IJCSC 0973-7391 Volume 9 • Issue 1 Sept 2017 - March 2018 pp. 34-41 available online at www.csjournals.com

Incremental Learning with Feature Shift Detection for Personalized E-mail Spam Filtering

Gopi Sanghani[1], Dr. Ketan Kotecha[2]
[1]Computer Engineering Department, Nirma University, Ahmedabad-382481, India.
[2]Parul University, Waghodia, Vadodara-391760, India.
gopisanghani@gmail.com, drketankotecha@gmail.com

Abstract: An ever increasing ratio of spam e-mails over legitimate e-mails and adversarial nature of spam e-mails lead to the requirement of employing spam filter that can be updated dynamically. Moreover, the discrimination criteria of spam and legitimate e-mails may vary for different users. This leads to the personalization of e-mail spam filter which automatically adapts individual user's characteristics. We present an incremental learning model using support vector machine for personalized e-mail spam filtering. We propose a novel feature distribution shift detection function to determine the necessity of updating the feature set and identifying new features with higher discriminating ability from an incoming set of e-mails. The proposed classifier is evaluated on ECML datasets. The results confirm the applicability of our model in the presence of feature distribution shift. Our incremental model achieves superior performance over a conventional batch learning model.

Keywords: Feature distribution shift, Incremental Learning, Personalized e-mail spam filter, Support Vector Machine.

1. Introduction

One of the major parameters in the success of Artificial Intelligence applications in real life is adaptability of algorithms towards incremental learning process. In a learning process, the current knowledge base is updated continually or at a regular interval of time that may modify the state of the current decision model. This requirement leads to the inclusion of incremental learning process with conventional learning algorithms. Incremental learning is a machine learning paradigm where the current model is modified whenever new example(s) appear over a period of time and contribute some different knowledge to the existing hypothesis. A very significant conventional formulation of machine learning algorithms is classification problems. Machine learning approaches for classification follow the generation of a discriminative model which learns from the available knowledge (the training set) and applies the learned hypothesis to classify unseen data (the test set). The ability to include additional training data when it becomes available and to re-learn from them is a prominent feature of the incremental learning process. In many of the real life classification problems, the input is given in the form of data streams, which eventually lead to the application of incremental learning process. The data stream classification techniques have to deal with the dynamically changing data distributions which cause the phenomenon of distribution shift or concept drift. In real concept drift the input data distribution i.e. feature distribution may not change, only the class label changes. And the virtual concept drift occurs as a result of the change in the underlying data distribution or unavailability of training data [11, 28].

In this research, we present an incremental learning framework for personalized e-mail spam filter with dynamic feature shift detection. Machine learning algorithms are extensively used in text classification problems. Personalized e-mail spam filter is the most prevalent application of automated binary text classification problems. Given a training set of n labeled sample e-mails T = {t1, t2, …, tn} and C = {cl, cs} denotes e-mail categories: legitimate and spam. The task is to learn discriminant function, which classifies previously unseen e-mails into one of the two categories based on their content. E-mail remains to be highly formalized, official and indispensable communication medium even after increasing use of social networking applications. However, the inevitable downside of it is a continuously growing ratio of unwanted and useless e-mails. A huge number of unsolicited e-mails are delivered to the internet users regardless of a personal or commercial level of interests. Content based e-mail spam filter is one of the highly challenging tasks due to two major reasons, the content of spam e-mails changes over a time, as a result of which feature distribution shift occurs in spam e-mails and differing criteria of individual users for spam and legitimate e-mail discrimination. The research work presented here addresses these challenges by applying incremental learning model with detection of feature distribution shift. The major outline of our work is given as follows:

1. To propose an incremental learning framework using support vector machine.

2. To propose a novel feature distribution shift detection function based on normalized average discriminative weight.

The rest of the paper is organized as follows. In Section 2 we summarize related work on personalized e-mail spam filter, incremental learning and concept drift detection. We present our incremental framework with dynamic feature shift detection for personalized e-mail spam filter algorithm in Section 3. Section 4 describes the in-depth experiments we perform and the discussion on results obtained. Section 5 concludes the paper with an insight into the future work.

