SEQUENTIAL HETEROGENEOUS FEATURE SELECTION FOR MULTI–CLASS CLASSIFICATION: APPLICATION IN GOVERNMENT 2.0

Imara Nazar* *Yasitha Warahena Liyanage** *Daphney–Stavroula Zois** *Charalampos Chelmis*†

*Electrical and Computer Engineering Department
†Computer Science Department
University at Albany, SUNY, Albany, NY, USA
Emails: {imohamednazar, yliyanage, dzois, cchelmis}@albany.edu

ABSTRACT

Herein, the problem of multi–class classification of participatory civil issue reports in crowdsourcing platforms is addressed. Specifically, an efficient method is proposed to guide the selection of heterogeneous features, so as to account for different information facets of the reported issue. An optimization framework is devised to select the minimum number of informative features from each feature set, and switch between feature sets when deemed necessary to achieve an accurate classification decision. Evaluation on real–world data from SeeClickFix, a government 2.0 platform, shows the ability of the proposed framework to classify civil issue reports with up to 92.6% accuracy, while using 99.82% less features than the state–of–the-art.

Index Terms— multiple feature sets, optimum feature selection, e–government

1. INTRODUCTION

Government 2.0 applications have facilitated government–citizen interactions by providing easy access and increasing transparency to public services through platforms such as SeeClickFix [1] in US, FixMystreet [2] in UK, Nonoville and IMcity [3] in Greece. According to the United Nations E–Government database, popularity of such platforms has increased significantly between 2008 and 2018 [4]. In order to ensure that the needs of every concerned citizen are met, dedicated officials have to manually assess issue reports and take action when necessary. A scalable automatic multi–class classification framework could be used to help officials in their decision–making process.

In this paper, we propose a framework to address the challenging problem of automatic multi–class classification of civil issue reports using multiple feature sets. Specifically, a reported civil issue may comprise fields such as title, description, tags or images, leading to a set of heterogeneous features extracted from textual and visual content, tags and/or metadata. Determining the class that each civil issue report belongs to requires evaluating the most informative features from each feature set. We formulate this task as a sequential hypothesis testing problem, where features are reviewed sequentially in each of the feature sets, by determining whether to continue the review process on the current feature set or not until the last feature set is reached. Our framework decides to stop and classify the issue when reviewing more features does not improve the accuracy of the classification decision. Our proposed algorithm is guaranteed to lead to an accurate decision by using the minimum number of features from each feature set specific to each civil issue report.

Automatic classification of civil issue reports has been explored recently, however, prior work has focused on binary [5–7] and multi–class classification [8], and urgency estimation [9, 10] of civil issue reports based on homogeneous features. In our prior work [11], we have used multiple feature sets but for binary classification. Extending this line of work to the multi–class setting is non–trivial. Beyond civil issue report classification, feature selection methods [12–16] produce a fixed set of features to be used during classification, common to all data instances. Recently, it has been shown that this approach is sub–optimal [8]. At the same time, selecting different subset of features from different sets per data instance has the potential to improve classification interpretability.

The rest of the paper is organized as follows. In Section 2, a stochastic optimization problem is defined to mathematically describe the task at hand. Section 3 derives the solution to this problem, while Section 4 discusses our proposed algorithm. Section 5 provides experimental results on a real–world dataset from the SeeClickFix platform. Section 6 concludes the paper, and discusses future research directions.

2. PROBLEM DESCRIPTION

We consider a set $\mathcal{Z}$ of civil issue reports, where each such report $z \in \mathcal{Z}$ is described by a vector $\mathbf{f} \triangleq [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_Q]^\top$,

This material is based upon work supported by the National Science Foundation under Grant No. ECCS–1737443.

978-1-7281-6662-9/20/$31.00

where Q denotes the total number of available feature sets (see also Fig. 1). Each $\mathbf{f}_q \triangleq [y_{q,1}, y_{q,2}, \dots, y_{q,N_q}]^\top, q \in \{1, \dots, Q\}$, represents a column vector of N_q features (e.g., intensity of image pixels, number of times a specific word appears in the description) in the qth feature set. We assume a total of $N = \sum_{q=1}^{Q} N_q$ features. Each issue z may belong to one of L possible hypotheses $H_i, i = 1, \dots, L$, with corresponding *a priori* probability p_i for each hypothesis. The conditional probability of the nth feature $y_{q,n}$ of the qth feature set under true hypothesis H_i is denoted by $p(y_{q,n}|H_i)$.

