ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Improvisation of Classification Accuracy for Online Purchasing Platform using Crowd sourcing

[1]Dr.P.Gokulakrishnan, [2]Dr.A.Thomas Paul Roy, [3]Dr.D.Suresh

Abstract— Crowdsourcing is a model where individuals cum organizations gather information such as ideas, micro-tasks, financial, voting related to goods and services from participants of large, open and rapidly-evolving nature. It involves usage of internet receive and distribute work between participants to get a collective result.The application of classification tasksin crowdsourcing is a counter step due to the increase in popularity of crowdsourcing market.Dynamic Label Acquisition and Answer Aggregation (DLTA) crowdsourcing framework accomplishes the classification task in a promising manner. But most of the existing works are not able to provide a proper budget allocation for labels because theydo not exploit the Label inference and acquisition phase.In addition, label mismatch and multi-label tasks are the other problems encountered in the existing works. To overcome, it is proposed to adopt Random Forest Algorithm (RFA) for classification in crowdsourcing. The objective of this work is to improve the crowdsourcing classification task efficiency with Dynamic Resource Algorithm.RFA is activated by constructing a multitude of decision tree at training time and results with the classes and itapplies a bagging technique to produce the ultimate result with highest accuracy.

Index terms: Crowdsourcing classification, Random Forest Algorithm,Bagging Techniques, Label Acquisition, Multi-label tasks.

I.INTRODUCTION

Crowdsourcing is a worthy platform to address tasks utilized by thousands of common employees or users (i.e., the crowd). Crowdsourcing is engaging a group or crowd to achieve a corporate goal, which may be efficiency, innovation, or problem solving. Public crowdsourcing platforms, like Amazon Mechanical Turk (MTurk), Crowd Flower and Up-work are exercising the crowdsourcing marketplace to ail individuals cum businesses for outsourcing their jobs and processes to a distributed workforce for virtual task performance.Crowdsourcing also profitsinformation management applications like information cleansing, information integration, data construction.

Consider the entity resolution as associate degree example, suppose a user (called the "requester") features a set of objects and needs to search out the objects that talk to an equivalent entity, maybe victimization completely different names. Though this downside has been studied for many years, ancient algorithms square measure still off from excellent.

To this finish, the requester 1st styles the tasks (e.g., a task for each combine of objects that asks staff to point whether the two objects talk to an equivalent entity.) or not Then the requester publishes the tasks on a crowd sourcing platform like AMT Crowd staff who square measure willing to perform such tasks(typically for pay or another reward) settle for the tasks, answer them and submit the answer and report them to the requester. The platform collects the answers and delivers to the requester.Because the crowd has discourse data and feature ability, crowd sourced entity resolution will improve the standard.

There are several other problems in crowdsourced data managementas depicted in Figure 1.