2. Related work

For a personalized e-mail spam filtering, a filter is employed by a single user as a client-side filter and messages identified as spam are usually sent to spam folder. Filtron [5], a personalized anti-spam filter based on machine learning text categorization paradigm, had been evaluated in real life scenario that confirmed the prominent role of machine learning techniques for anti-spam filtering. Gray & Haahr [1] present the concept of personalized, collaborative spam filtering named as CASSANDRA architecture. Cheng & Li [29] proposed combined supervised and semi-supervised Classifier using SVM for Personalized Spam Filtering. In the Two-tier Spam Filter Structure presented by Teng & Teng [27], e-mails classified as legitimate mails by the legitimate mail filter may pass, while the remaining e-mails are processed by the spam filter in an ordinary way. Chang, Yih, & McCann [17] designed a light-weight user model that is highly scalable and can be easily combined with a traditional global spam filter. Youn & McLeod [26] proposed an efficient spam e-mail filtering method using ontologies in which user profile ontology creates a blacklist of contacts and topic words. A personalized spam filter is presented by Junejo & Karim [16] using an automatic approach which built a statistical model of spam and non-spam words from the labeled training dataset and then updates it in two passes over the unlabeled individual user's inbox. The filter presented by Shams & Mercer [24] using natural language attributes, the majority of them being connected to stylometry aspects of writing.

Ghanbari & Beigy [4] proposed the algorithm called incremental RotBoost, an incremental learning algorithm based on ensemble learning. Hsiao and Chang [30] developed incremental cluster-based classification method, called ICBC with two phases. In the first phase, it clusters emails in each given class into several groups, and an equal number of features are extracted from each group. In the second phase, ICBC is capacitating with an incremental learning mechanism that can adapt itself to accommodate the changes of the environment in a fast and low-cost manner. Georgala, Kosmopoulos, and Paliouras [15] proposed Active Learning Approach using Incremental Clustering for spam filtering. The user provides the correct categories (labels) for the messages of the first batch and from then on the algorithm decides when to ask for a new label, based on a clustering of the messages that is incrementally updated. Taninpong and Ngamsuriyaroj [23] proposed an incremental spam mail filtering using Naïve Bayesian classification in which the sliding window concept is applied to keep the training set to a limited size and the training set is updated when new emails are received. In effect, features in the training set are incrementally updated so, the model would be adaptive to a new spam pattern. Incremental training algorithm using SVM has been successfully evaluated for personalized spam filtering in [7].

The performance of content based classification algorithms strongly relies upon the selection of most relevant and discriminative features. Over a time, relevancy of features varies in the presence of concept drift. Katakis et al. [9] proposed the incremental feature selection and feature based classification for textual data streams. The work of Delany et al. [25] present spam filtering system that uses example based machine learning techniques to track concept drift. In [10] katakis et al. presented ensemble classifiers to track recurring contexts which occur in e-mail spam filtering. Gomes et al. [12] presented a data stream classification system to address the challenges of learning recurring concepts in a dynamic feature space. Henke et al. [19] analyzed the evolution of features in the presence of concept drift for spam detection. Kmieciak and Stefanowski [18] introduced an approach to detect drift in unlabeled data and retrain a classifier using a limited number of labeled examples. Sheu et al. [14] presented a window based incremental learning technique for tracking concept drift in spam filtering by checking only header section of e-mails.

3. Proposed algorithm for feature shift detection in incremental learning

Over the past few years, many efficient e-mail spam filters are built and applied to block spam e-mails. Traditional server based spam filters are trained on generic mail corpus and then, commonly applied to a user's inbox to detect spam and legitimate e-mails. Many mail boxes facilitate user defined settings for the categorization of the important e-mails. Still, the end user remains highly dependent on the discrimination of e-mails characterized by the general training corpus. The essential advantage is users are relieved from the burden of processing thousands of unsolicited e-mails. But only global filters cannot optimally reflect individual user's characteristics while discriminating e-mails. As an extensive model, the personalized e-mail spam filter is required which facilitate robustness and should be adaptive to individual user's preferences. The discrimination criteria tend to change over a period of time. Also, the content pattern of spam e-mails can be described by the nature and appearance characteristics of e-mails. It can be characterized as regularly appearing spam e-mails, appearing for a short duration of time and appearing at some interval i.e. recurring context. So, there is a need to update the filter dynamically to tackle feature distribution shift that occurs because of adversarial patterns of spam e-mails.

