Automated Optimal Online Civil Issue Classification using Multiple Feature Sets

Imara Nazar*, Yasitha Warahena Liyanage*, Daphney–Stavroula Zois*, Charalampos Chelmis†
*Department of Electrical and Computer Engineering, †Department of Computer Science
University at Albany, SUNY, Albany, NY, USA
Emails: {imohamednazar, yliyanage, dzois, cchelmis}@albany.edu

***Abstract*—In this paper, the automatic classification of non-emergency civil issues in crowdsourcing systems is addressed in the case where multiple feature sets are available. We recognize that multiple feature sets can contain useful complementary information regarding the type of an issue leading to a more accurate decision. However, using all features in these sets may delay the decision. Since we are interested in reaching an accurate decision in a timely manner, an optimal way of selecting features from multiple feature sets is needed. To this end, we propose a novel approach that sequentially reviews available features and feature sets to decide whether the feature review process must continue in the current set or move to the next one. In the end, when all feature sets have been reviewed, the issue is classified using all available information. It is shown that the proposed approach is guaranteed to review the least number of features in all feature sets before reaching a decision, while the optimum decision rule is shown to minimize the average Bayes risk. Evaluation on real world SeeClickFix data demonstrates the ability to classify issues by reviewing 99.5% less features than state–of–the–art without sacrificing accuracy.**

I. Introduction

Improving the life standards of a society requires collective effort from both citizens and government officials. Concerns from citizens must be conveyed to the corresponding government officials and they should be uniformly monitored until the issue is resolved. Currently, crowdsourcing platforms such as SeeClickFix [1] in US, FixMystreet in UK [2], and Novoville and IMcity [3] in Greece, provide an extra communication channel for concerned citizens to report their issues directly to government agencies.

Increased use of internet and mobile applications enable users to report an issue with less effort. In order to ensure timely support, each reported issue must be classified accurately and quickly before sending it to the respective officials. For an example, in a reported civil issue, the information embedded in title, description, image, or tags, can potentially be useful to make an accurate decision (see Fig. 1). However, manually processing all the available information in a reported issue by city officials can lead to delays in processing. Considering the large number of data present, a classification method is needed that requires less human intervention. Even though automatic classification of issues has been previously studied, it is limited to the use of a single feature set [4]–[9]. In [10], different classifiers were trained on different feature sets, and issues were independently classified using these sets (i.e., each issue was classified using a specific classifier trained on a specific fixed feature set).

Integrating information content from multiple independent feature sets is one of the key challenges in reaching an accurate decision quickly. This has been studied within the context of various text classification applications. In [11]–[14], decisions from multiple feature selection algorithms are integrated into a final classification decision. On the other hand, [15], [16] study the problem of designing different classifiers that use different types of feature sets. In all these cases, all features from the selected feature sets were used for classification. Further, scalability and timeliness of such methods have largely been ignored.

To address the aforementioned challenges, we formulate the problem of automatic classification using multiple feature sets as a sequential hypothesis testing problem, in which features in multiple feature sets are quickly examined to classify the issue, while ensuring the risk of misclassification will be low. In this paper, we propose to start the review process with the first available feature set with features in this set being sequentially reviewed in the order of decreasing information content. At each step, our framework decides whether to stop reviewing the features of the current feature set and move to the next feature set or continue. When our framework decides to stop reviewing the features from the last feature set, it attempts to classify the issue based on the features examined thus far. The proposed approach is guaranteed to review the least number of features from each feature set before reaching a decision, while optimizing the Bayes risk.

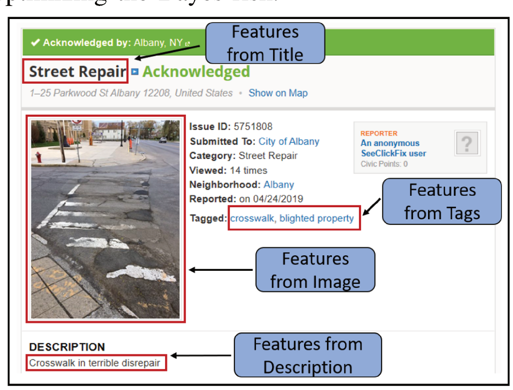

Fig. 1: A sample urban issue report in Albany, NY, from the SeeClickFix website.