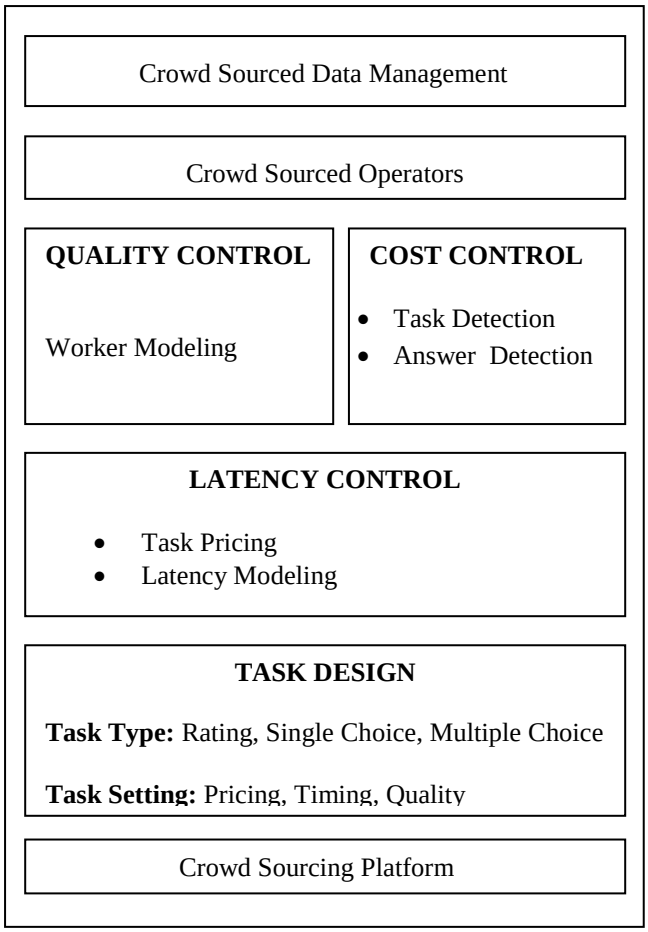

Figure 1: Overview of Crowdsourced data management

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020
[1]Professor, Department of CSE., PSNA College of Engineering and Technology, gokulakrishnan@psnacet.edu.in
[2]Professor, Department of CSE., PSNA College of Engineering and Technology, pauli.dgl@gmail.com
[3]Professor, Department of CSE., PSNA College of Engineering and Technology, sureshdgll@psnacet.edu.in

Internal Control:

Crowdsourcing could yield comparatively low-quality results or maybe noisy. For instances, a malicious employee could on purpose provide wrong answers. Staff could have completely different level of experience, associated a primitive employee is also incapable of accomplishing bound tasks.

To attain topquality, tolerate the crowd errors and infer high quality results from strident answers. The primary step of internal control is to characterize an employee's quality (called worker modeling). Then support the standard model of staff, there are many methods for boost quality. Eliminate the low quality staff (called employee elimination), assign a task to multiple staff and combine their answers (called answer aggregation), or assign a task to applicable staff (called task assignment).

Price Management:

The group isn't free, and if there are giant numbers of tasks, crowdsourcing may be pricey. For instance, in entity resolution, if there are ten thousand objects, there will be regarding fifty million pairs. Though the worth per try in one cent, it still takes millions of cash. There are many effective cost control techniques.

The primary is pruning, that uses pc algorithm to get rid of some surplus tasks then utilize the group to answer solely the mandatory tasks. The second is task choice that prioritizes those tasks to crowd source. The third is answer deduction that crowd sources the set of tasks and supported the answers collected from the group; deduce the results of alternative tasks. The fourth is sampling, that samples a set of tasks to crowd source. There also some specialized cost control techniques chiefly designed to optimize for specific operators.

Latency Management:

Crowdanswers could incur excessive latency for many reasons: for instance, man power is also distracted or untouchable, the tasks might not be appealing to enough staff, or the tasks it is not may be tough for many staffs. If the requester incorporates a time constraints necessary to regulate latency. There are many methods for latency management, primary is valuation. Typically a better value attracts a lot of staff and might scale back the latency. The second is latency modeling.

There are chiefly two latency models: the spherical model and applied mathematical model. (a) The spherical model leverages the concept that tasks maybe revealed in multiple rounds. If there are enough staff in crowdsourcing platform, the latency of respondent task in every spherical may be thought to be constant time. Therefore the latency is sculptured because the variety of rounds. (b) The applied mathematical model is additionally wont to be model latency that leverages the collected statistics from previous crowdsourcing tasks to create applied mathematics model that may capture the worker's point in time, the completion time, etc. These delivered models will then be wont to predict and may be regulated for expected latency. Three extra parts of crowd sourced knowledge management are: task style, crowd sourced operator style, and optimization.

Given a task (e.g. entity resolution), task style aims to style effective task sorts (e.g. making a YES/NO questions associated asking staff to pick out an answer). Task style additionally must set the properties of tasks, that isdeciding the cost, setting the time constraint, and selecting quality-control ways.

II.RELATED WORKS

LibonZheng proposed a DLTA Framework for Dynamic Crowd sourcing Classification Tasks [1]. The framework proceeds in a sequence of round robin (GRR) andalso with label inference and label acquisition. For each round it collected the answer of previous rounds and analyzes to perform proper budget allocation and delivers the resultant query to the crowd.The major advantage of the system is to propose a generative model for label collections and also depict the corresponding strategies for label inference and budget allocation and experimental results show that compared with existing methods. DLTA Approachdoes not exploit the label inference and acquisition phase. This results in inability of making a proper budget allocation for labels. Label mismatch, Multi-label task are some of the additional problems faced in the DLTA Framework.

Mohammad Asghariand CyrusShahabi proposed an On-line Task Assignment in Spatial Crowdsourcing[2]. Auction based algorithm (ABA) is one in which split the scheduling responsibilities and matching between the spatial crowd server and workers. Spatial crowdsourcing involves millions of workers and tasks and also allocate the task to a worker based on matching task to worker and computing a schedule for each worker. If each task is performed instantly, the challenge is scheduling the task. The On-line Task Assignment designed for the workers, who travel to a single location. But the other application needsa worker travel to more than one place. For example inUber application , the worker has to pick a passenger in one place and drop them in other place. So, it needs a extended auction-based framework.