3.1 Incremental Learning with SVM

Conventional learning models follow the assumption that training and test data observe the same statistical distribution of data. But in the presence of distribution shift, the decision model learned from the training data may not appropriately classify the new examples and so, the classifier performance is degraded. The increased error rate indicates the requirement of updating the classification model. Our incremental learning framework enables the classifier to learn new information derived from the incoming set of examples and at the same time, it holds the previously acquired knowledge. Our algorithm is developed as follows: the filtering process is carried out over three passes. The first pass is performed using conventional batch training, with n labeled examples, that generates the discriminant function F(x). Pass II comprises a series of testing phases in which small batches of incoming unlabeled e-mails are given to identify true labels. Pass III is carried out by activating incremental retraining whenever the performance criteria are violated. In order to handle the feature distribution shift, we propose a novel feature shift detection function to be applied before activating an incremental training, which is explained in the subsequent subsection. The flow of our incremental algorithm is presented in fig. 1.

Many content based spam filters apply machine learning techniques, of which support vector machines have shown consistently superior performance. SVM was initially applied for spam categorization by Drucker, Wu, and Vapnik [8]. Since then various extensions and online and active learning approaches have been presented by many researchers because of SVM's good generalization ability and higher classification accuracy. Support vector machines [3] are supervised machine learning algorithm also known as optimal margin classifiers. SVM algorithm maps input vectors into a feature space of higher dimension and constructs an optimized hyper plane for generalization. In binary classification problem, a data set X contains n labeled example vectors {(x1, y1) . . . (xn, yn)}, where xi be the input vector in the input space, with corresponding binary labels yi ϵ {-1, 1}. Let φ(xi) be the corresponding vectors in feature space, where φ(xi) is the implicit kernel mapping and let k(xi, xj) = φ(xi) . φ(xj) be the kernel function, implying a dot product in the feature space. The optimization problem for a soft-margin SVM is,

$$\min_{w,b}\left(0.5\|w\|^2 + C\sum_i \zeta_i\right) \quad (1)$$

Subject to the constraints yi (w.x + b) = 1 − ξi and ξi ≥ 0 where w is the normal vector of the separating hyper plane in feature space and C > 0 is a regularization parameter controlling the penalty for misclassification. Equation (1) is referred to as the primal equation. From that, the Lagrangian form of the dual problem is:

$$w(\alpha) = \max_\alpha\left(\sum_i \alpha_i - 0.5\sum_{i,j}\alpha_i\alpha_j y_i y_j k(x_i, x_j)\right) \quad (2)$$

Subject to 0≤αi≤ C. This is a quadratic optimization problem that can be solved efficiently using algorithms such as Sequential Minimal Optimization (SMO) [13]. Many αi go to zero during optimization and the remaining xi corresponding to those αi > 0 are called support vectors. If l is the number of support vectors and αi > 0 for all i, with this formulation, the normal vector of the separating plane w is calculated as:

$$w = \sum_i \alpha_i y_i x_i \quad (3)$$

The classification f(x) for a new sample vector x can be determined by computing the kernel function of x with every support vector:

$$f(x) = sign\left(\sum_i \alpha_i y_i k(x, x_i) + b\right) \quad (4)$$

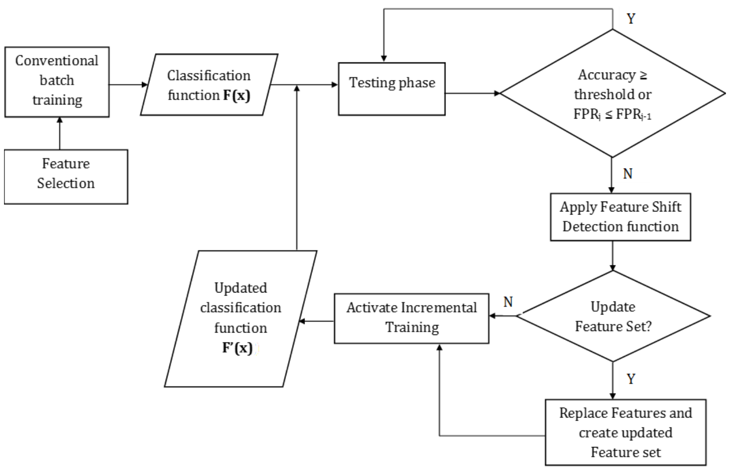

Figure 1. Incremental model with feature shift detection for personalized e-mail spam filtering

