

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329262080>

Trust me, my neighbors say it's raining outside: ensuring data trustworthiness for crowdsourced weather stations

Conference Paper · November 2018

DOI: 10.1145/3276774.3276792

CITATION

1

READS

48

3 authors:



Alexander Byron Chen
University of Virginia

3 PUBLICATIONS 1 CITATION

SEE PROFILE



Madhur Behl
University of Virginia

44 PUBLICATIONS 337 CITATIONS

SEE PROFILE



Jonathan L. Goodall
University of Virginia

104 PUBLICATIONS 1,362 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



DataNet Federation Consortium [View project](#)



Overcoming Social and Technical Barriers for the Broad Adoption of Smart Stormwater Systems [View project](#)

Trust Me, My Neighbors Say It's Raining Outside: Ensuring Data Trustworthiness for Crowdsourced Weather Stations

Alexander B. Chen
Civil and Environmental
Engineering
University of Virginia
Charlottesville, Virginia
abc8fq@virginia.edu

Madhur Behl
Computer Science
University of Virginia
Charlottesville, Virginia
madhur.behl@virginia.edu

Jonathan L. Goodall
Civil and Environmental
Engineering
University of Virginia
Charlottesville, Virginia
goodall@virginia.edu

ABSTRACT

Decision making in utilities, municipal, and energy companies depends on accurate and trustworthy weather information and predictions. Recently, crowdsourced personal weather stations (PWS) are being increasingly used to provide a higher spatial and temporal resolution of weather measurements. However, tools and methods to ensure the trustworthiness of the crowdsourced data in real-time are lacking. In this paper, we present a Reputation System for Crowdsourced Rainfall Networks (RSCRN) to assign trust scores to personal weather stations in a region. Using real PWS data from the Weather Underground service in the high flood risk region of Norfolk, Virginia, we evaluate the performance of the proposed RSCRN. The proposed method is able to converge to a confident trust score for a PWS within 10-20 observations after installation. Collectively, the results indicate that the trust score derived from the RSCRN can reflect the collective measure of trustworthiness to the PWS, ensuring both useful and trustworthy data for modeling and decision-making in the future.

CCS CONCEPTS

• **Information systems** → *Sensor networks*; • **Human-centered computing** → *Reputation systems*;

KEYWORDS

Crowdsourcing, Sensor Networks, Reputation System, Trust, Rainfall Networks

1 INTRODUCTION

Personal weather stations (PWSs) are being increasingly used to provide a higher spatial and temporal resolution of rainfall measurements [5]. This spread of crowdsourced data collection using sensor-packed Internet of Things (IoT) devices is leading to more real time, hyper-local stormwater management for the built environment based on sophisticated flood prediction

models [1]. Consider the example of one of the largest crowdsourced weather networks - Weather Underground - which includes over 250,000 PWSs owned and operated by private citizens. Weather Underground experienced a growth from 7,000 PWSs to the 0.25 million stations in less than 10 years. This growth can be attributed to the availability of user-friendly and affordable off-the-shelf weather stations and a growing number of weather enthusiasts and citizen scientists.

However, there are several barriers associated with the use of crowdsourced rainfall data from PWS networks [8]. Unlike the stations owned and managed by experts, it is difficult and cumbersome to enforce standardization of PWSs at large scales. At best, certain "best practices" can be recommended to hobbyist for data collection and maintenance, but these practices are also difficult to enforce. According to the description on their website, Weather Underground performs fairly simple (average based) quality control to only temporarily removes stations from the network that fail the quality control checks [2], leaving the trustworthiness of data unspecified. Traditional quality control and quality assurance methods require a wealth of historical data measured with validated equipment, which are usually unavailable for the PWS networks.

Consequently, in this paper we present a reputation system for crowdsourced rainfall networks (RSCRN). The proposed method computes and assigns a trust score for the PWS in an online manner, and therefore capturing the integrity of published data for modelers and decision-makers.

2 PROBLEM DEFINITION

The goal of RSCRN is to assign trust scores to each PWS based on the data it provides to the network. This trust score represents a collective measure of trustworthiness of the station, which can be beneficial for decision-making and risk analysis in the future. In this section, we present the problem formulation for assigning trust scores to PWSs using the RSCRN. The overview of the RSCRN workflow is shown in Figure 1.