Further, we define the cost coefficient $c_{q,n} > 0$ to denote the time and effort spent reviewing the nth feature in the qth feature set. We consider switching costs denoted by coefficient $s_{q,q'} > 0$ to capture the cost of switching from feature set q to feature set q'. We also consider the misclassification cost $M_{ij} \geq 0$, which represents the cost of selecting hypothesis H_j when the true hypothesis is $H_i, i \neq j$.

Our proposed framework reviews features and feature sets sequentially. While reviewing features in a specific feature set, we select whether to continue and review the next feature or move to the next available feature set. For simplicity, we begin our review process from the first available feature set. Our proposed sequential review process comprises a collection $(R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)})$ of random variables, where $R_q \in \{1, 2, \dots, N_q\}, q \in \{1, 2, \dots, Q\}$, indicates the feature that the framework reviews last in the qth feature set. Further, random variable $D_{(R_1,\dots,R_Q)}$ represents the assignment of issue z to one of the hypotheses after the end of the review process. While reviewing features on the qth feature set, the decision to stop depends only on the accumulated information until R_q. Our objective is to find variables $R_1, \dots, R_Q$ and $D_{(R_1,\dots,R_Q)}$ to accurately classify each civil issue report z, while minimizing the cost incurred from reviewing features and switching between feature sets, i.e.:

$$\min_{R_1,\dots,R_Q,D_{(R_1,\dots,R_Q)}} J(R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)}), \tag{1}$$

where

$$J(R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)}) = E\Bigg\{ \sum_{q=1}^{Q} \sum_{n=1}^{R_q} c_{q,n} + \sum_{q=1}^{Q-1} s_{q,q+1} + \sum_{j=1}^{L} \sum_{i=1}^{L} M_{ij} P(D_{(R_1,\dots,R_Q)} = j, H_i) \Bigg\}. \tag{2}$$

The first two terms in Eq. (2) denote the cost of reviewing features belonging to different feature sets and the corresponding switching costs, whereas the last term corresponds to the average Bayes error of the classification strategy used. At this point, we consider the *a posteriori* probability vector $\pi_{q,n} \triangleq [\pi^1_{q,n}, \pi^2_{q,n}, \dots, \pi^L_{q,n}]$, where $\pi^i_{q,n} \triangleq p(H_i|y_{1,1}, \dots, y_{1,R_1}, y_{2,1}, \dots, y_{q,n})$ that corresponds to the sufficient statistic of the accumulated information. This posterior probability vector can be iteratively computed via

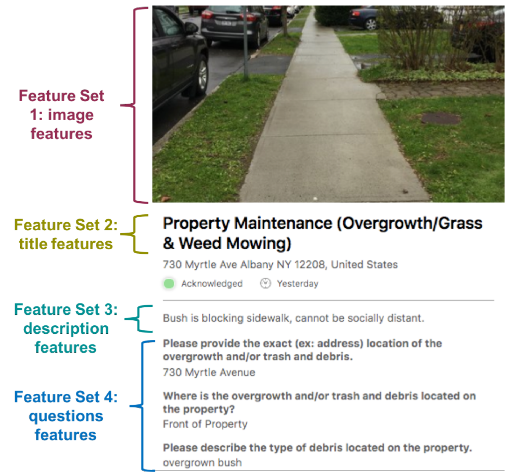

Fig. 1. Exemplary reported civil issue on the SeeClickFix platform; possible different feature sets (e.g., image, description) are also illustrated.