Syed et al. [20, 21] proposed the framework for incremental learning of support vector machine with a new batch of data and set of support vectors, which precisely represents the separating hyper plane. In this research, we present the incremental algorithm by applying periodic retraining of support vector machine with a small number of additional training examples and an existing support vector set. An important property SVM possesses is, a set of support vectors represents the feature space and class boundaries in a very concise manner. So, incremental SVM can be trained by preserving support vectors and adding them to the next batch of incoming examples [2, 22].

3.2. Feature shift detection and feature set update

Due to a constant change in the spam patterns, the classification model requires the dynamic updates to maintain and upgrade the filter performance. Our incremental learning framework incorporates the feature shift detection function before activating re-training of a classifier. This function identifies the necessity of updating the feature set, also it derives new features to be included in the original feature set. Thereby reducing the overhead which may otherwise be caused every time before re-training. The feature shift detection & feature set update algorithm is described in fig. 2. The algorithm generates a new subset of features from the re-training set of spam and legitimate e-mails with true labels. From this set distinct features are found with their respective discriminative weight. Each such distinct feature's discriminative weight is compared with the average discriminative weight avDm of the feature set used during previous training of the classifier. If features with higher discriminative weight than avDm exist than the current feature set is updated as shown in step 5 in table 1. Modifying the feature set causes the classifier to effectively re-learn the change of data distribution from the new set of examples. In the case of e-mails, spammers also try to elude filters and frequently change spammy features, though over a long span of time traditional features and latest features both are essential for efficient filtering. Updating feature space before activating incremental training is required to include new features with higher discriminating ability. The re-training with updated feature set results in modified discriminant function F'(x), which further classifies the incoming examples.

4. Experiments and Results

The proposed incremental learning model using SVM for personalized e-mail spam filter has been evaluated by performing detailed experiments on ECML datasets [6]. Support vector machine is a discriminative classifier that directly learns the boundary between classes. It learns a decision boundary that maximizes the distance between samples of the two classes.

Input:

1. Current feature set $FS_{current}$
2. Re-training set $RS_{i+1} = MCM_i \cup SV_i \cup TS_k$

   Where,

   $MCM_i \leftarrow$ miss-classified e-mails with true labels

   $SV_i \leftarrow$ Set of support vectors

   $TS_k \leftarrow$ Set of testing e-mails for which incremental training is to be activated

Output: Updated feature set $FS_{updated}$

1 Initialize $FS_{updated} = FS_{current}$

2 Generate new subset of features FS_{new} from retraining set RS_{i+1}

3 Calculate the average discriminative weight per feature *avDm* from $FS_{current}$

4 Find each distinct feature F_d from FS_{new} such that,

$$F_d \in FS_{new} \setminus FS_{current}$$

Add the feature to the another set call distinct feature set $FS_{distinct}$ as follows:

If $dm(F_d) \geq avDm$

then $FS_{distinct} = FS_{distinct} \cup F_d$

5 If $FS_{distinct} \neq$ NULL then update $FS_{updated}$ as,

For each feature F_j in $FS_{distinct}$

$FS_{updated} = FS_{updated} \cup F_j$

$FS_{updated} = FS_{updated} - \{F_k \mid F_k$ has the lowest discriminative weight$\}$

6 Return $FS_{updated}$

Figure 2. Algorithm for Feature Shift Detection and Feature Set Update

4.1. Experimental Data and Implementation Detail

The ECML task-A and task-B datasets were made publically available during 2006 ECML-PKDD Discovery Challenge. The dataset contains the processed form of e-mails where each feature in an e-mail is represented by an id-count pair. Features are extracted in form of token id and email is represented as a feature vector with term frequencies. We apply sequential minimal optimization (SMO) algorithm for training SVM. SMO algorithm is designed and used to avoid solving a quadratic programming (QP) problem of SVM model. SMO algorithm follows decomposition approach and at every step, it analytically updates the coefficient of two vectors selected heuristically. To find the discriminative weight of features we use Information Gain [31] feature selection function.

