2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

# Smart and Connected Water Resource Management via Social Media and Community Engagement

Long H. Nguyen,* Rattikorn Hewett,* Akbar S. Namin,* Nicholas Alvarez,* Cristina Bradatan,† Fang Jin*
* Department of Computer Science, Texas Tech University
† Department of Sociology, Anthropology, and Social Work, Texas Tech University
{long.nguyen, rattikorn.hewett, akbar.namin, nicholas.alvarez, cristina.bradatan, fang.jin}@ttu.edu

*Abstract*—**Water is a critical natural resource that has significant impacts on human living and society. Growing population and energy consumption exacerbate the scarcity of water and our ability to manage this resource. This demonstration paper presents WaterScope, a smart and connected platform for water resource management, which integrates multiple data sources such as water level data, social media data, and water related articles. Furthermore, the tool enables forecasting underground water levels, identifying water concerns, sharing knowledge and expertise among stakeholders, and thus bringing new insights to our understanding and insights of the water supplies and resource management. The prototype engages water stakeholders who face problems of similar nature but deal with the problem in an ad-hoc and isolated manner. The interactive WaterScope platform targets creating an interconnected virtual community that aims to improving water supply resilience.**

## I. Introduction

Water resource problems are extremely complex due to their scope, scale, and interconnection between multiple systems crossing diverse disciplinary and social boundaries. Several problems arise with water supply, usage, conservation, and treatment restraint. However, in a realistic operational environment, these operations are run and managed under different institutions and business entities in an isolated and independent manner. For example, the current water stakeholders do not have a fast lane to access the overall water system. The amount of water that vendors pump depends heavily on the market need, which results in an unbalanced water supply and demand. Furthermore, residential areas need guidance if they are to adopt more economical habits when consuming water. Due to the chaotic situation, it is highly valuable to build a smart and connected water platform for water resource management.

### *A. A Motivational Case Study: Water Shortage in Texas*

The semi-arid Texas High Plains, is an intensive agricultural production area. Approximately 50% of the cotton planted in the region has to be irrigated due to the persistent shortage of water resources. Thus, connecting and integrating the isolated information among water stakeholders in order to optimize water strategy is especially urgent across the West Texas region. We pay special attention to underground water supply as this is the main source of water supply in this area.

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

### *B. Contribution*

We have developed a WaterScope [1] prototype platform to collect, integrate, store and manage a variety of dynamic and heterogeneous water-related datasets such as individual water level data, weather, social media data, and water knowledge-base data resources. Rather than treating each water dataset in isolation, the different dataset behaviors, correlations and interactions are considered in detail and integrated seamlessly utilizing deep learning algorithms to model and predict the water resource level. In particular, the following features make WaterScope a smart and connected platform beneficial for various water stakeholders.

- WaterScope provides a Bird's-eye view of Texas's underground water supply on the map. For each well, users can observe the historical water level changes and post any comments or concerns.
- WaterScope forecasts the future underground water supply using deep learning-based approaches for time series analysis, also called Long Short Term Memory (LSTM) to predict underground (well) water level for the next month. The prediction feature helps in water conservation and protection of the region's water supply.
- WaterScope extracts the concern flow from different water stakeholders in Texas by mining social media data, which helps understanding the key issues during different time periods.
- WaterScope offers an information sharing platform, which can automatically classify users' posts and questions, experiences, and connects water stakeholders in different domains.
- WaterScope helps create an interconnected virtual community that is working towards the common goal of improving water supply resilience in otherwise disconnected groundwater-dependent rural economies.