Bayes' rule. Specifically, after reviewing the nth feature from the qth feature set, the posterior probability vector $\pi_{q,n}$ is given by:

$$\pi_{q,n} = \frac{\pi_{q,n-1} \operatorname{diag}(\Delta_{q,n}(y_{q,n}))}{\pi_{q,n-1} \Delta^\top_{q,n}(y_{q,n})}, \tag{3}$$

where $\Delta_{q,n}(y_{q,n}) = [p(y_{q,n}|H_1), p(y_{q,n}|H_2), \dots, p(y_{q,n}|H_L)]$, and $\operatorname{diag}(\mathbf{v})$ denotes a diagonal matrix whose entries are the elements of vector $\mathbf{v}$. The a posteriori probability vector is initialized as $\pi_{1,0} = [p_1, p_2, \dots, p_L]$ and $\pi_{q+1,0} = \pi_{q,R_q}$ for $q \in \{1, \dots, Q-1\}$. Using the *a posteriori* probability vector $\pi_{q,n}$, we can rewrite the cost function in Eq. (2) in a more compact form. Namely, exploiting Eq. (3) and $x_{(R_1,\dots,R_Q)} = \sum_{n_1=1}^{N_1} \cdots \sum_{n_Q=1}^{N_Q} x_{(n_1,\dots,n_Q)} \mathbb{1}_{\{R_1=n_1,\dots,R_Q=n_Q\}}$, where $\mathbb{1}_A$ is the indicator function for event A (i.e., $\mathbb{1}_A = 1$ when A occurs, and $\mathbb{1}_A = 0$ otherwise), the average cost in Eq. (2) can be rewritten as follows:

$$J(R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)}) = E\Bigg\{ \sum_{q=1}^{Q} \sum_{n=1}^{R_q} c_{q,n} + \sum_{q=1}^{Q-1} s_{q,q+1} \Bigg\} + E\Bigg\{ \sum_{j=1}^{L} \pi_{Q,R_Q} M_j^\top \mathbb{1}_{\{D_{(R_1,\dots,R_Q)}=j\}} \Bigg\}, \tag{4}$$

where $M_j \triangleq [M_{1,j}, M_{2,j}, \dots, M_{L,j}]$.

3. PROPOSED FRAMEWORK

In our framework, classification of a civil issue report z is performed after inspecting all available feature sets. In order to acquire the optimum classification strategy $D_{(R_1,\dots,R_Q)}$, we search for a lower bound for the last term of Eq. (4) that does not depend on variables $R_1, R_2, \dots, R_Q$. Since $D_{(R_1,\dots,R_Q)}$

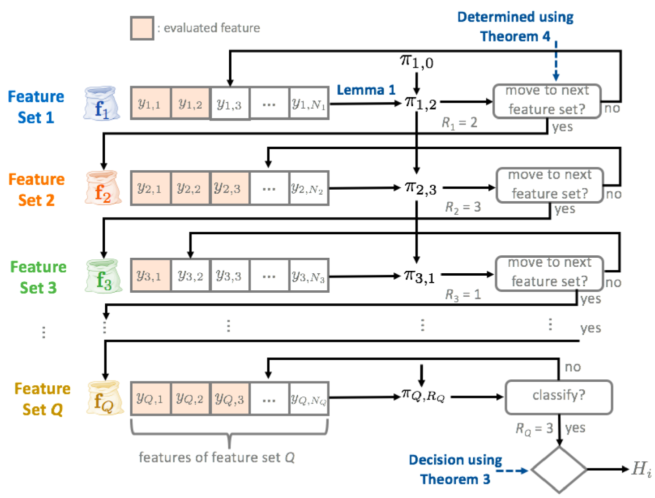

Fig. 2. Graphical representation of the proposed algorithm.

contributes only to this part of the average cost, the optimum classification strategy $D_{(R_1,\ldots,R_Q)}$ for fixed $R_1,\ldots,R_Q$ can then be found, as shown in Theorem 1.

