WHAT MATTERS THE MOST? OPTIMAL QUICK CLASSIFICATION OF URBAN ISSUE REPORTS BY IMPORTANCE

Yasitha Liyanage[⋆], *Mengfan Yao*[†], *Christopher Yong*[†], *Daphney–Stavroula Zois*[⋆], *Charalampos Chelmis*[†]

[⋆]Electrical and Computer Engineering Department, [†]Department of Computer Science
University at Albany, SUNY, Albany, NY, USA

ABSTRACT

Civic engagement platforms such as SeeClickFix and FixMyStreet have revolutionized the way citizens interact with local governments to report and resolve urban issues. However, recognizing which urban issues are important to the community in an accurate and timely manner is essential for authorities to prioritize important issues, allocate resources and maintain citizens' satisfaction with local governments. To this end, a novel formulation based on optimal stopping theory is devised to infer urban issues importance from ambiguous textual, time and location information. The goal is to optimize recognition accuracy, while minimizing the time to reach a decision. The optimal classification and stopping rules are derived. Furthermore, a near–real–time urban issue reports processing method to infer the importance of incoming issues is proposed. The effectiveness of the proposed method is illustrated on a real–word dataset from SeeClickFix, where significant reduction in time–to–decision without sacrificing accuracy is observed.

***Index Terms*—** participatory civil issues, issue urgency, government 2.0, optimal stopping theory, quickest detection

1. INTRODUCTION

In recent years, "Government 2.0" applications [1,2] and civic engagement platforms have not only enabled citizens to actively participate in collecting, analyzing and sharing knowledge about their local environments (e.g., measure air quality [3], map fuel consumption on city streets [4], predict bus arrival times [5]), but also interact with local governments to resolve urban issues, such as potholes and noise complaints (e.g., SeeClickFix [6] and FixMyStreet [7]). At the same time, local governments can gain a better understanding of the urban issues faced by their communities, as long as reported issues are timely processed and addressed to maintain citizens' participation in urban issue monitoring [6, 8, 9].

Currently, reported issues are acknowledged and assessed by a city official for routing to the appropriate agency. Needless to say, this approach does not scale. Methods for the automatic classification [10–12] and identification of duplicate urban issues [13] have been recently proposed. Such approaches ignore citizens implicit endorsement of urban issues that are "important" to them. However, recognizing the importance of an issue early on can assist local governments in prioritizing important issues to better serve their citizens. To this end, the problem of automatically detecting signs of danger [14] and/or inferring the "urgency" of urban issues [15] reported by concerned citizens has recently been explored. Such supervised methods require large–scale annotation to achieve good accuracy. Moreover, the scalability and timeliness of such methods have largely been ignored. Finally, simple text– and emotion–based features are often considered, ignoring important spatial and temporal factors that have the potential to facilitate important urban issues recognition.

To address the challenges associated with identifying "important" issues early on, we formulate this problem as an optimal stopping problem, in which features extracted from an issue report are sequentially evaluated to infer its importance as fast as possible without sacrificing accuracy. We show that the optimal solution has a very intuitive structure: (i) features are sequentially evaluated starting from the most informative, (ii) at each step, the framework decides whether to stop the process, and (iii) once stopped, a given issue is assigned an importance value based on features examined thus far. Based on this framework, we devise a method to infer the importance of reported issues from ambiguous information such as textual description, reported time, and location in near–real–time. The optimal number of features used by our method depends on the cost representing the time and effort evaluating each feature and the classification quality. Thus, our approach provides a viable, realistic and timely solution to the recognition and prioritization of important urban issues by efficiently utilizing computational resources rather than blindly relying on the same fixed set of features for all issues, as done by state–of–the–art classifiers.

2. PROBLEM FORMULATION

2.1. Description

We consider a set $\mathcal{I}$ of issues, where each issue $i \in \mathcal{I}$, that has been reported by a concerned citizen, consists of a title, de-

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

scription, address, timestamp, photo(s), and comment(s) from other citizens. A feature vector $\mathbf{f}_i = [f_1, f_2, \ldots, f_K]^T$, where K denotes the total number of features extracted from a given report, is used to represent each issue i, which may belong to one of two hypotheses $H_{\mathcal{H}}$ or $H_{\mathcal{L}}$, the true hypothesis that i is of high or low importance, respectively.