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

II. Problem Formulation

We consider a set $\mathcal{I}$ of issues. Each issue $i \in \mathcal{I}$ is described by a vector $\mathbf{f} \triangleq [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_Q]^T$ of heterogeneous features, where $\mathbf{f}_q \triangleq [y_{q,1}, y_{q,2}, \dots, y_{q,N_q}]^T$ is a column vector consisting of N_q features that describe the qth feature set. We assume that we have access to a total of $N = \sum_{q=1}^{Q} N_q$ features. Each issue i belongs to one out of two hypotheses, $H_{\mathcal{C}_1}$ or $H_{\mathcal{C}_2}$, where $H_{\mathcal{C}_1}$ ($H_{\mathcal{C}_2}$) denotes that issue i is of type $\mathcal{C}_1$ ($\mathcal{C}_2$).

For each feature $y_{q,n}$, the probability $p(y_{q,n}|H_{\mathcal{C}_1})$ (similarly $p(y_{q,n}|H_{\mathcal{C}_2})$) of the evaluation of the nth feature that belongs to the qth feature set when the true hypothesis is $H_{\mathcal{C}_1}$ (similarly for true hypothesis $H_{\mathcal{C}_2}$) is empirically calculated from training data. Further, the *a priori* probability $P(H_{\mathcal{C}_1}) = p$ of issue i being an instance of $\mathcal{C}_1$ is also estimated empirically. The probability of issue i being an instance of $\mathcal{C}_2$ can be computed as $P(H_{\mathcal{C}_2}) = 1 - p$. We define the cost coefficient $c_{q,n} > 0$, $n \in \{1, 2, \dots, N_q\}, q \in \{1, 2, \dots, Q\}$ to denote the time and effort spent evaluating 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 reviewing features in the qth feature set to reviewing features in the q'th feature set. Further, we consider the misclassification cost $M_{mj} \geq 0$, which represents the cost of selecting type j when the true hypothesis is $H_m, m \in \{\mathcal{C}_1, \mathcal{C}_2\}$. Here, $j \in \{1, 2, \dots, L\}$, and L denotes the number of decision choices (*e.g.,* $\mathcal{C}_1, \mathcal{C}_2$).

In order to calculate the belief encoded in each issue i, we propose to review features and feature sets sequentially. More specifically, at each step, we have to select between continuing to review features in the current feature set or move to the next feature set. We also need to decide if we are ready to proceed with a classification decision or continue the feature review process. In any case, during this process, any decision we make is based on the accumulated information thus far, the cost of reviewing additional features as well as the cost of switching between feature sets. For simplicity, we begin our review process from the first available feature set and continue to the rest feature sets assuming an extant predefined order. In our future work, we plan to study the case where the order of the feature sets is not predefined and needs to be selected optimally. Our proposed sequential review process comprises a collection $\{R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)}\}$ of random variables. In particular, random variable $R_q \in \{1, 2, \dots, N_q\}, q \in \{1, 2, \dots, Q\}$, indicates the feature in the qth feature set that the framework decides to stop. Further, random variable $D_{(R_1,\dots,R_Q)}$ represents the decision we reach after concluding the review process. It depends on $R_q, q \in \{1, 2, \dots, Q\}$, and takes values in the set $\{1, 2, \dots, L\}$. For instance, when $L = 2$, $D_{(R_1,\dots,R_Q)} = 1$ denotes that issue i is of type $\mathcal{C}_1$, while $D_{(R_1,\dots,R_Q)} = 2$ denotes that issue i is of type $\mathcal{C}_2$. We note that while reviewing the features that belong to the qth feature set, the decision to stop the review process at any feature in this set depends only on the accumulated information until the random variable R_q. Equivalently, features (or feature sets) that may be examined in the future are not used.