2.1 Problem Formulation

We consider a PWS network (e.g. Norfolk, Virginia) having N stations. Each station has a different data availability based on when they were installed, or when they reported any data. For a given set of X days, we find the number of stations (say n) which reported some data for each of the X days. The rainfall measurements from these n stations are collected into a matrix P , with X rows for each of the days, and n columns for each of the stations.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BuildSys '18, November 7–8, 2018, Shenzhen, China
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5951-1/18/11... \$15.00
<https://doi.org/10.1145/3276774.3276792>

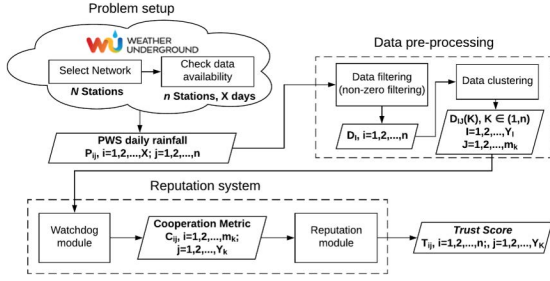


Figure 1: Overview of the RSCRN workflow: N stations from a PWS network is first collected from the Weather Underground network; based on the data availability, the collected PWS daily rainfall data-set is then filtered and clustered into sub-data sets; the reputation system finally receives the sub-datasets and outputs trust score for each PWS in the network.

2.2 Data pre-processing

The data pre-processing comprises of two steps - first, we filter the data such that we obtain the maximum number of days for which non-zero rainfall values were reported by the PWSs, and second, we cluster the stations into groups based on the reported data and their relative location.

2.2.1 Data filtering. To filter the data, we define a non-zero rainfall measurement index Z_i , ($i = 1, 2, \dots, X$) such that:

$$Z_i = \sum_{j=1}^n 1(P_{ij} \neq 0)$$

Each Z_i can take a value in the set $\{0, 1, 2, \dots, n\}$. Simply put, for each row of the matrix P , we count the number of stations which report some non-zero rainfall value. Therefore, we obtain values Z_1, Z_2, \dots, Z_X , for each row where any value $Z_i \in 0, 1, \dots, n$.

In doing so the original data-set P is filtered into n different sub-dataset D_I ($I = 1, 2, \dots, n$) using:

$$D_I = \text{matrix rows in } P \text{ where } Z_i = I$$

Each of the sub-dataset D_I will have a length of Y_I days that $Y_I = \text{number of matrix rows in } P \text{ where } Z_i = I$.

The resulting sub-dataset D_I is a matrix with Y_I rows and n columns of rainfall measurement from n stations such that, at least I ($I \leq n$) stations have non-zero daily rainfall measurement in these Y_I days. For example, D_{11} consists, for example $Y_{11} = 78$ days of rainfall measurements during which 11 out of n stations have reported non-zero rainfall.

2.2.2 Data clustering. Next, the k-Means clustering algorithm applied to divide n stations into k clusters using the sub-dataset D_i . The k-Means is an iterative algorithm that finds the center of each cluster, where the within-cluster sum of squares to this center is minimized [6]. The optimized number of clusters k is determined based on the convergence of the input. By using k-means clustering for PWSs based on the reported rainfall data and their relative location in terms of

latitude and longitude, we aim to partition them into stations which report similar data. It is expected that stations which are located close to each other should fall within the same cluster. The resulting clustered sub-dataset, with stations displaying similar rainfall characteristics and close proximity (thus reporting similar data), will be the input for the reputation system.

3 REPUTATION SYSTEM FOR CROWDSOURCED RAINFALL NETWORKS

Reputation can be considered as a collective measure of trustworthiness (in the sense of reliability) based on the referrals or ratings from members in a community. In this paper, we implement a beta reputation system [7] to PWS data to assign trust scores to each station. The reputation system consists of two modules: (i) a watchdog module for detecting cooperative and non-cooperative behaviors of stations within the network, and (ii) a reputation module for managing and representing reputation information.