Although the definition of importance is subjective, issues that are likely to have a profound effect on many people are intuitively expected to receive high attention (e.g., quantified by the number of views and/or comments), be reported from different sources, and/or persist for a large period of time. In our modeling, we use the number of votes and comments received to quantify urban issue importance (see Section 5).

We pose the challenge of automatically inferring the importance of an issue as a quickest detection problem. Specifically, the proposed framework sequentially evaluates features $f_n, n \in \{1, 2, \ldots, K\}$, one at a time, deciding at each step between stopping and continuing the evaluation process based on accumulated information thus far and the cost of reviewing additional features. Cost coefficient $c_n > 0$, $n \in \{1, \ldots, K\}$, represents the value of time and effort spent evaluating the nth feature. We also consider misclassification costs $M_{kj} \geqslant 0, k \in \{\mathcal{H}, \mathcal{L}\}, j \in \{1, \ldots, L\}$, where M_{kj} represents the cost of assigning importance j to issue i when the true hypothesis is H_k, and L represents the number of available decision choices. For example, when $L = 3$, $j = 1$ may correspond to "high importance issue", $j = 2$ may denote "low importance issue", and $j = 3$ may indicate "human expert inspection required". We factor misclassification costs into our approach to quantify the relative importance of detection errors.

Our proposed sequential evaluation process comprises a pair (R, D_R) of random variables. Random variable R (referred to as *stopping time* in decision theory) takes values in the set $\{0, \ldots, K\}$, and indicates the feature that the framework stops at. Random variable D_R takes values in the set $\{1, \ldots, L\}$ and denotes the possibility to select among L choices. For each feature f_n, the probability $p(f_n|H_{\mathcal{H}})$ (similarly $p(f_n|H_{\mathcal{L}})$) of the evaluation of the nth feature to observe value f_n when the true hypothesis is $H_{\mathcal{H}}$ (similarly for true hypothesis $H_{\mathcal{L}}$) is empirically computed from training data. The *a priori* probability $P(H_{\mathcal{H}}) = p$ of i being a high importance issue is also estimated empirically. The probability of i being a low importance issue can be computed as $P(H_{\mathcal{L}}) = 1 - p$. Assuming for simplicity that features f_n are *independent under each hypothesis* $H_k, k = \{\mathcal{H}, \mathcal{L}\}$, the conditional joint probability of $\{f_1, \ldots, f_n\}$ is given as $P(f_1, \ldots, f_n|H_k) = \prod_{l=1}^{n} p(f_l|H_k)$. Even though validation of this assumption is beyond the scope of this paper, we find our proposed method to work well in practice. Both the decision to stop at stage n (i.e., the event $\{R = n\}$), and the selection of possibility j (i.e., $D_R = j$) depend only on the accumulated information $\{f_1, \ldots, f_R\}$.

2.2. Optimization Problem

To minimize the number of features considered for inferring the importance of an issue without sacrificing accuracy, the stopping time R and the classification rule D_R have to be selected. To this end, we define the following cost function:

$$J(R, D_R) = \mathbb{E}\left\{ \sum_{n=1}^{R} c_n + \sum_{j=1}^{L} \sum_{k=\mathcal{H},\mathcal{L}} M_{kj} P(D_R = j, H_k) \right\}. \quad (1)$$

The former expression regularizes the number of features, whereas the latter penalizes the average cost of our classification rule. Our goal can be interpreted as finding the minimum average cost with respect to both random variables R and D_R, i.e., $\min_{R, D_R} J(R, D_R)$, to derive the optimal stopping and classification rules. To prove that the optimal rule is to stop at R, we will first show how to obtain the optimum classification rule D_R for any given stopping time R. Once the optimal classification rule has been established, the resulting cost becomes only a function of R, and can thus be optimized with respect to R. Since D_R depends only on the accumulated information $\{f_1, \ldots, f_R\}$, the *a posteriori probability* $\pi_n \triangleq P(H_{\mathcal{H}}|f_1, \ldots, f_n)$ is a sufficient statistic of the accumulated information, and must be updated as more features are evaluated as shown in Lemma 1.