Megan K.O ' Brien and Christian Poellabauer proposed a Detecting Errors in Crowdsourcing Smart Phone Sensor Data[3], Which provide the guide line for developing effective techniques to identify and remove the label errors in smart phone sensor data. It works on label errors using Supervised Learning Algorithm called Data Corruption and Ensemble algorithm (DCE) to limit the generation of falseoutputs in the presence of label errors. The label error is occur if the label iscorrupted accidently or deliberately. It solves the issue by Supervised Learning for auction recognition. It converts training scheme into four groups of sub classifiers. The sub classifiers does not specify the following attributes

a) The simulated data set may not fully reflect real world label errors.

b) The model of the router could be flawed.
c) The data set has an imbalanced distribution.
d) Small data set may limit the classifier's performance.

Ting Wu et al proposed a Object Identificationwith Pay-As-You-Go crowdsourcing [4]. It deploys the new crowdsourcing paradigm for OI tasks, named Adaptive Worker Assignment. It address the two problems

a) Assigning the questions to the best workers with the consideration of the Pay-As-You-Go payment scheme.
b) Design termination criteria and verify the crowd sourced results from previous workers.

The first issue solved by near-optimal algorithm which determines a near-optimal set of workers to be crowd sourced , and greedy strategy used for select and crowd sources a near-optimal set of workers. For the second problem we need to determine, when to stop asking questions regarding each worker. To find this majority voting foraggregating crowdsourced data and imposed a threshold so that terminate crowdsourcing. Each workerterminates the process, only if the probability of answer is being correct and also no less than the given threshold.

XianoDuan and Keisha Tajima proposed a Hierarchical Reorganization For Improving The Classification Accuracy in crowdsourcing[5]. It focuses on the accuracy of multi classclassification tasks in crowd sourcing by Worker allocation algorithm (WAA). It recognize each task into hierarchical classification tasks and assign the workers to appropriate sub-tasks in the hierarchy. The main objective of the work is to assign workers to tasks they are good at. Because, different workers are good at different categories. It converts flat classification tasks into hierarchical classification tasks and assign workers to sub-tasks that they are good at. It is only the idea for improving classification accuracy. So, the validation of the idea is done by real task execution on the crowdsourcing platform.

AntonellaFrisiello et al defined a Gamified crowdsourcing for Disaster Risk Management[6]. It proposes aGamified Strategy for crowd sourced disaster risk management services aimed to increase awareness, engagement, and change people behaviors. It is aimed to strengthen the citizen engagement towards topics related to co-operative environmental monitoring, risk awareness and natural hazards. The goals of Gamified strategy is citizen's awareness, citizen's engagement in reporting , report validation. It includes 3actors (players, spectators, observers). The review applies the main ingredients given by current literature, and stimulating an active attitude and self-protection behaviors. It monitors data collection for natural hazards through crowdsourcing, which reduce impacts in terms of human and economic losses in case of natural disasters. The future work includes implementation and evaluation of gamified mobile application for crowd sourcing. The evaluation is done by user experience, engagement and quality of the contents produced.

RaffaellaGuida et alproposed a Remote Sensing and Crowdsourcing [7]. In this work crowdsourcing is expressed as remote sensing project to manage water quality in Africa, Where the truths are collected from the training people in local communities. The Remote Sensing Algorithm (RSA) or Detection algorithm is used to validate the ground truths. It proceeds in two perspectives:

a) Monitor the water qualityof its main reservoir.
b) Monitor the incidence of water borne diseases in children.

In comparison analysis not all the data have been returned. The next step is correlate the results on quality of water sample and answers to the questionnaires submitted by the children.
Hien et al define a Acquisition and Analysis of Data For Disaster Response[8]. The increasing popularity of mobile devices, crowdsourcing data collection and analysis emerged as ascalable solution. The Acquisition and Analysis phase address the two challenges.

a) Transmission under band width scarcity caused by damaged communication network and prioritizing the visual data collection.
b) Analyze the acquired data intimely manner.

The analytical model is provided to quantify the visual awareness of a video based on its meta data. For acquiring most relevant data under bandwidth constrains, leads to visual awareness maximization problem.Future work includes study of crowdsourcing technologies which includes both the analysts at the command centre and the controlled workers at the disaster site to answer some open questions, justifies who to ask and where to collect the data in disasters.