Our goal is to select random variables $R_1, \dots, R_Q$ and decision rule $D_{(R_1,\dots,R_Q)}$ to accurately classify each issue i, while minimizing the cost incurred from reviewing individual features and switching between feature sets. To this end, we propose to solve the following optimization problem:

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

where

$$J(R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)}) \triangleq E\Bigg\{ \sum_{q=1}^{Q} \sum_{n=1}^{R_q} c_{q,n} + \sum_{q=1}^{Q-1} s_{q,q+1} + \sum_{j=1}^{L} \sum_{m=\{\mathcal{C}_1,\mathcal{C}_2\}} M_{mj} P(D_{(R_1,\dots,R_Q)} = j, H_m) \Bigg\}. \quad (2)$$

The first two terms in Eq. (2) denote the cost of reviewing the features that belong to different feature sets and the corresponding switching costs, whereas the last term penalizes the average cost of our classification rule.

In order to solve the optimization problem defined in Eq. (1), we first define the *a posteriori* probability $\pi_n^q \triangleq p(H_{\mathcal{C}_1}|y_{1,1}, \dots, y_{1,R_1}, \dots, y_{q,1}, \dots, y_{q,n})$, which corresponds to the accumulated information up to and including feature $y_{q,n}$. Lemma 1 shows how to compute π_n^q iteratively.

Lemma 1. *The posterior probability after reviewing the nth feature in the qth feature set ($y_{q,n}$) is:*

$$\pi_n^q = \frac{\pi_{n-1}^q P(y_{q,n}|H_{\mathcal{C}_1})}{\pi_{n-1}^q P(y_{q,n}|H_{\mathcal{C}_1}) + (1 - \pi_{n-1}^q) P(y_{q,n}|H_{\mathcal{C}_2})}, \quad (3)$$

where $\pi_{n-1}^q, n \in \{1, 2, \dots, N_q\}$, is the posterior probability after the review of the $n-1$ feature in the qth feature set. The initialization terms are given as $\pi_0^1 = p$ and $\pi_0^{q+1} = \pi_{R_q}^q$ for $q \in \{1, \dots, Q-1\}$.

Lemma 2. *Using Lemma 1 and the fact that $x_{(R_1,\dots,R_Q)} = \sum_{n_1=1}^{N_1} \dots \sum_{n_Q=1}^{N_Q} x_{(n_1,\dots,n_Q)} \mathbb{1}_{\{R_1=n_1,\dots,R_Q=n_Q\}}$, 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 written compactly as:*

$$J(R_1, \dots, R_Q, D_{(R_1,\dots,R_Q)}) = E\Bigg\{ \sum_{q=1}^{Q} \sum_{n=1}^{R_q} c_{q,n} + \sum_{q=1}^{Q-1} s_{q,q+1} \Bigg\} + E\Bigg\{ \sum_{j=1}^{L} \big(M_{\mathcal{C}_1 j} \pi_{R_Q}^Q + M_{\mathcal{C}_2 j} (1 - \pi_{R_Q}^Q) \big) \mathbb{1}_{\{D_{(R_1,\dots,R_Q)}=j\}} \Bigg\}. \quad (4)$$

III. Optimal Strategies

Here, we solve the optimization problem in Eq. (4) to derive the optimal stopping and classification strategies.

A. Classification Strategy

In our framework, the classification of issue i is performed only at the final stopping time R_Q after reviewing all available feature sets. In order to obtain the optimal classification rule $D_{(R_1,\dots,R_Q)}$, an independent of stopping times $R_1, R_2, \dots, R_Q$, lower bound for the last term of Eq. (4) is needed. Since $D_{(R_1,\dots,R_Q)}$ contributes only to this portion of the average cost, the optimal classification rule $D_{(R_1,\ldots,R_Q)}$ for given stopping times $R_1,\ldots,R_Q$ can then be derived. Theorem 3 provides such bound.