Lemma 1. *The posterior probability π_n when the nth feature is evaluated to generate outcome f_n, and $\pi_0 = p$, is:*

$$\pi_n = \frac{p(f_n|H_{\mathcal{H}})\pi_{n-1}}{\pi_{n-1} p(f_n|H_{\mathcal{H}}) + (1 - \pi_{n-1}) p(f_n|H_{\mathcal{L}})}. \quad (2)$$

Lemma 1 and the fact that $x_R = \sum_{n=0}^{K} x_n \mathbb{1}_{\{R=n\}}$ for any sequence of random variables $\{x_n\}$, where $\mathbb{1}_A$ is the indicator function for event A, allow us to rewrite the average cost in Eq. (1) as:

$$J(R, D_R) = \mathbb{E}\left\{ \sum_{n=1}^{R} c_n + \sum_{j=1}^{L} \left(M_{\mathcal{H}j} \pi_R + M_{\mathcal{L}j} (1 - \pi_R) \right) \mathbb{1}_{\{D_R = j\}} \right\}. \quad (3)$$

3. OPTIMAL SOLUTION

An independent of D_R lower bound for Eq. (3) can be derived by observing that D_R contributes only to a portion of the average cost. Theorem 2 provides such bound, which also gives rise to the optimal classification rule.

Theorem 2. *For any classification rule D_R given stopping time R, $\sum_{j=1}^{L} \left(M_{\mathcal{H}j} \pi_R + M_{\mathcal{L}j} (1 - \pi_R) \right) \mathbb{1}_{\{D_R = j\}} \geqslant g(\pi_R)$, where $g(\pi_R) \triangleq \min_{1 \leqslant j \leqslant L} \left[M_{\mathcal{H}j} \pi_R + M_{\mathcal{L}j} (1 - \pi_R) \right]$. The optimal rule is defined as follows:*

$$\mathcal{D}_R^{optimal} = \arg\min_{1 \leqslant j \leqslant L} \left[M_{\mathcal{H}j} \pi_R + M_{\mathcal{L}j} (1 - \pi_R) \right]. \quad (4)$$

From Theorem 2, $J(R, D_R^{optimal}) \leqslant J(R, D_R)$, since the optimal classification rule results to the smallest average cost. Based on the this fact, Eq. (3) can be written as follows:

$$\widetilde{J}(R) \triangleq J(R, D_R^{optimal}) = \mathbb{E}\left\{ \sum_{n=1}^{R} c_n + g(\pi_R) \right\}, \quad (5)$$

which depends only on the stopping time R. To obtain the optimal stopping rule R, we need to solve the following optimization problem:

$$\min_{R\geqslant 0}\widetilde{J}(R) = \min_{R\geqslant 0}\mathbb{E}\left\{\sum_{n=1}^{R}c_n + g(\pi_R)\right\}, \tag{6}$$

which constitutes a classical problem in optimal stopping theory for Markov processes [16]. We derive our optimal stopping rule as described in Theorem 3 based on the observation that (i) the optimum rule will consist of a maximum of $K+1$ stages since $R\in\{0,1,\dots,K\}$, and (ii) the solution we seek must also be optimum, if instead of the first stage we start from any intermediate stage and continue toward the final stage [17].

Theorem 3. *For $n = K-1,\dots,0$, the function $\bar{J}_n(\pi_n)$ is related to $\bar{J}_{n+1}(\pi_{n+1})$ through the equation:*

$$\bar{J}_n(\pi_n) = \min\Bigg[g(\pi_n), c_{n+1} + \sum_{f_{n+1}} A_n(f_{n+1})\times \bar{J}_{n+1}\left(\frac{p(f_{n+1}|H_{\mathcal{H}})\pi_n}{A_n(f_{n+1})}\right)\Bigg], \tag{7}$$

where $A_n(f_{n+1}) \triangleq \pi_n p(f_{n+1}|H_{\mathcal{H}}) + (1-\pi_n)p(f_{n+1}|H_{\mathcal{L}})$ and $\bar{J}_K(\pi_K) = g(\pi_K)$.