4.2. Results and Discussions

Experiments on ECML datasets are conducted to analyze the incremental personalized e-mail spam filter performance in the presence of distribution shift. ECML task-A dataset contains 4000 emails with 2000 spam and 2000 legitimate emails. The size of training dataset ECML task-B is 100 only with 50 spam and 50 legitimate emails. Test data for task-A contains 3 user inboxes with 2500 emails each. And task-B test data consists of 15 different user inboxes with 400 emails each. Both ECML task-A and task-B training and test datasets follow the similarity in the proportion of spam and legitimate e-mails only. The source of data in training and testing follow a different distribution. We conduct our experiments in two parts. The first part focuses on observing the behavior of the incremental learning model over a conventional batch learning model. Generally, the classifier learns the statistical distribution of data during training and the discriminant function is generated. The discriminant function can efficiently classify un-labeled e-mails from testing data till the test data follows the same distribution. When distribution shift occurs, the discriminant function is updated by activating incremental training to maintain and improve the filter performance. In ECML Datasets training data and testing data follow a completely different distribution. So, SVM's conventional training results into a static filter, which perform less efficiently in classifying the test datasets. But if the filter is incrementally updated by adding a small number of additional examples taken from test data, the filter performance is significantly improved.

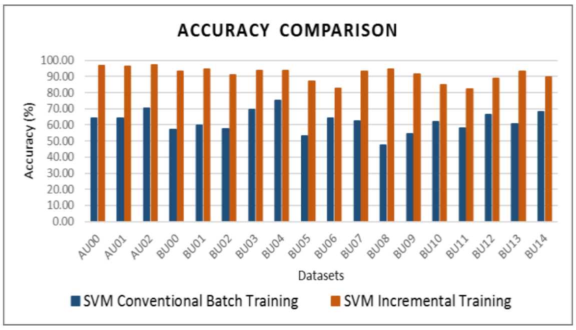

Figure 3(a). ECML Dataset - Classification Accuracy Results

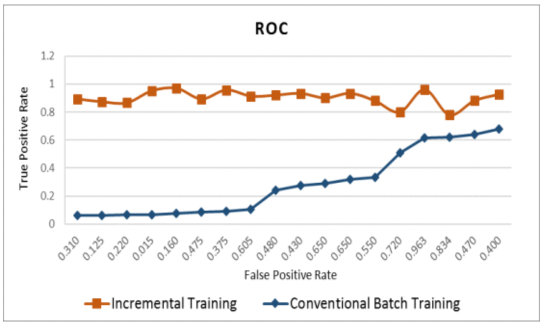

Figure 3(b). ECML Dataset - ROC

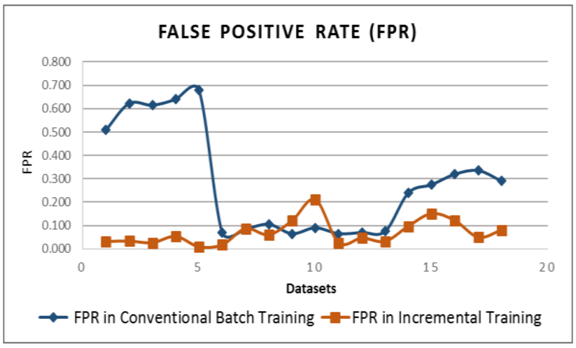

Figure 4(a). ECML Dataset - False positive rate Results

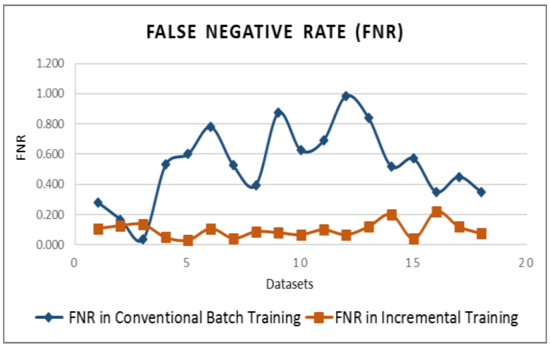

Figure 4(b). ECML Dataset - False negative rate Results

