure and Applied Mathematics 118(14), 17–26 (2018).
12. Ghareb, A.S., Bakar, A.A. and Hamdan, A.R.: Hybrid Feature Selection based on enhanced Genetic Algorithm for Text. Elsevier Journal for Expert Systems with Applications 18, 21-44 (2015).
13. Grandini, M., Bagli, E. and Visani, G.: Metrics for Multi-Class Classification: An Overview. Computer Science, Mathematics, arXiv: 221112671, (2020).
14. Horn, D., Demircioglu, A., Bischl, B., Glasmachers, T. and Weihs, C. A comparative study on large scale kernelized support vector machines. Journal of Advances in Data Analysis and Classification. Springer Springer-Verlag Berlin Heidelberg (2016).
15. Jayadeva, Khemchandani, R. & Chandra, S. Twin Support Vector Machines for Pattern Classification, IEEE Trannsaction on Pattern Analysis and Machine Intelligence, 29 (5): 905-10. (2007)
16. Ju, X., Tian Y., Liu, D. and Qi, Z. Nonparallel Hyperplanes Support Vector Machine for Multi-class Classification. Elsevier: International Conference On Computational Science (51), pp. 1574-1582. (2015)
17. Kumar, A., Dabas, V. and Hooda, P. Text classification algorithms for mining unstructured data: a SWOT analysis. International Journal of Information Technology. DOI: https://doi.org/10.1007/s41870-017-0072-1. (2017).
18. Kuo, K.M., Yalley, P., Kao, Y. and Huang, C.H.: A Multi-Class Classification Model for Supporting the Diagnosis of Type II Diabetes Melitus. PeerJ, 8, e9920, (2020).
19. Lausser, L., Schmid, F., Schirra, L-R., Wilhelm, A.F.X. and Kestler, H.A. Rank-based Classifier for Extremely High-dimensional Gene Expression Data. Journal of Advances in Data Analysis and Classification. Springer-Verlag Berlin Heidelberg. (2016).
20. Li, L., Zhao, K., Sun, R., Gan, J., Yuan, G. and Tong, L. Parameter-Free Extreme Learning Machine for Imbalanced Classification. Journal of Neural Processing Letters. Springer Science+Business Media, LLC, part of Springer Nature (2020).
21. Matthew, B. and Sohini, R.: A Generalized Flow for Multi-class and Binary Classification Tasks: An Azure ML Approach. IEEE International Conference on Big Data (Big Data), 1728-1737 (2015).
22. Pei, H., Lin, Q., Yang, L. and Zhong, P. A novel semi-supervised support vector machine with asymmetric squared loss . Journal of Advances in Data Analysis and Classification. Springer Nature 2020: Springer-Verlag GmbH Germany. (2020).
23. Rathgamage, D. and Duleep, P.W.: Multiclass Classification Using Support Vector Machines. Electronic Theses and Dissertations, 1845, (2018).
24. Rathgamage, D. and Iacob, I.E.: DCSVM: Fast Multi-class Classification using support Vector Machines. Springer: International Journal of Machine Learning and Cybernetics, (2019).
25. Saigal, P. and Khanna, V.: Multi-category News Classification using Support Vector Machine based Classifiers. Springer Nature Applied Sciences, 2 (3), (2020).
26. Salter-Townshend, M. and Murphy T.B. Mixtures of biased sentiment analysers. Journal of Advances in Data Analysis and Classification. Springer-Verlag Berlin Heidelberg. (2013).
27. Sueno, H.T., Gerardo, B.D. and Medina R.P.: Multi-class Document Classification using Support Vector Machine (SVM) Based on Improved Naïve Bayes Vectorization Technique. International Journal of Advanced Trends in Computer Science and Engineering, 9 (3), 3937-3944, (2020).
28. Yaakub, M.R., Latiffi, M.I.A. and Zaabar, L.S.: A Review on Sentiment Analysis Techniques and Applications. IOP Conference Series: Materials Science and Engineering, 55, (2019).
29. Yang, L. and Wang, L. A class of semi-supervised support vector machines by DC programming. Journal of Advances in Data Analysis and Classification. Springer-Verlag Berlin Heidelberg. (2013).
30. Zhao, X., Barber, S., Taylor, C. C., and Milan, Z. Interval forecasts based on regression trees for streaming data. Journal of Advances in Data Analysis and Classification. Springer Nature 2019: Springer-Verlag GmbH Germany. (2019).
31. Zhu, W. and Zhong, P. Minimum Class Variance SVM+ for data classification. Journal of Advances in Data Analysis and Classification. Springer Springer-Verlag Berlin Heidelberg (2015).

Nur Suhailayani Suhaimi a lecturer (PhD) at Universiti Teknologi MARA, Malaysia. One of research member in Data Mining and Optimization Lab of UKM. Area of interests are optimization, Sentiment Analysis and Data Mining.

Zalinda Othman an associate professor (PhD) at Universiti Kebangsaan Malaysia. An active senior lecturer in teaching and research for Faculty of Information Science and Technology. One of research member in Data Mining and Optimization Lab of UKM. Area of interests are optimization, IT in Manufacturing and Data Mining.

Mohd Ridzwan Yaakub an associate professor (PhD) at Universiti Kebangsaan Malaysia. An active senior lecturer in teaching and research for Faculty of Information Science and Technology. Leader of Sentiment Analysis Research Lab. Area of interests are Sentiment Analysis, Opinion Mining and Artificial Intelligence.