The optimal stopping rule stemming from Eq. (7) has a very intuitive structure. Specifically, at each stage n, there are two options given π_n: (i) stop examining features and select optimally between the L possibilities, or (ii) continue and evaluate the next feature. The cost of stopping is $g(\pi_n)$, whereas the cost of continuing is $c_{n+1} + \sum_{f_{n+1}} A_n(f_{n+1})\bar{J}_{n+1}\left(\frac{p(f_{n+1}|H_{\mathcal{H}})\pi_n}{A_n(f_{n+1})}\right)$.

4. CIVIC ALGORITHM

In this section, we describe CIvIC, a novel algorithm to automatically Classify urban Issues into Importance Categories based on Lemma 1 and Theorems 2 and 3. Initially, the posterior probability π_0 of an issue is set to the prior probability p of a reported issue being important, and the two terms inside the minimization of Eq. (7) are compared. If they are equal, CIvIC assigns the appropriate importance value to the issue under examination based on the optimal rule of Eq. (4). Otherwise, the first feature is evaluated and the posterior probability π_1 is updated according to Eq. (2). CIvIC repeats these steps until either it decides to stop, at which case it uses $< K$ features, or all features are evaluated to assign the appropriate importance to the issue under examination.

Some practical considerations follow. We use a smoothed maximum likelihood estimator to estimate $p(f_n|H_k), k = \mathcal{H},\mathcal{L}, n = 1,\dots,K$, from training data as follows $\hat{p}(f_n|H_k) = \frac{N_{n,k}+1}{N_k+V}$, where $N_{n,k}$ denotes the number of issues that give rise to outcome f_n and belong to hypothesis H_k, N_k denotes the total number of issues in the training dataset that belong to hypothesis H_k and V is the maximum value of the nth feature. We estimate the *a priori* probabilities as $P(H_k) = \frac{N_k}{N_{\mathcal{H}}+N_{\mathcal{L}}}, k = \mathcal{H},\mathcal{L}$. Quantizing the interval $[0,1]$ enables the efficient computation of a $K\times d$ matrix, where each row corresponds to K values $\bar{J}_n(\pi_n), n = 0,1,\dots,K-1$, computed using Eq. (7) for different d values of $\pi_n\in[0,1]$. Since this computation requires only *a priori* information, it can be conducted once offline. Hence, the complexity of calculating $\bar{J}_n(\pi_n)$ is independent from the actual number of issues, which can be huge. Finally, different features can hinder or facilitate the quick recognition of the importance of an issue. Consider for example the case of two features f_1 and f_2, where f_1 indicates that the issue refers to 'noise complaints', and f_2 is the number of tags in a report. The type of an issue can potentially simplify the process of evaluating importance compared to the number of tags. As a result, if feature f_2 was to be examined first, it would be very probable for feature f_1 to be examined as well to improve the chances of accurate classification. Alternatively, if f_1 was to be evaluated first, CIvIC could reach a decision using one feature only. To avoid the computational complexity of evaluating all $K!$ possible feature orderings, we sort features in increasing order of the sum of type I and II errors scaled by the cost coefficient of the n feature to promote low cost features that at the same time are expected to result in few errors.

5. CASE STUDY: THE SEECLICKFIX PLATFORM

In this section, we illustrate the performance of CIvIC on a real–world dataset of $2,195$ issues collected from the SeeClickFix platform[1] for the metropolitan area surrounding Albany, the capital of the U.S. state of New York, spanning a time period between Jan 5, 2010 and Feb 10, 2018. We consider $2,594$ features directly extracted from issues' title, description, address, and reported time. Specifically, we tokenized sentences into unigrams, removed punctuation (e.g., periods, commas, and apostrophes), stopwords (e.g., "a", "the", "there"), and digits (e.g., "8th", "31st"), and stemmed each word to its root (e.g., replace "parked" with "park"). Feature values correspond to the number of times a specific word appears in the issue report. We excluded words present in $\geq 95\%$ and $\leq 2\%$ of all issues, respectively. We also considered the logarithm of the number of: 1) words plus one (both for the title and the description), 2) exclamation marks plus one, and 3) uppercase letters plus one. We divided addresses into three geographical dimensions based on the radial distance between issues (namely, 0.5, 1, and 2 miles). We also divided reported time into three dimensions (i.e., 1, 7, and 15 days) based on the time gap in days between issues.