Theorem 3. *For any classification rule $D_{(R_1,\ldots,R_Q)}$ given stopping times $R_1,R_2,\ldots,R_Q$, we can show that:*

$$\sum_{j=1}^{L}\left(M_{\mathcal{C}_1 j}\pi_{R_Q}^{Q}+M_{\mathcal{C}_2 j}(1-\pi_{R_Q}^{Q})\right)\mathbb{1}_{\{D_{(R_1,\ldots,R_Q)}=j\}} \geqslant g(\pi_{R_Q}^{Q}), \tag{5}$$

where $g(\pi_{R_Q}^{Q}) \triangleq \min_{1\leqslant j\leqslant L}\left[M_{\mathcal{C}_1 j}\pi_{R_Q}^{Q}+M_{\mathcal{C}_2 j}(1-\pi_{R_Q}^{Q})\right]$. The optimal rule is defined as follows:

$$D_{(R_1,\ldots,R_Q)}^{optimal} = arg\min_{1\leqslant j\leqslant L}\left[M_{\mathcal{C}_1 j}\pi_{R_Q}^{Q}+M_{\mathcal{C}_2 j}(1-\pi_{R_Q}^{Q})\right]. \tag{6}$$

From Theorem 3, we deduce that $J(R_1,\ldots,R_Q,D_{(R_1,\ldots,R_Q)}^{optimal}) \leqslant J(R_1,\ldots,R_Q,D_{(R_1,\ldots,R_Q)})$, since the optimal classification rule results to the smallest average cost. Based on this observation, Eq. (4) can be rewritten as follows:

$$\tilde{J}(R_1,\ldots,R_Q) \triangleq J(R_1,\ldots,R_Q,D_{(R_1,\ldots,R_Q)}^{optimal}) \tag{7}$$

$$= \min_{D_{(R_1,\ldots,R_Q)}} J(R_1,\ldots,R_Q,D_{(R_1,\ldots,R_Q)}) \tag{8}$$

$$= \mathbb{E}\left[\sum_{q=1}^{Q}\sum_{n=1}^{R_q} c_{q,n}+\sum_{q=1}^{Q-1} s_{q,q+1}+g(\pi_{R_Q}^{Q})\right], \tag{9}$$

where the last term depends only on R_Q.

B. Stopping Strategy

To determine the optimum value of $\tilde{J}(R_1,\ldots,R_Q)$ with respect to $R_q, q\in\{1,\ldots,Q\}$, the following optimization problem must be solved:

$$\min_{R_1\geqslant 0,\ldots,R_Q\geqslant 0}\mathbb{E}\left[\sum_{q=1}^{Q}\sum_{n=1}^{R_q} c_{q,n}+\sum_{q=1}^{Q-1} s_{q,q+1}+g(\pi_{R_Q}^{Q})\right]. \tag{10}$$

The solution to the optimization problem given in Eq. (10) can be found via dynamic programming, as shown in Theorem 4.

Theorem 4. *For $n=\{N_q-1,\ldots,0\}$, stopping or continuing the feature review process is decided based on:*

$$\bar{J}(\pi_n^q) = \min\left\{\bar{J}_s(\pi_n^q),\bar{J}_c(\pi_n^q)\right\}, \tag{11}$$

where $J_s(\pi_n^q)$ and $J_c(\pi_n^q)$ denote the expected cost of stopping and continuing the feature review process, respectively, after reviewing the nth feature in the qth feature set. Furthermore, these functions are described by the following set of equations:

$$\bar{J}_c(\pi_n^q) = c_{q,n+1}+\sum_{y_{q,n+1}}\bar{J}(\pi_{n+1}^q)\big(\pi_n^q P(y_{q,n+1}|H_{\mathcal{C}_1}) + (1-\pi_n^q)P(y_{q,n+1}|H_{\mathcal{C}_2})\big), \tag{12}$$

$$\bar{J}_s(\pi_n^q) = s_{q,q+1}+c_{q+1,1}+\sum_{y_{q+1,1}}\bar{J}(\pi_1^{q+1}) \times\left(\pi_n^q P(y_{q+1,1}|H_{\mathcal{C}_1})+(1-\pi_n^q)P(y_{q+1,1}|H_{\mathcal{C}_2})\right), \tag{13}$$