Theorem 1 *For any classification strategy $D_{(R_1,\ldots,R_Q)}$ given variables $R_1, R_2, \ldots, R_Q$, it can be shown that:*

$$\sum_{j=1}^{L} \pi_{Q,R_Q} M_j^\top \mathbb{1}_{\{D_{(R_1,\ldots,R_Q)}=j\}} \geqslant g(\pi_{Q,R_Q}), \qquad (5)$$

where $g(\pi_{Q,R_Q}) \triangleq \min_{1\leqslant j\leqslant L} \pi_{Q,R_Q} M_j^\top$. The optimum strategy is then:

$$D^{optimal}_{(R_1,\ldots,R_Q)} = \arg\min_{1\leqslant j\leqslant L} \pi_{Q,R_Q} M_j^\top. \qquad (6)$$

Since the optimum classification strategy gives rise to the smallest average Bayes error, we note that

$$J(R_1,\ldots,R_Q,D^{optimal}_{(R_1,\ldots,R_Q)}) \leqslant J(R_1,\ldots,R_Q,D_{(R_1,\ldots,R_Q)}). \qquad (7)$$

Based on this observation, Eq. (4) can be rewritten as follows:

$$\begin{aligned}\widetilde{J}(R_1,\ldots,R_Q) &\triangleq J(R_1,\ldots,R_Q,D^{optimal}_{(R_1,\ldots,R_Q)}) \\ &= \min_{D_{(R_1,\ldots,R_Q)}} J(R_1,\ldots,R_Q,D_{(R_1,\ldots,R_Q)}) \\ &= E\Bigg\{\sum_{q=1}^{Q}\sum_{n=1}^{R_q} c_{q,n} + \sum_{q=1}^{Q-1} s_{q,q+1} + g(\pi_{Q,R_Q})\Bigg\},\end{aligned} \qquad (8)$$

where the last term depends on π_{Q,R_Q}.

Next, the optimum value of $\widetilde{J}(R_1,\ldots,R_Q)$ with respect to $\{R_1,\ldots,R_Q\}$ can be found by solving the following optimization problem:

$$\min_{R_1\geqslant 0,\ldots,R_Q\geqslant 0} E\Bigg\{\sum_{q=1}^{Q}\sum_{n=1}^{R_q} c_{q,n} + \sum_{q=1}^{Q-1} s_{q,q+1} + g(\pi_{Q,R_Q})\Bigg\}. \qquad (9)$$

Specifically, given the structure of the problem, the optimum solution can be found using dynamic programming [17], as shown in Theorem 2.

Theorem 2 *For $n = \{N_q-1,\ldots,0\}$, stopping or continuing the feature review process is decided based on:*

$$\bar{J}(\pi_{q,n}) = \min\{\bar{J}_w(\pi_{q,n}), \bar{J}_c(\pi_{q,n})\}, \qquad (10)$$

where $\bar{J}_w(\pi_{q,n})$ and $\bar{J}_c(\pi_{q,n})$ denote the expected cost of stopping and continuing the feature review process, respectively, after reviewing the nth feature in the qth feature set. These functions are described by the following set of equations:

$$\bar{J}_c(\pi_{q,n}) = c_{q,n+1} + \sum_{y_{q,n+1}} \bar{J}(\pi_{q,n+1})\pi_{q,n}\Delta^\top_{q,n+1}(y_{q,n+1}), \qquad (11)$$

$$\begin{aligned}\bar{J}_w(\pi_{q,n}) &= s_{q,q+1} + c_{q+1,1} + \sum_{y_{q+1,1}} \bar{J}(\pi_{q+1,1})\pi_{q,n} \\ &\quad \times \Delta^\top_{q+1,1}(y_{q+1,1}),\end{aligned} \qquad (12)$$

$$\bar{J}_w(\pi_{Q,n}) = g(\pi_{Q,n}), \qquad (13)$$

where the last term denotes the expected cost of stopping the feature review process at the Qth feature set.

The optimal conditions derived from Eqs. (10)–(13) have a very intuitive structure. Specifically, when reviewing the nth feature of the qth feature set, our framework faces two options given $\pi_{q,n}$: (i) stop examining the current feature set, or (ii) continue with the next feature in the current feature set. For $q \in \{1,2,\ldots,Q-1\}$, when a decision to stop examining the current feature set is made, our framework switches to the $(q+1)$the feature set and considers the cost of using the first feature in that set. Once our framework decides to stop during the Qth feature set, a selection is made among the L hypotheses.