[1] https://seeclickfix.com/albany-county

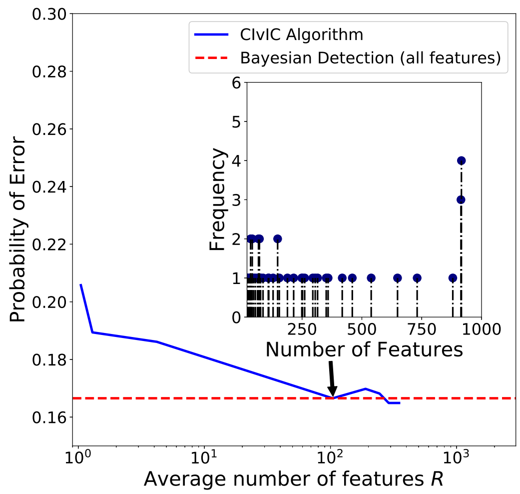

Fig. 1: Probability of error versus average number of features used. Inset shows the distribution of number of features used by CIvIC to classify each issue's importance for an average of ~ 104 features.

For each issue, feature values correspond to the number of issues falling into the 3×3 grid of geographical and time dimensions. Finally, we considered a neighborhood categorical variable encoded by the unit vector $\mathbf{e}_i, i \in \{1, \ldots, 28\}$ with 1 in the ith position and zero elsewhere, and a binary variable indicating if the reported day is on weekend (i.e., 1 if the day is on weekend and 0 otherwise).

We compare CIvIC's performance to (i) a standard Bayesian detection method [18] that uses all available features, (ii) prior work (i.e., Support Vector Machine with feature selection (SVM–FS) [14]) with variations, i.e., linear (SVM–linear) and Gaussian (SVM–Gaussian) kernels, and dimensionality reduction with PCA (SVM–PCA), and (iii) tree–based classifiers, namely Random Forest (RF) with maximum tree depth 5 and 10, and XG Boosting, which have been shown to achieve good performance in classification tasks, while being relatively fast compared to other classification models [19, 20]. In our experiments, $L = 2$ (i.e., issues are of high, $H_{\mathcal{H}}$, or low, $H_{\mathcal{L}}$, importance) and varying feature costs $c_n \in \{0, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.25\}$ and misclassification costs $M_{\mathcal{H}1} = M_{\mathcal{L}2} = 0$, $M_{\mathcal{H}2} = M_{\mathcal{L}1} = 1$ are considered. Five–fold cross validation results are reported. Importance of an issue is based on the number of votes and comments received; we discretized importance based on predefined thresholds, i.e., an issue belongs to $H_{\mathcal{H}}$ if number of votes $V > \overline{V}$ and number of comments $C > \overline{C}$, otherwise it belongs to $H_{\mathcal{L}}$. To test the robustness of our algorithm, we considered 4 scenarios of varying thresholds $\overline{V}$ and $\overline{C}$. An approximately balanced training and testing dataset was created in each case by randomly undersampling the majority class. We observed that both CIvIC and all baselines exhibited similar performance in all cases. Due to space limitations, we report results only for $\overline{V} = \overline{C} = 15$.

Fig. 1 shows the error probability achieved by CIvIC as the average number of features used increases. The error probability achieved by the standard Bayesian method that uses all available features is also included for comparison. As expected, when the average number of features used is small, CIvIC exhibits large error probability. However, as this number increases, performance improves dramatically. The inset in Fig. 1 illustrates the number of features used by CIvIC to recognize an issue's importance for an average number of ~ 104 features achieving the same performance as the standard Bayesian method. Table 1 summarizes the performance of CIvIC compared to baselines. Among all baselines, the Bayesian detection approach that uses all features achieves the highest accuracy (83.3%) and precision (81.9%), while CIvIC can achieve same accuracy and precision using on average ~ 104 and ~ 289 features (i.e., 96% and 88.8% reduction in number of features), respectively. Last but not least, SVM–Gauss achieves the highest recall (91.6%), but requires ~25 times as many features for a mere 3% improvement in recall while sustaining 7.7% and 4.4% degradation in precision and accuracy, respectively, compared to CIvIC.

Table 1: Performance comparison of CIvIC with baselines.