ShyamalaRamachandranand .Sasireka presented Descriptive Study and Analysis of Crowdsourcing Technique [9]. It involves mobile crowdsourcing techniques like Task Assignment Based Methods, Group Based Requirement System, and Green Mobile Crowd Sensing Based Techniques.Descriptive Study and Analysis of Crowdsourcing Technique provide an elaborative analysis and discussion are made with evaluation metrics, utilized datasets, employed methods, implementation and energy consumption, publication year.

Eventually analyze the research gaps and issues of various mobile crowdsourcing techniques.The limitations of Task Assignment Based Technique is quality of service metric is not considered. The major challenge in Green Mobile Crowd Sensing Based Technique is does not measure the energy saving for geographical sensors. The major challenge in other Mobile Crowd SensingTechniques is prior Knowledge and accurate model is not considered.The future

work includes, further scope for mobile crowd sensing techniques by considering issues and research gaps.

C.Bielskiet al proposeda Coupling Early Warning Services, Crowdsourcing andModeling for Improved Decision Support and Wildfire Emergency Management[10]. This describes the Wildlife Monitoring for Emergency Management System in Wildlife Disasters. It is developed for European Forest Fire Disasters, Based on the integration of information from different sources, data processing chains and decision support systems. The system design is implemented by Data Processing and Analysis Chain phase.

The information management is controlled by Processing Description phase. It includes weather forecasting, the fire weather index, monitoring hotspots, confirmation of wildfire, wildfire hazards nowcasting and forecasting, wildfire risk impact mapping, wildfire disaster decision support, burned area mapping, citizen involvement. The future work includes new algorithms and available data sources, which easily adopted and ingested to improve Wild Fire Crisis Management.

NingXuet al proposed a Ontological Bagging Approach for Image Classification of Crowd sourced Data[11]. It proceeds by ontological bagging algorithm(OBA), it works by learning the most weak attributes for different semantic levels by multiple instance learning and the error propagation of hierarchical classifiers are reduced by bagging idea. The main advantage is, it learns discriminative features on each level of ontology using multiple instance learning, and classifies categories from coarse to fine semantic grains based on these features, Which mimics the human visual system.The future work is, to evaluate the work on otherrealistic image data sets.

Markus herb et al proposed a Crowd Sourced Semantic Edge mapping for autonomous vehicles [12]. It proceeds by a Novel method to derive a detailed high definition maps by crowdsourcing data using commodity sensors. It uses multi-session feature based visual SLAM to align sub maps recorded by individual vehicles on a central backend server. The future work is to evaluate a method on real world data, which contain high level of detail and metric accuracy.

Sumit Mishra et al proposed a Non-dominated sorting for Worker Selection in Crowd Sourced platforms[13]. The aim of the work is to perform a non-dominated sorting of the workers based on the requirement.From this set of ordered workers domination count is used to select the best set of workers that can perform the task. The future work includes to generalize the concept of band by incorporating the workers of the first to k − 1th level while selecting the workers from kth level.

Xiangpeng et al proposed a [14], which improves the navigation of vehicles in urban areas and collects the highamount of data by on-board and infrastructure based sensors for evaluate traffic network statuses.The main objective is to depict a real-time route planning algorithms, which determines the best trajectory in a real-time depends on the frequent data inputs. The future work is to focus on designing data processing technique at the back-end level to deal with the unavailability and errors of the reported data.

III.PROPOSED SYSTEM

Crowdsourcing techniques are focuses on improving the efficiency and classification accuracy of crowdsourcing applications.The Figure 2 shows the Architecture of the proposed work.Architectural works represents set of concepts, which includes their principles, elements and components.Proposed system has three modules, they are

A. Pre-Processing module
B. Feature Extraction module
C. Classification module

A. PRE-PROCESSING

Creating the platform for online purchasing is the initial process. Pre-processing deals with creating

- Process of Admin
- Registration of User
- Purchasing of Product
- Getting reviews.

Process of admin:

Admin plays a major role and has the responsible of adding reliable products as per corresponding categories.User can buy their required products and can post the review about the products.