Fig. 3(a) shows the comparison of accuracy in batch learning and incremental learning model using SVM. Another traditional characteristic usually requires the comparatively large number of training examples to train discriminative function completely. ECML task-B contains only 100 training examples though; the incremental training enables the filter to substantially improve the results.The second part of an experiment is carried out with an aim of detecting the feature shift distribution that occurs when the distribution of data is non-stationary and the test data is generated from different underlying distributions than the training data. The discriminative models have strong discrimination ability but their modeling capability is limited since they are focusing on classification boundaries as in SVM. In ECML datasets test datasets are taken from individual user's mailbox. The feature space representing each individual test data folder is different than the feature space generated from training data. Therefore, the need arises to personalize the filter and to retrain it in an updated feature space. Our proposed feature shift detection function determines the necessity for updating feature set and identifies new features with higher discriminative ability. These new features are included in the feature set by replacing the old features with lowest discriminative weight in order to keep the same feature dimension.

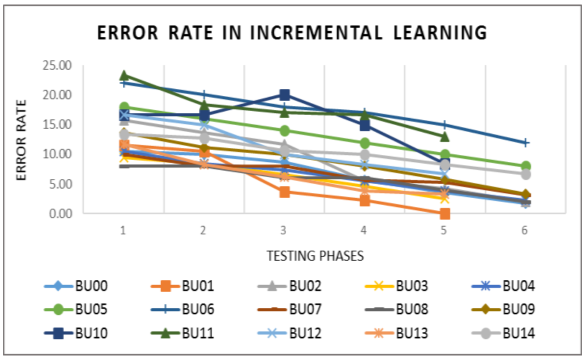

Figure5 (a). Error Rate for ECML – B dataset

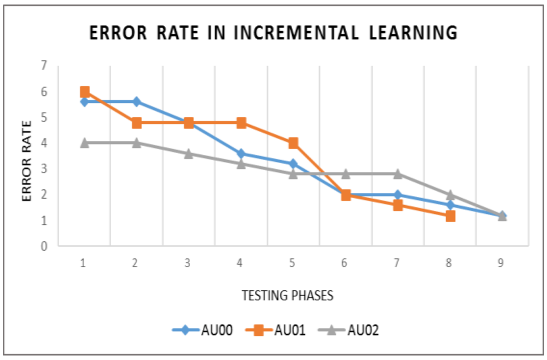

Figure5 (b). Error Rate for ECML – A dataset

Table 1: ECML Datasets Results

Dataset	**Inbox**	**SVM Incremental Training with updated Feature Set**			**SVM Conventional Batch Training**		
		Precision	Recall	F1 Measure	Precision	Recall	F1 Measure
ECML TASK-A	USER00	0.97	0.95	0.96	0.59	0.72	0.65
	USER01	0.95	0.94	0.95	0.57	0.83	0.68
	USER02	0.96	0.96	0.97	0.61	0.96	0.75
ECML TASK-B	USER00	0.94	0.92	0.93	0.42	0.47	0.45
	USER01	0.97	0.92	0.95	0.37	0.40	0.38
	USER02	0.94	0.83	0.88	0.76	0.22	0.34
	USER03	0.92	0.96	0.94	0.85	0.48	0.61
	USER04	0.94	0.92	0.93	0.85	0.61	0.71
	USER05	0.87	0.86	0.87	0.66	0.13	0.21
	USER06	0.80	0.86	0.83	0.81	0.38	0.51
	USER07	0.97	0.89	0.93	0.83	0.31	0.45
	USER08	0.95	0.93	0.94	0.18	0.02	0.03
	USER09	0.96	0.91	0.94	0.68	0.16	0.26
	USER10	0.93	0.97	0.95	0.67	0.48	0.56
	USER11	0.91	0.80	0.85	0.61	0.43	0.50
	USER12	0.89	0.88	0.88	0.67	0.65	0.66
	USER13	0.95	0.91	0.93	0.62	0.55	0.58
	USER14	0.93	0.83	0.88	0.69	0.65	0.67