4. ALGORITHM

The theoretical results presented in Section 3 enable the design of an efficient algorithm for sequential heterogeneous feature selection for mulit–class classification of civil issue reports. Next, we discuss the main steps of the proposed algorithm and its computational complexity. Initially, the posterior probability $\pi_{1,0}$ is set to $[p_1, p_2, \ldots, p_L]$ and the two terms inside the minimization operator in Eq. (10) are compared. If the first term is greater than the second term, our algorithm reviews the first feature from the first feature set; otherwise, our algorithm moves to the next feature set. In any case, the posterior probability is updated using Eq. (3) to represent the new knowledge acquired from reviewing a particular feature. This process is repeated until all features sets are reviewed, in which case classification is performed using

$\sum_{q=1}^{Q} R_q$ features using the classification strategy in Eq. (6). We underscore that even though our algorithm moves from one feature set to the next, this does not necessarily imply that it will review any features from that feature set (i.e., $R_q = 0$ for some q). Fig. 2 provides a graphical representation of the proposed algorithm. An interesting property of our algorithm is that it can use different number of features to classify different civil issue reports (see also Section 5), improving not only classification accuracy but also model interpretability.

Our proposed algorithm consists of an off–line (training) and an online (testing) part. During training, key parameters such as the conditional probabilities $p(y_{q,n}|H_i)$ and the prior probabilities p_i are estimated from the available data (see Section 5). Additionally, a $b \times \sum_{q=1}^{Q} N_q$ matrix is generated using Theorem 2, where each row corresponds to $\sum_{q=1}^{Q} N_q$ values $\bar{J}(\pi_{q,n}), n = 0, \ldots, N_q - 1, q = 1, \ldots, Q$ for all possible b vectors of $\pi_{q,n}$. Among these two steps, the computational complexity of the later step prevails and is $\mathcal{O}(L|\mathcal{Y}|b)$, where $\mathcal{Y}$ denotes the set of feature values and $|\cdot|$ its cardinality. During testing, our algorithm can classify a civil issue report in $\mathcal{O}(L \times \sum_{q=1}^{Q} N_q)$, since updating the posterior probability using Eq. (3) is $\mathcal{O}(L)$ due to carrying out a dot product between a pair of L–dimensional vectors.

5. EXPERIMENTS

Our method is evaluated on a real–world dataset collected from the SeeClickFix[1] platform for the capital of the state of New York, between Jan 5, 2010 and Feb 10, 2018. Specifically, our dataset consists of 529 civil issue reports, each of which belongs to one of four hypotheses, namely "Parking Enforcement", "Code Violation", "Traffic Signal Repair" and "Signs (Missing, Needed, or Damaged)". Each civil issue report is described by two feature sets, where the first feature set contains 99 features extracted from the report's title and the second feature set contains 1507 features extracted from the report's description. Each feature is extracted by tokenizing sentences into unigrams, removing punctuation, stopwords, and digits, and stemming each word to its root (e.g., replace "parking" with "park"). A feature value corresponds to the number of appearances of a specific word in an issue, with words being present in $\geq 95\%$ and $\leq 2\%$ of all issues excluded.

Our experiments were performed for $L = 4$, $c_n = c \in \{0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-1}, 0.2, 0.3\}$, $M_{i,j} = 1, \forall i \neq j$ and $M_{i,i} = 0$. To avoid overfitting, reported results are based on five–fold cross–validation. We compare the performance of our framework to (i) standard Bayesian detection [18] that utilizes the top 1, 5, 10, 50, and all available features ordered using our proposed ordering method (see below), (ii) Support Vector Machine with feature selection (SVM–FS) [19] with linear (SVM–L) and Gaussian (SVM–G) kernels, and PCA (SVM–PCA) for dimensionality reduction, and (iii) Random Forest (RF) with maximum tree depths $d = 5, 10$, and XG Boosting (XG–B) [20, 21].