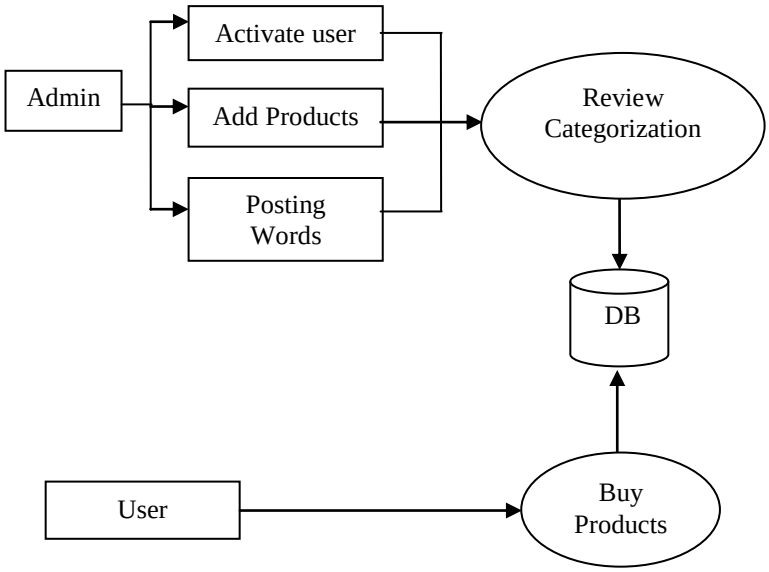

Figure 2:Architectureof Proposed System

The admin has all privileges including the ability to block the user's account and add products.

Registration of User:

User Registration modulerequires credential of the user like username, password.

Purchasing of Product:

ISSN: 2278-4632
Vol-10 Issue-6 No. 8 June 2020

Product Purchasemoduleallowsthe user to view the listed items that are provided by the admin.

Getting Reviews:

User can post a review after bought a product, in terms of realizing the features of the product.

B.FEATURE EXTRACTION

Feature Extraction can be done by C4.5 (J48) Algorithm, Which defines the product added by the admin, isbelonging to Electronics or Appliances or Accessories, etc.Also provide the sales platform includes getting Reviews. C4. 5 works by creates decision tree using the expected values of the class.

C.CLASSIFICATION

In classification,the getting reviews are classified intopositive and negative comments. Classification is achieved by Random Forest Algorithm, Whichis operates by constructing a multitude of decision tree at training time andalso resultsthe mode of classes.

Working Principle:

Random Forest applies "Bootstrap Aggregating" or "Bagging". The training set and their response is given by (1) and (2).

Given training set **T=** $t_1, t_2 \ldots t_n$ **(1)**

Response **R=** $r_1, r_2 \ldots r_n$ **(2)**

Bagging repeatedly selects (B times) a random samples with replacement of the given trainingset.

For b=1, 2 …B

1. Sample with replacement, n training examples from T, R, call these as T_b, R_b.
2. Train a classification f_b on T_b, R_b.

After training, predictions for unseen samples t' can be made by averaging the prediction from the entire individual regression tree ont' by equation (3).

$$\hat{\hat{f}} = \sum_{b=1}^{B} f_{b\,(t')} \quad (3)$$

The next step is "Feature Bagging", for correlation handling.If one or few features are very strong predictor for response variable, these features will be selected in many of the B trees, causing them to become correlated, An analysis of how bagging and random subspace contribute to Accuracy gains under different condition.

IV.RESULTS AND DISCUSSION

Experimental Setup:

Net Beans is used for this implementation with the help of **MySql** and **JSP**. MySql is open source relational database management system, used to accessing and managing the DB. MySql database is accessed by many programming languages with language specific APIs contains libraries .JSP (Java Server Pages) is a programming technology in server side used for creates a platform- independent web based applications and also helps developers to creates dynamically generated web pages based on HTML, XML ,etc.

An initial experiment is conducted to verify the effectiveness of the proposed work in terms of accuracy. Proposed algorithm works by realizing the comments given by the end users, with the help of decision trees. Accuracy of classification is achieved by no of positive comments and negatives comments. The proposed algorithm achieves 87% accuracy in the crowd sourcing classification, which is higher than that of the existing algorithms.