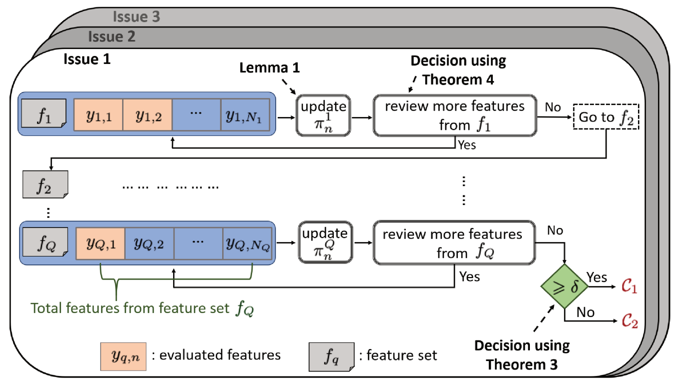

Fig. 2: Graphical representation of proposed approach.

where $q\in\{1,\ldots,Q\}$ and $\bar{J}_s(\pi_n^Q) = g(\pi_n^Q)$ denotes the expected cost of stopping the feature review process at the Qth feature set.

The optimal conditions described by Eq. (11) have a very intuitive structure, i.e., stop at the stage where the cost of stopping is smaller than the cost of continuing. After reviewing the nth feature in the qth feature set, our method faces two options given π_n^q: (i) stop reviewing features that belong to the current feature set, or (ii) continue and review the next feature in this set. For $q\in\{1,2,\ldots,Q-1\}$, a decision to stop reviewing features in the qth feature set implies that we switch to the $q+1$th feature set, and continue to review features from this feature set starting from the first feature. On the other hand, once the feature review process terminates in the Qth feature set, we proceed with optimally selecting between the L possibilities. Since we wish to find the optimal solution, we select the option that gives rise to the minimum cost.

Fig. 2 shows a graphical representation of our proposed approach. Initially, the posterior probability for the 1st feature set (π_0^1) is set to the prior probability p and features from the 1st feature set are sequentially reviewed. After the review of each feature in the 1st feature set, the posterior probability is updated according to Lemma 1. Next, our proposed approach evaluates the expressions in Theorem 4 to decide if the feature review process in the 1st feature set must be continued or not. If the answer is negative, then our approach moves on to the 2nd feature set, initializes the posterior probability (π_0^2) for this set to the final posterior probability of the 1st feature set ($\pi_{R_1}^1$), and sequentially reviews features optimally. This process continues until our approach switches to the last (Qth) feature set, at which stage when the feature review process terminates, our approach proceeds with classifying the issue using the optimal classification strategy given in Eq. (6).

C. Practical Considerations and Implementation

From an implementation point of view, the values of functions $\bar{J}(\pi_n^q), q\in\{1,\ldots,Q\}, n=\{N_q-1,\ldots,0\}$ are computed once offline using Theorem 4 by uniformly quantizing π_n^q in the interval $[0,1]$. Further, the conditional probabilities for each feature in each of the feature sets are estimated from training data as follows:

$$\hat{p}(y_{q,n}|H_{\mathcal{C}_1}) = \frac{N(y_{q,n}, \mathcal{C}_1)}{\sum_{y'_{q,n}} N(y'_{q,n}, \mathcal{C}_1)}, \quad (14)$$

$$\hat{p}(y_{q,n}|H_{\mathcal{C}_2}) = \frac{N(y_{q,n}, \mathcal{C}_2)}{\sum_{y'_{q,n}} N(y'_{q,n}, \mathcal{C}_2)}, \quad (15)$$

where $N(y_{q,n}, \mathcal{C}_1)$ and $N(y_{q,n}, \mathcal{C}_2)$ denote the number of issues of type $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively, that give rise to outcome $y_{q,n}$ after extracting, and reviewing the nth feature in the qth feature set. We also estimate the *a priori* probabilities as follows:

$$[P(H_{\mathcal{C}_1}), P(H_{\mathcal{C}_2})]^T = [p, 1-p]^T = \left[\frac{N_{\mathcal{C}_1}}{N_{\mathcal{C}_1} + N_{\mathcal{C}_2}}, \frac{N_{\mathcal{C}_2}}{N_{\mathcal{C}_1} + N_{\mathcal{C}_2}}\right]^T, \quad (16)$$

where $N_{\mathcal{C}_1}$ and $N_{\mathcal{C}_2}$ denote the number of type $\mathcal{C}_1$ and type $\mathcal{C}_2$ issues in the training set, respectively.

IV. Numerical Results

We illustrate the performance of our algorithm on a real–world dataset of 263 civil issues, spanning a time period between Jan 5, 2010 and Feb 10, 2018, for the capital of the state of New York, collected from SeeClickFix[1]. We performed experiments to classify between one out of two hypotheses, i.e., (i) Signs (missing, needed, or damaged) ($\mathcal{C}_1$), and (ii) Parking Enforcement ($\mathcal{C}_2$). Each issue is described by two feature sets, where the first feature set contains $1,017$ features extracted from the description and the second feature set contains 61 features extracted from the title. Each feature in these two feature sets is extracted by tokenizing sentences into unigrams, removing punctuation (e.g., periods, commas, and apostrophes), stopwords (e.g., "a", "the", "there"), and digits (e.g., "8th", "31st"), and stemming each word to its root (e.g., replace "parked" with "park"). A feature value corresponds to the number of appearances of a specific word in the issue report, with words being present in $\geq 95\%$ and $\leq 2\%$ of all issues excluded.

For each feature set, conditional probabilities were estimated from the training dataset. Features belonging to each feature set were sorted in increasing order of the sum of type I and type II errors to ensure the most informative features from each feature set are reviewed first. The prior probability of an issue belonging to a certain hypothesis is also estimated from the training dataset. The number L of decision choices is set to two and experiments were performed for varying feature costs $c_n \in \{0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-1}, 0.41\}$. We also set misclassification costs to constant values, i.e., $M_{\mathcal{C}_1 1} = M_{\mathcal{C}_2 2} = 0$, $M_{\mathcal{C}_1 2} = M_{\mathcal{C}_2 1} = 1$. To avoid overfitting, reported results are based on five–fold cross validation.

Our proposed approach starts by sequentially reviewing features from the first feature set. During the feature review process, it recursively updates the posterior probability π_n^1 by incorporating the new knowledge incurred from reviewing feature $y_{1,n}$, and decides if more features should be reviewed from the first feature set. Once it decides to stop reviewing features from the first feature set, it switches to the second feature set and continues the feature review process. Similarly to the case of the first feature set, it recursively updates the posterior probability π_n^2. At the same time, it decides whether to continue reviewing more features from the second feature set or not. Once it decides to stop reviewing features from the second feature set, each issue is classified either as type $\mathcal{C}_1$ or type $\mathcal{C}_2$ using the optimal classification strategy.

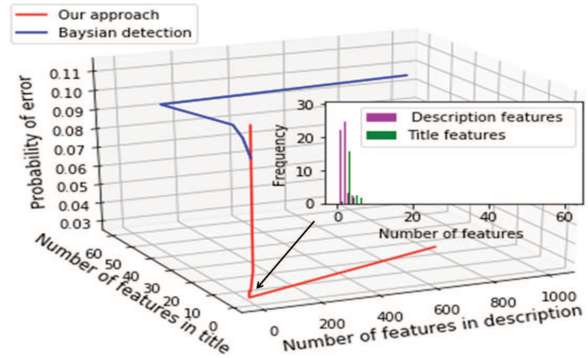

Fig. 3: Probability of error as a function of the number of reviewed features from description feature set and title set. Inset shows the distribution of number of features used by our proposed method to classify issues when probability of error is lowest.