3.0.1 Watchdog Module. The input of the watchdog module is a sub-dataset cluster D_I^K with Y_I rows and m columns computed from section 2.2.2, representing Y_I days of rainfall measurement from m stations. The watchdog module computes a *cooperative metrics* for every m stations in each epoch ($i = 1, 2, \dots, Y_I$), denoted as C_{ij} , by an outlier detection method. Outlier detection methods can be broadly classified as either model-based or consensus-based techniques [4]. Consensus-based technique is a suitable method for detecting outliers based on nothing but the data reported from the PWSs.

We use the robust averaging algorithm [3] as the outlier detection method. Robust averaging is a type of average value to minimize the impact of malicious and faulty sensors by assigning smaller weights coefficient. This is explained as follows: We first assign an initial default (uniform) weight $w_{i,j}^{l=1} = \frac{1}{m}$ to station j at iteration $l = 1$, where m is the number of station in the clustered data. Next, we compute the robust average $RA_l = \sum_{i=1}^N w_{i,j}^k \cdot P_{i,j}$. We then compute the deviation $v_{i,j}^k = (p_i - \sum_{i=1}^N w_{i,j}^k \cdot p_{i,j})^2$ of station i 's measurement P_i from robust average RA_l . Finally, we compute the new robust weight at iteration $l + 1$, defined as

$$w_{i,j}^{l+1} = \frac{\frac{1}{v_{i,j}^k + \epsilon}}{\sum_{i=1}^N \frac{1}{v_{i,j}^k + \epsilon}} \quad (1)$$

the algorithm continues iterating until the convergence $|w_{i,j}^l - w_{i,j}^{l+1}| < \lambda$ is observed, i.e the robust weights converge to a value with difference less than λ .

The *cooperative metric* is then defined as

$$C_{ij} = \frac{W_{ij} - \overline{W_i}}{\sigma(W_i)} \quad (2)$$

where $\overline{W_i}$ and $\sigma(W_i)$ are the average and standard deviation of the i -th row of the robust weight matrix. This metric demonstrates the level of deviation of the final robust weight from the

naive weight. The resulting *cooperative metric* is used as an input for reputation module.

3.0.2 Reputation Module. The reputation module assumes neutral *trust scores* for every station without the knowledge of any past behaviors (data in this case). As described in the previous section, the *cooperative metric* can be interpreted as a measure of the PWS deviation from the robust average for each epoch. The reputation module integrates these metrics and updates the *trust score* T_{ij} for every m stations in each epoch ($i = 1, 2, \dots, Y_i, j = 1, 2, \dots, m$).

The beta reputation system [7] is adopted for its advantage of simplicity, flexibility, and ability to counter most arbitrary sensor faults [4]. The beta reputation system takes binary outcomes as input, and computes trust scores by updating the beta probability density functions (PDF). The beta PDF is a continuous family of distribution functions indexed by two parameters α and β . It is denoted by $beta(p|\alpha, \beta)$ can be expressed using the gamma function Γ as follows:

$$beta(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

where $0 < p <= 1, \quad \alpha, \beta > 0$ (3)

The expectation value of the beta distribution is given by

$$E(p) = \alpha / (\alpha + \beta)$$

where $0 < E(p) < 1$ (4)

For all stations at epoch $i = 1$, before any data is collected, the prior distribution is assumed to be an uniform beta PDF with $\alpha_1 = 1, \beta_1 = 1$ and $E(p)_1 = 0.5$. This can be interpreted as assuming neutral trust for these stations, which means that the relative frequency of a positive behavior for these stations in the future is 0.5. After observing new data, the posterior distribution will be the beta PDF with updated α and β parameter. In this paper, these parameters are updated using the cooperative metrics C_{ij} , provided from the watchdog module:

$$\begin{aligned} \alpha_{i+1,j} &= \alpha_{ij} + C_{ij} & \text{if } C_{ij} >= 0 \\ \beta_{i+1,j} &= \beta_{ij} + |C_{ij}| & \text{if } C_{ij} < 0 \end{aligned}$$
 (5)

Given the cooperative metrics, the expectation value of the posterior beta PDF becomes $E(p)_{i+1,j} = \frac{\alpha_{i+1,j}}{\alpha_{i+1,j} + \beta_{i+1,j}}$. This indicates the relative frequency of the positive behavior for each station, which is updated to a new value at every epoch. Finally, the *trust score* T_{ij} is computed by re-scaling the expectation value to 0 to 10 for each station j at epoch i , i.e.

$$T_{ij} = 10 \cdot E(p)_{ij}$$
 (6)

We next implement the proposed reputation system method to a real world dataset obtained from personal weather stations, and perform a comprehensive case study.