To estimate the conditional probabilities for each feature, we use a smoothed maximum likelihood estimator as follows:

$$\hat{p}(y_{q,n}|H_i) = \frac{N(y_{q,n}, i) + 1}{\sum_{y'_{q,n}} N(y'_{q,n}, i) + V}, \qquad (14)$$

where $N(y_{q,n}, i)$ denotes the number of civil issue reports belonging to hypothesis H_i that give rise to outcome $y_{q,n}$ after reviewing the nth feature in the qth feature set. V is the maximum outcome among all features. We estimate the *a priori* probability for each hypothesis i as follows:

$$P(H_i) = \frac{N_i}{\sum_{i=1}^{L} N_i}, \qquad (15)$$

where $i = 1, \ldots, L$. The values of function $\bar{J}(\pi_{q,n})$ are computed once offline using Theorem 2 for all possible combinations of $\pi^i_{q,n}, i = 1, \ldots, L$, by quantizing the interval $[0, 1]$ in increments of 0.1 such that $\sum_{i=1}^{L} \pi^i_{q,n} = 1$. This results in an efficient computation of $b \times \sum_{q=1}^{Q} N_q$ matrix, where b denotes the number of generated $\pi_{q,n}$ and N_q corresponds to the number of features in the qth feature set. In order to avoid the computational complexity of evaluating $N_q!$ possible feature orderings in each feature set q, we sort the features in increasing order of the sum of type I and type II errors, scaled by the cost coefficient c_n. This ordering promotes the most informative features that also have low cost.

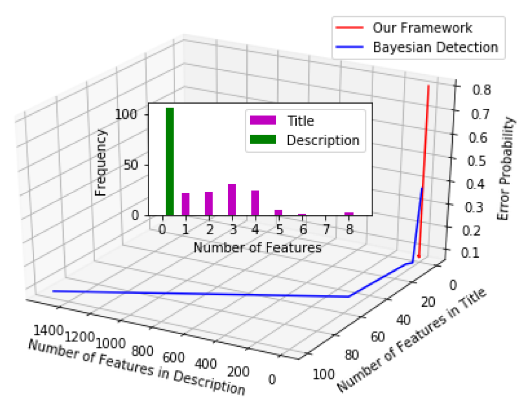

Fig. 3. Error probability as a function of the number of reviewed features from the civil issue report's description and title. The inset illustrates the distribution of the number of features used by our proposed framework when the error probability is the lowest.

Fig. 3 illustrates the error probability of our proposed algorithm as a function of the number of reviewed features from the title and description feature sets. The performance of the standard Bayesian detection is also given. We observe

[1] https://seeclickfix.com/

that our framework exhibits a large error probability when the number of features reviewed is small. However, the probability of error reduces dramatically when features from the title feature set are reviewed. Further, we observe that on average approximately 3 features are enough to classify the issue with same error confidence as Bayesian detection, where all the features from both features sets are reviewed. A careful analysis of the results indicates that the features mostly used by our algorithm correspond to the number of times the following words appear in the title {"park", "sign", "code", "violation", "traffic", "signal", "miss"}. Depending on the features reviewed by our algorithm for each civil issue report, we can thus explain the associated classification decision and improve the interpretability of the results. For instance, if the word "park" is included in a civil issue report, then with high probability the report belongs to the "Parking Enforcement" class. Fig. 4 illustrates our algorithm's operation on three different civil issue reports.

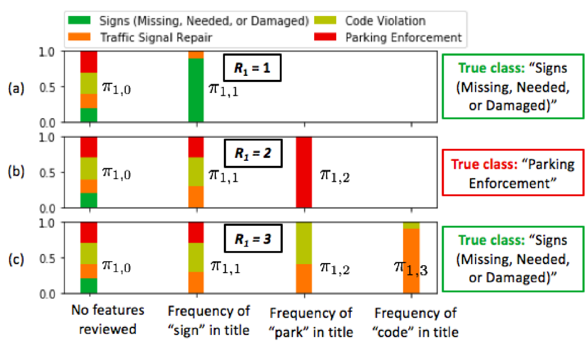

Fig. 4. Posterior probability evolution as more features are reviewed (left to right) for three civil issue reports (a)–(c). For instance, in (b), all classes are initially equiprobable , "Signs (Missing, Needed or Damaged)" becomes unlikely when the frequency of "sign" in the title is low, and "Parking Enforcement" becomes the most probable class with supporting evidence in the form of the frequency of term "park". Civil issue reports (a) and (b) are correctly classified, whereas (c) is incorrectly classified as "Traffic Signal Repair".