We compare the performance of our approach to (i) a standard Bayesian detection method [17] that uses the top 1, 5, 10, 50 features ordered using the proposed ordering technique, as well as all available features, (ii) prior work, i.e., Support Vector Machine with feature selection (SVM–FS) [18] 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), which have been shown to achieve good performance while being relatively fast compared to other classification models [19], [20].

Fig. 3 shows the error probability achieved by our proposed approach and the standard likelihood ratio test (also shown in Fig. 3 as Bayesian detection) as a function of the number of reviewed features from the first and second feature sets. Our approach exhibits a large error probability when the number of features reviewed by each feature set are small and this goes down significantly as more features are reviewed. The inset in Fig. 3 shows the number of features reviewed from each feature set when the probability of error of our approach reaches the lowest value. We observe that on average approximately 2 features from the first feature set and 3 features from second feature set must be reviewed before classifying an issue.

Table I summarizes the performance of our approach with respect to the baselines. In all baselines, "Average" indicates the average value of the performance metric (i.e., Accuracy, Precision, Recall) when the two feature sets are used independently to compute the corresponding metric. On the other hand, "Combined" indicates the value of the performance

[1] https://seeclickfix.com/

TABLE I: Performance comparison of our approach with baselines.

	Parameters	Accuracy	Precision	Recall	Avg. # feat. Set 1	Avg. # feat. Set 2
Our Approach	$c = 0.41$	0.884	0.899	0.949	0	0.8
	$c = 0.10$	0.958	0.977	0.936	0	1
	$c = 10^{-3}$	0.969	1	0.938	0	2.825
	$c = 10^{-4}$	**0.973**	**1**	**0.948**	**1.73**	**3.175**
	$c = 10^{-5}$	0.973	1	0.948	11.01	3.325
	$c = 0$	0.957	1	0.916	645.63	3.128
Bayesian Detection	All (Average)	0.904	0.936	0.861	1017	61
	All (Combined)	0.969	1	0.938	1078	
	Top 50 (Average)	0.901	0.915	0.873	50	50
	Top 50 (Combined)	0.973	1	0.947	50	
	Top 10 (Average)	0.889	0.893	0.872	10	10
	Top 10(Combined)	**0.973**	**1**	0.948	10	
	Top 5 (Average)	0.893	0.899	0.880	5	5
	Top 5 (Combined)	0.969	0.992	0.948	5	
	Top 1 (Average)	0.901	0.919	0.864	1	1
	Top 1 (Combined)	0.958	0.977	0.936	1	
SVM	SVM–L (Average)	0.933	0.947	0.907	1017	61
	SVM–L (Combined)	0.973	0.993	0.957	1078	
	SVM–G (Average)	0.813	0.900	0.623	1017	61
	SVM–G (Combined)	0.943	**1**	0.888	1078	
	SVM–FS (Average) [18]	0.899	0.899	0.898	6	6
	SVM–FS (Combined) [18]	0.958	0.961	0.957	6	
	SVM–PCA (Average)	0.916	0.941	0.875	118	
	SVM–PCA (Combined)	0.966	0.992	0.941	109	
RF	d=5 (Average)	0.884	0.906	0.839	1017	61
	d=5 (Combined)	0.946	0.993	0.906	1078	
	d=10 (Average)	0.890	0.902	0.870	1017	61
	d=10 (Combined)	0.958	0.993	0.925	1078	
XG–B	All (Average)	0.903	0.908	0.881	1017	61
	All (Combined)	**0.977**	0.969	**0.987**	1078	

metric when the two feature sets are fused to one, i.e., both title and description feature sets are considered as one single feature set. Among all baselines, XG–B (Combined) using all features achieves the highest accuracy and recall, but requires ~ 215 times as many features in total as our approach for a mere 0.41% and 4.11% improvement, respectively. We notice that using the "Combined" feature set always gives rise to better results than using two feature sets independently. Our approach reaches a similar performance as Baysian detection with Top 10 features (Combined) while reviewing only ~ 1.73 and ~ 3.17 features from the first and second feature set, respectively. SVM–G (Combined) achieves perfect precision as our approach, reviewing however ~ 215 times more features.