The second part of an experiment is carried out with an aim of detecting the feature shift distribution that occurs when the distribution of data is non-stationary and the test data is generated from different underlying distributions than the training data. The discriminative models have strong discrimination ability but their modeling capability is limited since they are focusing on classification boundaries as in SVM. In

ECML datasets test datasets are taken from individual user's mailbox. The feature space representing each individual test data folder is different than the feature space generated from training data. Therefore, the need arises to personalize the filter and to retrain it in an updated feature space. Our proposed feature shift detection function determines the necessity for updating feature set and identifies new features with higher discriminative ability. These new features are included in the feature set by replacing the old features with lowest discriminative weight in order to keep the same feature dimension.
Classification results show that incremental training of SVM on a set of support vectors and a new set of examples allows obtaining and significantly improving the accuracy of the filter. This also confirms the role of support vectors for incrementally training SVM by using the old solution as a starting point to find an updated solution. Table 1 shows precision, recall and F1 measure, the most common performance measures in binary classification. Fig. 3(a) shows the comparison of accuracy achieved in incremental training and conventional batch training of SVM. Fig. 3(b) shows ROC curve for comparison of true positive rate (TPR) vs. false positive rate (FPR). The occurrence of False Positives (FP), i.e. legitimate emails classified incorrectly as spam degrades the filter performance. An FP is significantly more harmful than a False Negative (FN) i.e. a spam email incorrectly classified as legitimate. Very low FPR is achieved in incremental SVM learning. Fig. 4(a) and (b) show the precise comparison for FPR and FNR in both the incremental and batch learning models. Fig. 5(a) and (b) present the error rate in incremental learning model. The error rate decreases over time as a number of testing phases increases.

5 Conclusion and Future work

Email spam filtering on a personalized level has been one of the most challenging classification tasks in the presence of distribution shift. We employed incremental learning model so, the discriminant function learns the modified distribution of data and the filter is updated dynamically. The proposed incremental learning algorithm outperforms over a batch learning model and achieved very low false positive rate. Also, the feature distribution shift detection function effectively determines when to update the feature set and new features with higher discriminative ability. Experimental results confirm the applicability of our unique approach using incremental learning of SVM with the heuristically updating feature set for improving the efficiency and consistently maintaining the performance of the classifier. The future work addresses to develop an algorithm which derives a feature discrimination weight and continually monitors the change in feature patterns for the prediction of a shift.

Acknowledgment

We are thankful to the Nirma University for providing resources and other facilities to carry out this research work.

References

[1]. A. Gray and M. Haahr, "Personalised, collaborative spam filtering", in Proceedings of 1st Conference on Email and Anti-Spam, 2004.

[2]. A. Shilton and M. Palaniswanmi, "Incremental learning of Support Vector Machines", IEEE transactions on neural networks, vol. 16, pp. 456-61, 2005.

[3]. C. Cortes and V. Vapnik, Support-vector networks. Machine Learning, 20, 273–297, 1995.

[4]. E. Ghanbari and H. Beigy, "Incremental RotBoost algorithm: An application for spam filtering". Journal Intelligent Data Analysis archive, 19 (2), 2015, pp. 449-468, IOS Press Amsterdam, The Netherlands.

[5]. E. Michelakis, I. Androutsopoulos, G. Paliouras, G. Sakkis, and P. Stamatopoulos, "Filtron: a learning-based anti-spam filter", In Proc.1st Conf. on Email and Anti-Spam; 2004.

[6]. ECML-PKDD: Discovery challenge. http://www.ecmlpkdd2006.org/challenge.html (2006).

[7]. G. Sanghani and K. Kotecha, "Personalized spam filtering using incremental training of support vector machine", International Conference on Computing, Analytics and Security Trends (CAST), Pune, pp. 323-328. 2016.

[8]. H. Drucker, D. Wu, and V. Vapnik, "Support vector machines for spam categorization", IEEE Transactions on Neural Networks, 10(5), 1048–1054, 1999.

[9]. I. Katakis, G. Tsoumakas, and I. Vlahavas, "On the utility of incremental feature selection for the classification of textual data streams". In 10th Panhellenic Conference on Informatics (PCI 2005), pages 338-348. Springer-Verlag.