Table 1 illustrates the performance of our algorithm compared to baselines. In order to evaluate the classification performance, we use micro–averaged accuracy, macro–averaged precision and macro– averaged recall. From a computation point of view, micro–averaged accuracy uses the cumulative number of true positives, true negatives, false positives and false negatives per each class. Macro–averaged performance metrics are computed as the unweighted average of the associated performance index in each class. In all baselines, the "Average" indicates the unweighted average of the performance index when the title and description feature sets are used independently, whereas "Combined" indicates the performance measure when both title and description feature sets are considered into a single feature set. Among all baselines,

Table 1. Performance comparison with baselines.

	Parameters	**Accuracy**	**Precision**	**Recall**	**Avg. # feat.**	
					Set 1	Set 2
Our Algorithm	$c = 0.3$	0.3362	0.2208	0.3914	0.4584	0
	$c = 0.2$	0.9205	0.9179	0.9215	2.4080	0.1132
	$c = 0.1$	**0.9262**	**0.9270**	**0.9293**	**2.7910**	**0**
	$c = 10^{-3}$	0.9205	0.9230	0.9243	3.9314	0.0037
	$c = 0$	0.9205	0.9230	0.9243	11.0050	0.0037
Bayesian Detection	All (Average)	0.9205	0.9230	0.9243	99	1507
	All (Combined)	0.9205	0.9230	0.9243	1606	
	Top 50 (Average)	0.9205	0.9230	0.9243	50	50
	Top 50 (Combined)	0.9205	0.9230	0.9243	50	
	Top 10 (Average)	0.9205	0.9230	0.9243	10	10
	Top 10 (Combined)	0.9205	0.9230	0.9243	10	
	Top 5 (Average)	0.9318	0.9260	0.9313	5	5
	Top 5 (Combined)	0.9111	0.9177	0.9143	5	
	Top 1 (Average)	0.6181	0.5049	0.5782	1	1
	Top 1 (Combined)	0.4539	0.3230	0.4757	1	
SVM	SVM-L (Average)	0.8875	0.8730	0.8837	99	1507
	SVM-L (Combined)	**0.9697**	0.9612	0.9663	**1606**	
	SVM-G (Average)	0.8487	0.8631	0.8503	99	1507
	SVM-G (Combined)	0.9678	**0.9639**	0.9688	**1606**	
	SVM-FS (Average)	0.6637	0.7334	0.6987	24	10
	SVM-Fs (Combined)	0.9470	0.9496	0.9467	6	
	SVM-PCA (Average)	0.8478	0.8633	0.8493	11	190
	SVM-PCA (Combined)	0.96	0.95	0.96	1606	
RF	$d = 5$ (Average)	0.8497	0.8567	0.8530	99	1507
	$d = 5$ (Combined)	0.9583	0.9452	0.9617	1606	
	$d = 10$ (Average)	0.8771	0.8618	0.8710	99	1507
	$d = 10$ (Combined)	0.9696	0.9627	**0.9708**	**1606**	
XG-B	All (Average)	0.8789	0.8653	0.8726	99	1507
	All (Combined)	0.96	0.96	0.96	1606	

SVM–L (Combined) achieves the highest accuracy, while requiring $\sim$ 575 times more features than our framework. SVM–G (Combined) and RF $d = 10$ (Combined) use all features to achieve the highest precision and recall respectively among all methods. Combining the two feature sets results in better performance for all the baselines as opposed to using the two feature sets independently. In Bayesian Detection, classification performance increases until it converges to a fixed value. Against state–of–the–art, our proposed algorithm achieves 99.82% savings in feature use with less than 5% loss in performance, while improving model interpretability.