V. Conclusions and Future Work

In this work, a sequential hypothesis testing framework was proposed to address the problem of automatic civil issues classification in an accurate and timely manner. Specifically, an optimization problem was defined in terms of the cost of reviewing features, cost of switching between feature sets and the Bayes risk associated with the decision rule. The optimal classification was shown to minimize the average Bayes risk, while the optimal stopping strategy is guaranteed to review the least number of features in all feature sets before reaching a decision. Evaluation on a real–world dataset from SeeClickFix showed that accurate classification can be performed while reducing the number of features used by up to 99.5% compared to state–of–the–art. In our future work, we plan to consider the case where the order of the feature sets is not explicitly known.

References

[1] I. Mergel, "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 *Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV '07. New York, NY, USA: ACM, 2007, pp. 72–80. [Online]. Available: http://doi.acm.org/10.1145/1328057.1328076

[3] A. Bousios, D. Gavalas, and L. Lambrinos, "Citycare: Crowdsourcing daily life issue reports in smart cities," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, July 2017, pp. 266–271.

[4] Y. Sano, K. Yamaguchi, and T. Mine, "Category estimation of complaint reports about city park," in *2015 IIAI 4th International Congress on Advanced Applied Informatics*, July 2015, pp. 61–66.

[5] D.-S. Zois, C. Yong, C. Chelmis, A. Kapodistria, and W. Lee, "Improving monitoring of participatory civil issue requests through optimal online classification," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 2034–2038.

[6] Y. W. Liyanage, D. Zois, C. Chelmis, and M. Yao, "Automating the classification of urban issue reports: an optimal stopping approach," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 3137–3141.

[7] P. S. Kochhar, F. Thung, and D. Lo, "Automatic fine-grained issue report reclassification," in *2014 19th International Conference on Engineering of Complex Computer Systems*, Aug 2014, pp. 126–135.

[8] C. Masdeval and A. Veloso, "Mining citizen emotions to estimate the urgency of urban issues," *Inf. Syst.*, vol. 54, no. C, pp. 147–155, Dec. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.is.2015.06.008

[9] Y. Liyanage, M. Yao, C. Yong, D. Zois, and C. Chelmis, "What matters the most? optimal quick classification of urban issue reports by importance," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2018, pp. 106–110.

[10] N. Beck, "Classification of issues in the public space using their textual description and geo-location," 2015.

[11] D. Koller and M. Sahami, "Hierarchically classifying documents using very few words," Stanford InfoLab, Technical Report 1997-75, February 1997, previous number = SIDL-WP-1997-0059. [Online]. Available: http://ilpubs.stanford.edu:8090/291/

[12] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," *Inf. Sci.*, vol. 181, pp. 1138–1152, 2011.

[13] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, April 2005.

[14] K. Chen, L. Wang, and H. Chi, "Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 03, pp. 417–445, 1997.

[15] S. Stüker, F. Metze, T. Schultz, and A. H. Waibel, "Integrating multilingual articulatory features into speech recognition," in *INTERSPEECH*, 2003.

[16] R. Rose and P. Momayyez, "Integration of multiple feature sets for reducing ambiguity in asr," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 4, April 2007, pp. IV–325–IV–328.

[17] H. L. Van Trees, "Detection, estimation, and modulation theory, part i: Detection, estimation, and linear modulation theory." 1968.

[18] S. Hirokawa, T. Suzuki, and T. Mine, "Machine learning is better than human to satisfy decision by majority," in *Proceedings of the International Conference on Web Intelligence*, ser. WI '17. New York, NY, USA: ACM, 2017, pp. 694–701. [Online]. Available: http://doi.acm.org/10.1145/3106426.3106520

[19] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016. [Online]. Available: http://arxiv.org/abs/1603.02754