4 CASE STUDY

Norfolk, Virginia, USA is a city located on the coast at the south-eastern corner of Virginia that houses the world's largest naval base, Naval Station Norfolk. It is also one of the most vulnerable cities to coastal flooding due to land subsidence

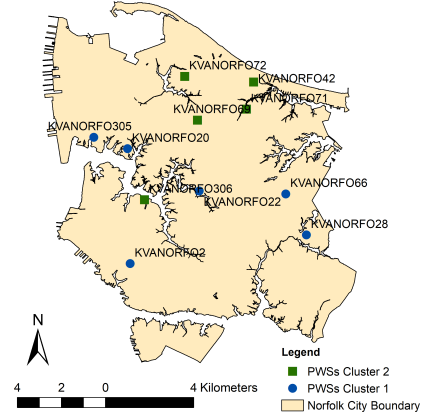


Figure 2: PWSs location in the Norfolk, Virginia region and the PWSs clustering results.

and sea level rise, making it an ideal case study for testing the proposed RSCRN.

4.1 Data Availability

A data-set of 397 days of daily rainfall measurements from $n = 11$ stations during the period from 1/1/2017 to 2/1/2018 were obtained using Weather API made available by Weather Underground. The data-set were collected as a matrix P for this case study. In order minimize the effect of temporal variability, the rainfall measurements were re-sampled from 5-10 minutes to a daily time scale. The data availability for the collected PWS rainfall data sources is shown in Figure 2.

4.2 Data pre-processing

The collected PWS data-set is first pre-processed using the methodology mentioned in Section 2.2.

4.2.1 Data filtering implementation. The rainfall measurements matrix P containing $X = 397$ daily rainfall measurements from $n = 11$ stations, were filtered into $I = 11$ sub-datasets D_I .

4.2.2 Data clustering implementation. k-Means clustering was implemented in Python using scikit-learn library [9] for the sub-dataset D_0 to D_{11} using reported rainfall data, latitude and longitude of each stations. Consider the convergence (the cluster results remain the same for all sub-dataset D_0 to D_{11}) and the minimum stations within a cluster (at least two stations within a cluster), the number of the cluster was set to $k = 2$. Figure 2 shows 11 stations were clustered into two groups. The results of the clustering are consistent with the proximity of stations within the same cluster i.e. stations within the same cluster tend to provide similar rainfall characteristics.

4.3 Results

The reputation system mentioned in Section 3 is implemented to evaluate the trustworthiness of PWS data. The first group of cluster of filtered sub-dataset D_{11}^1 , which consists of $Y_{11} = 78$

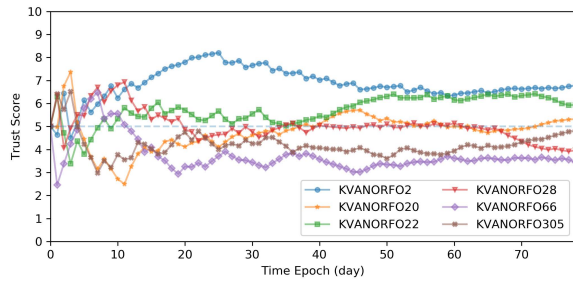


Figure 3: PWSs trust score evolution over time

days of rainfall measurements that at all stations ($m_1 = 6$) have non-zero rainfall measurement was used here for demonstrating the implementation for the reputation system. Each day of the rainfall measurement is treated as an epoch. As described in Section 2, the reputation system is initialized with a neutral trust score $T_{i0} = 5, \alpha = 1, \beta = 1$ for each stations prior to the evaluation. The trust score for each station is then updated based on their rainfall measurements feeding in on every epoch.