[10].I. Katakis, G. Tsoumakas, and I. Vlahavas, "Tracking recurring contexts using ensemble classifiers: An application to email filtering", Knowledge and Information Systems, Springer London; 22(3): 371–391, 2010. IEEE Transactions on Neural Networks and Learning Systems, 25(1):95–110, 2014.

[11].J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, "A survey on concept drift adaptation". ACM Computing Surveys (CSUR), 46 (2014) 1-37.

[12].J. Gomes, M. Gaber, P. Sousa, and E. Menasalvas. Mining recurring concepts in a dynamic feature space.

[13].J. Platt, "Fast training of support vector machines using sequential minimal optimization," in Advances in Kernel Methods - Support Vector Machines, MIT Press, Cambridge, pp. 185-208, 1999.

[14].J-J Sheu, K-T Chu, N-F Li, and C-C Lee, "An efficient incremental learning mechanism for tracking concept drift in spam filtering". PLoS ONE12(2): e0171518. 2017. https://doi.org/10.1371/journal.pone.0171518.

[15].K. Georgala, A. Kosmopoulos, and G. Paliouras, "Spam Filtering: an Active Learning Approach using Incremental Clustering". 2014, WIMS Thessaloniki, Greece Copyright is held by the owner/author(s).Publication rights licensed to ACM.

[16].K. Junejo, A. Karim, "Robust personalizable spam filtering via local and global discrimination modeling", Knowledge and Information Systems, Vol. 34, No. 2, pp. 299-334, 2013.

[17].M. Chang, W. Yih, and R. McCann, "Personalized spam filtering for gray mail". In CEAS-08: Proceedings of 5th Conference on Email and Anti-Spam, 2008.

[18].M. Kmieciak and J. Stefanowski, "Handling sudden concept drift in Enron message data streams," Control and Cybernetics, vol. 40, no. 3, pp. 667–695, 2011.

[19].M´arcia Henke, Eduardo Souto, and Eulanda M dos Santos, "Analysis of the evolution of features in classification problems with concept drift: Application to spam detection," in IFIP/IEEE International Symposium on Integrated Network Management (IM). IEEE, 2015, pp. 874–877.

[20].N. Syed, H. Liu, and K. Sung, "Handling concept drift in incremental learning with support vector machines", In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99), San Diego, CA, USA.

[21].N. Syed, H. Liu, and K. Sung, "Incremental learning with support vector machines". In Proceedings of the Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJCAI-99), Stockholm, Sweden.

[22].P. Laskov, C. Gehl, S. Krüger, and K-R. Müller, "Incremental support vector learning: analysis, implementation and applications". J Mach Learn Res 7: 1909–1936, 2006.

[23].Phimphaka Taninpong and Sudsanguan Ngamsuriyaroj, " Incremental Adaptive Spam Mail Filtering Using Naïve Bayesian Classification". Proc. 10th ACIS International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009, pp. 243-248.

[24].R. Shams and R. Mercer, "Personalized spam filtering using natural-language attributes", 12th IEEE International Conference on Machine Learning Applications, Florida, USA, 2013.

[25].S. J. Delany, P. Cunningham, A. Tsymbal, and L. Coyle, "A case-based technique for tracking concept drift in spam filtering". Knowledge-Based Systems, 18(4–5):187–195, 2005.

[26].S. Youn, and D. McLeod, "Spam decisions on gray e-mail using personalized ontologies", SAC, page 1262-1266, ACM, 2009.

[27].Teng Wei-Lun and Teng, Wei-Chung, "A personalized spam filtering approach utilizing two separately trained filters", in WI-IAT 08: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society

[28].Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. Machine Learning, 23(1), 69-101.

[29].V. Cheng, C. Li, "Personalized spam filtering with semi-supervised classifier ensemble", in WI-06: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society, pages 195–201, 2006.

[30].W. -F. Hsiao, and T.-M. Chang, "An Incremental Cluster-Based Approach to Spam Filtering". Expert Systems with Applications, 34(3), 1599-1608, 2007.

[31].Y. Yang and J. Pedersen, "A comparative study on feature selection in text categorization", in Proceeding ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning, Morgan Kaufmann Publishers, San Francisco, USA, pp. 412-420, 1997.