Method	Accuracy	Precision	Recall	Avg. # feat.
CIvIC ($c = 0.25$)	0.794	0.785	0.818	1.05
CIvIC ($c = 10^{-1}$)	0.811	0.789	0.854	1.29
CIvIC ($c = 10^{-2}$)	0.814	0.783	0.873	4.19
CIvIC ($c = 10^{-3}$)	**0.833**	**0.801**	**0.889**	**104.10**
CIvIC ($c = 10^{-4}$)	0.830	0.807	0.870	189.78
CIvIC ($c = 10^{-5}$)	0.832	0.811	0.867	244.99
CIvIC ($c = 10^{-6}$)	**0.835**	**0.819**	0.864	289.59
CIvIC ($c = 0$)	**0.835**	**0.819**	0.864	350.34
Bayes [18]	0.833	**0.819**	0.860	2,594
SVM-FS [14]	0.746	0.701	0.810	20
SVM-linear	0.806	0.801	0.815	2,594
SVM-Gaussian	0.796	0.739	**0.916**	2,594
SVM-PCA	0.825	0.791	0.886	208
RF (depth=5)	0.815	0.779	0.883	2,594
RF (depth=10)	0.820	0.784	0.886	2,594
XG Boosting	0.827	0.801	0.873	2,594

6. CONCLUSION

In this work, the problem of automatic recognition of the importance of urban issues in civic engagement platforms is addressed. Specifically, a novel formulation based on optimal stopping theory is proposed, where the optimization function is defined in terms of the cost of evaluating features and the Bayes risk associated with the classification rule. The optimal classification and stopping rules are derived and a near–real–time algorithm, CIvIC, is devised that implements the optimal rules. Evaluation on a real–world dataset from the SeeClickFix civic engagement platform confirms CIvIC's reduced time–to–detection and accurate recognition performance. In future work, we plan to extend our framework to enable multi–valued importance recognition and devise appropriate learning–to–rank approaches to dynamically order incoming urban issues requests.

7. REFERENCES

[1] S. A. Chun, S. Shulman, R. Sandoval, and E. Hovy, "Government 2.0: Making Connections Between Citizens, Data and Government," *Info. Pol.*, vol. 15, no. 1,2, pp. 1–9, April 2010.

[2] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," 2006.

[3] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: participatory urban sensing using a network of handheld air quality monitors," in *7th ACM conference on embedded networked sensor systems*. ACM, 2009, pp. 349–350.

[4] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "GreenGPS: a participatory sensing fuel-efficient maps application," in *8th international conference on mobile systems, applications, and services*. ACM, 2010, pp. 151–164.

[5] P. Zhou, Y. Zheng, and M. Li, "How long to wait? Predicting bus arrival time with mobile phone based participatory sensing," *IEEE Transactions on Mobile Computing*, vol. 13, no. 6, pp. 1228–1241, 2014.

[6] Ines Mergel, "Distributed democracy: Seeclickfix. com for crowdsourced issue reporting," 2012.

[7] 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*. ACM, 2007, pp. 72–80.

[8] A. L. Kavanaugh, E. A. Fox, S. D. Sheetz, S. Yang, L. T. Li, D. J. Shoemaker, A. Natsev, and L. Xie, "Social media use by government: From the routine to the critical," *Government Information Quarterly*, vol. 29, no. 4, pp. 480–491, 2012.

[9] D. C. Brabham, "A model for leveraging online communities," *The participatory cultures handbook*, vol. 120, 2012.

[10] 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.

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

[12] 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*, October 2018.

[13] M. Budde, J. D. M. Borges, S. Tomov, T. Riedel, and M. Beigl, "Improving Participatory Urban Infrastructure Monitoring through Spatio-Temporal Analytics," in *3rd ACM SIGKDD International Workshop on Urban Computing*. ACM, 2014.

[14] 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, ACM.

[15] C. Masdeval and A. Veloso, "Mining citizen emotions to estimate the urgency of urban issues," *Information systems*, vol. 54, pp. 147–155, 2015.

[16] A. N. Shiryaev, *Optimal Stopping Rules*, vol. 8, Springer Science & Business Media, 2007.

[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] 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," in *ACM International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.