The trust score evolution of PWSs over time is shown in Figure 3. It can be observed that, for the first few epochs of the evaluation, the trust scores computed from the reputation system are more sensitive to either cooperative or non-cooperative behavior of the station, causing the trust score to fluctuate significantly. As the cooperative metric are being collected from the PWSs over time, the reputation system takes previous reported data (and trust scores) into account when computing subsequent scores. This shows that the trust toward the PWS was built not only considering its current behavior but also aggregating its past behaviors. *The final trust scores derived from the RSCRN can reflect the degree of trustworthiness of each PWSs, ensuring both useful and trustworthy data for modeling and decision-making.*

Notice that the assigned trust scores are relative to the consensus mechanism, and therefore while it is important to converge to a correct trust score value over time, it is also important to analyze how fast does the score converge to the final value. The evaluation of the convergence of the trust scores is discussed in how many epochs (days) of data are required for reaching a certain convergence threshold. The convergence threshold is defined as the maximum score difference during the score evolution. It can be seen that the trust score of each stations reached a 0.5 score convergence after receiving 10 to 20 days of rainfall measurements, a 0.3 score convergence after 15 to 35 days, and finally reached a 0.15 score convergence after 50 to 70 days. The results of score convergence is shown in Fig 4. *It can be inferred that for a new station in a region, the proposed reputation method can assign an accurate trust score within 10-20 observations of the personal station coming online.*

5 CONCLUSION

This paper presents a reputation system for crowdsourced rainfall networks (RSCRN). The results of the trust score evolution of show that the reputation system successfully identified the

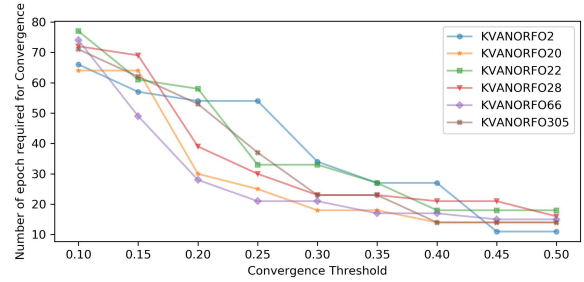


Figure 4: Number of epoch required for reaching different convergence threshold

cooperative and non-cooperative behaviors of the PWSs over time, then gradually converge to a trust score after receiving 10-20 observations of rainfall measurements from the station.

In ongoing and future research, we aim to evaluate the robustness of the proposed RSCRN through simulations of including artificial stations with faulty data. We also aim to incorporate domain knowledge from hydrology and rainfall characteristics into the reputation scheme to validate the results of the RSCRN. Ultimately, application-level benefits can be provided by taking the derived trust score of the RSCRN into account for hydrological and meteorological model simulation and prediction.

ACKNOWLEDGMENTS

This research is supported by the National Science Foundation under Grant No. CBET-1735587.

REFERENCES

- [1] T Bedrina, A Parodi, A Quarati, and A Clematis. 2012. ICT approaches to integrating institutional and non-institutional data services for better understanding of hydro-meteorological phenomena. *Natural Hazards and Earth System Science* 12, 6 (2012), 1961–1968. <https://doi.org/10.5194/nhess-12-1961-2012>
- [2] The Weather Channel. 2018. What is a Gold Star weather station? (2018). https://feedback.weather.com/customer/en/portal/articles/2924672-what-is-a-gold-star-weather-station-?b_id=17298
- [3] Chun Tung Chou, Aleksandar Ignjatovic, and Wen Hu. 2013. Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults. *IEEE Transactions on Parallel and Distributed Systems* 24, 8 (2013), 1525–1534.
- [4] Saurabh Ganeriwal, Laura K Balzano, and Mani B Srivastava. 2008. Reputation-based framework for high integrity sensor networks. *ACM Transactions on Sensor Networks (TOSN)* 4, 3 (2008), 15.
- [5] Mohammad Ghareisifard and Uta Wehn. [n. d.]. ([n. d.]). <https://doi.org/10.1016/J.JHYDROL.2016.01.036>
- [6] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108.
- [7] Audun Josang and Roslan Ismail. 2002. The beta reputation system. In *Proceedings of the 15th bled electronic commerce conference*, Vol. 5. 2502–2511.
- [8] C. L. Muller, L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R. R. Leigh. 2015. Crowdsourcing for climate and atmospheric sciences: Current status and future potential. (sep 2015), 3185–3203 pages. <https://doi.org/10.1002/joc.4210>
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.