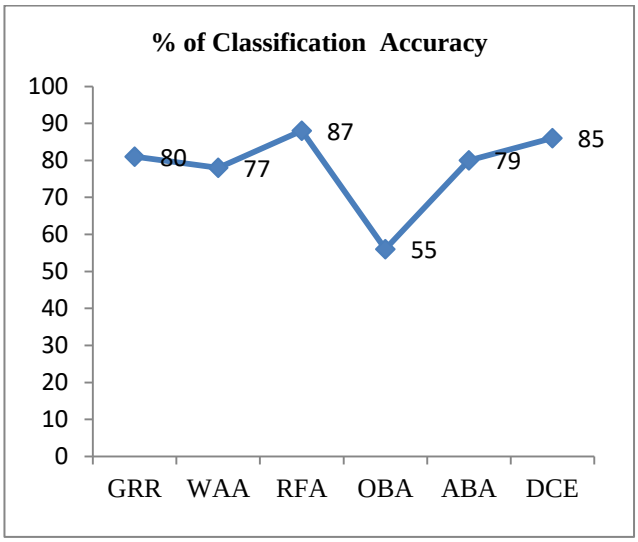

Figure 3: Comparison between RFA with other algorithms

The abbreviations of the existing algorithms are justified in chapter II (Related Works). RFA is compared with these algorithms only on the basis of accuracy as shown in the Figure 3.

V.CONCLUSION

Crowdsourcing is obtaining the information or input of a particular task given by the large number of peoplewhom are involved in services, typically via the Internet. The application of the classification tasksenablestheincreasing popularity of crowdsourcing market.Random Forest Algorithm (RFA) used for classification in crowdsourcingand improvement of efficiency is accomplished by Dynamic Resource algorithm.This makes a best analysis for classification in crowdsourcing and improves the classification accuracy in crowdsourcing applications. When compared to other algorithms,it achieves higher percentage of accuracy in crowdsourcing classification. The classified reviews are analyzed by to verify the quality of the product or items. The performance of the proposed work achieves a promising result of accuracy of 87 percentage.In future, it is proposed to device

this framework for a variety of applications in different crowdsourcing platforms.

VI.REFERENCES

[1]. LibinZheng, Lei Chen,DLTA: A Framework for Dynamic Crowd sourcing Classification Tasks. VOL 31, NOV 5, MAY 2019.

[2]. Mohammad Asghari, Cyrus Shahabi, On On-line Task Assignment in Spatial Crowd sourcing, 2017.

[3]. Xiao Bo, Christian Poellabauer, Detecting Label Errors in Crowd-Sourced Smartphone Sensor Data. International Workshop on Social Sensing, 2018.

[4]. Ting Wu, Chen Jason Zhang, Lei Chen, Pan Hui and Siyuan Liu, Object Identification with Pay-As-You-Go Crowd sourcing, 2016 IEEE International Conference on Big Data (Big Data).

[5]. XiaoniDuan, Keishi Tajima, Improving Classification Accuracy in Crowd sourcing through Hierarchical Reorganization,978-5386-2715-0/17, IEEE International Conference on Big data,2017.

[6]. Antonella Frisiello, Quynh Nhu Nguyen, Claudio Rossi, Fabrizio Dominici,GamifiedCrowd sourcing for Disaster Risk Management,2017 .

[7]. RaffaellaGuida, Peter T.B. Brett and Salman S. Khan,remote sensing and crowd sourcing, IEEE 978-1-4799-1114-1/3.

[8]. Hien To, Seon Ho Kim, Cyrus Shahabi,Effectively Crowd sourcing the Acquisition and Analysis of Visual Data for Disaster Response, IEEE International Conference on Big Data (BIG DATA) , 2015.

[9]. Dr.ShyamalaRamachandran, V. Sasireka2, Descriptive Study and Analysis of Crowd Sourcing Techniques in Mobile Social Media Networks, IEEE Transactions on Mobile Computing, 2017.

[10]. C.Bielski, V.O'Brien, C.Whitmore, K.Ylinen,I.Juga, Coupling Early Warning Services, Crowdsourcing, and Modelling for Improved Decision Support and Wildfire Emergency Management,2017.

[11]. NingXu, Jiangping Wang, An Ontological Bagging Approach for Image Classification of Crowd sourced Data, National Science Foundation Grant DBI 10-62351, and U.S.

[12]. Markus herb, Tobias weiherer, Nassir Navab, Federico Tombari, Crowd Sourced Semantic Edge mapping for autonomous vehicles , IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019.

[13]. Sumit Mishra, AkashYadav, Ashok Singh Sairam, Worker Selection in Crowd Sourced platforms using Non-dominated Sorting, 978-1-7281-1895-6/19,IEEE Region 10 Conference (TENCON), 2019.

[14]. Xiangpeng wan, Hakim Ghazzai, YehiaMassoud, Real Time Navigation in Urban Areas Using Mobile Crowd Sourced Data, 978-1-5386-8396-5/19, IEEE International Systems Conference (SysCon), 2019.