6. CONCLUSION

In this paper, the problem of multi–class classification of civil issue reports is addressed when multiple feature sets are present. Specifically, the objective is to accurately classify each civil issue report, while minimizing the cost incurred from reviewing features and switching between feature sets. The optimum solution of the associated optimization problem leads to a framework that sequentially reviews the least number of informative features and switches between feature sets when appropriate. The review process ends at the last feature set and is guaranteed to lead to the minimum misclassification cost. Experiments on a real–world dataset illustrates that the proposed framework results in an accurate classification of civil issue reports, while saving up to 99.82% features compared to the state–of–the–art.

Currently, we are in the process of assessing the performance of our proposed algorithm on a variety of datasets. As part of our future work, we plan to theoretically derive the average number and associated standard deviation of features per feature set needed to reach an accurate classification decision. Furthermore, we will consider various extensions of our existing framework, including but not limited to switching back/forth between feature sets, performing classification at any step of the process, and optimizing the order by which feature sets are reviewed.

7. REFERENCES

[1] Mergel, I., "Distributed Democracy : SeeClickFix.Com for Crowdsourced Issue Reporting," Tech. Rep., 2012.

[2] S. F. King and P. Brown, "Fix My Street or else: Using the Internet to Voice Local Public Service Concerns," in *1st International Conference on Theory and Practice of Electronic Governance*, 2007, pp. 72–80.

[3] A. Bousios, D. Gavalas, and L. Lambrinos, "CityCare: Crowdsourcing Daily Life Issue Reports in Smart Cities," in *IEEE Symposium on Computers and Communications (ISCC)*, July 2017, pp. 266–271.

[4] "https://publicadministration.un.org/egovkb/en-us/Data-Center," accessed 10-17-2019.

[5] Y. Sano, K. Yamaguchi, and T. Mine, "Category Estimation of Complaint Reports about City Park," in *4th International Congress on Advanced Applied Informatics*, July 2015, pp. 61–66.

[6] N. Beck, "Classification of Issues in the Public Space Using Their Textual Description and Geo-Location Masterarbeit," 2015.

[7] D.-S. Zois, C. Yong, C. Chelmis, A. Kapodistria, and W. Lee, "Improving Monitoring of Participatory Civil Issue Requests through Optimal Online Classification," in *52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 2034–2038.

[8] Y. Warahena Liyanage, D.-S. Zois, C. Chelmis, and M. Yao, "Automating the Classification of Urban Issue Reports: an Optimal Stopping Approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 3137–3141.

[9] Y. Warahena Liyanage, M. Yao, C. Yong, D.-S. Zois, and C. Chelmis, "What Matters the Most? Optimal Quick Classification of Urban Issue Reports by Importance," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov. 2018, pp. 106–110.

[10] C. Masdeval and A. Veloso, "Mining Citizen Emotions to Estimate the Urgency of Urban Issues," *Information Systems*, pp. 147–155, Dec. 2015.

[11] I. Nazar, Y. Warahena Liyanage, D.-S. Zois, and C. Chelmis, "Automated Optimal Online Civil Issue Classification using Multiple Feature Sets," in *53rd Asilomar Conference on Signals, Systems, and Computers*, 2019.

[12] Girish Chandrashekar and Ferat Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[13] John P Cunningham and Zoubin Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 2859–2900, 2015.

[14] Peng Zhou, Xuegang Hu, Peipei Li, and Xindong Wu, "Ofs-density: A novel online streaming feature selection method," *Pattern Recognition*, vol. 86, pp. 48–61, 2019.

[15] Kui Yu, Xindong Wu, Wei Ding, and Jian Pei, "Towards scalable and accurate online feature selection for big data," in *2014 IEEE International Conference on Data Mining*. IEEE, 2014, pp. 660–669.

[16] Xindong Wu, Kui Yu, Wei Ding, Hao Wang, and Xingquan Zhu, "Online feature selection with streaming features," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 5, pp. 1178–1192, 2012.

[17] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, Athena Scientific, 2005.

[18] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*, John Wiley & Sons, 2004.

[19] S. Hirokawa, T. Suzuki, and T. Mine, "Machine Learning is Better Than Human to Satisfy Decision by Majority," in *International Conference on Web Intelligence*, 2017, pp. 694–701.

[20] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[21] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.