## II. The WaterScope Framework

There are three main data sources fed into WaterScope: water level data, twitter data and water related knowledge, as shown in Figure 1. The water sensor data is used as a statistical means to measure the water level (the vertical height of the underground water layer compared to the bottom of the water table). Missing data are interpolated before passing them into

[1] http://myweb.ttu.edu/fjin/projects/west-tx-water/

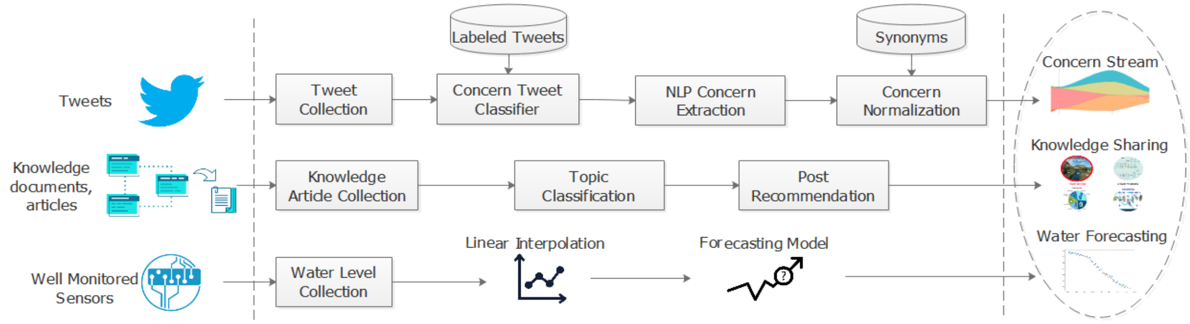



Fig. 1. The WaterScope framework.

the water forecasting model. Furthermore, tweets are ingested to gain insight about water stakeholders concerns, such as water administration, farmers and residents.

The raw tweets will undergo through pre-processing steps and then feed into the concern Tweets classifier to determine whether they are concern tweets. For the labelled concern tweets, concern items need to be extracted utilizing the Stanford Natural Language Processing library [1] and further normalized to ensure the uniqueness of concern. WaterScope also allows water users to post questions, stories, and knowledge, and then automatically classify posts into several categories via topic modeling, and further recommend posts to users based on their interest and relevant. Finally, all these processes are combined into the visualization components to help understand and comprehend water resource management.

## III. Methodology

### A. Datasets and Preprocessing

The datasets are composed of 20 years of water level data from 1995 to 2005 in Texas along Ogallala aquifer, and 3,005 tweets data from April 2016 until March 2018 posted by the Texas Water Development Board. Some of the water level data had missing values due to sensors being out of order. WaterScope applies linear interpolation for the missing values to ease the processes and analysis. Additionally, 204 water related knowledge documents were collected from American Water Work Association website[2] and used as the training set for topic classification, enabling the knowledge sharing feature of the system.

### B. Water Level Forecasting

Underground water levels are dynamically influenced by many factors, including precipitation, weather conditions, water usage and even the water level of neighboring areas. To accurately reflect how each of these factors affect underground water is a fundamental and challenging issue. The Long Short Term Memory (LSTM) network is effective in time series analysis due to its capabilities in extracting both long term and short term patterns, and solving the problem of exploding and vanishing gradients from the earlier Recurrent Neural Network (RNN) system [2]. Therefore, we trained an LSTM network to predict water levels as a quantitative measure for decision makers to plan ahead with regards to water uncertainty.

[2]https://www.awwa.org/

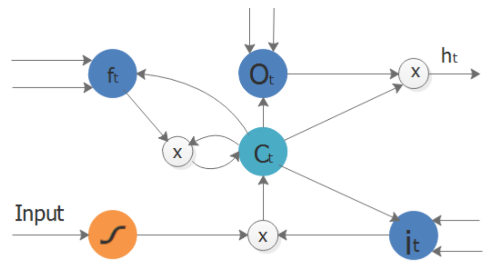



Fig. 2. The LSTM unit architecture.

In addition to the memory cell, the input and output gates, we utilized a customized implementation of the LSTM unit that includes a forget gate in order to address the problem of a continual input stream which is a major characteristics of well water time series data [3]. Figure 2 presents the design of this customized LSTM unit.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{2}$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{3}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{4}$$

$$h_t = o_t tanh(c_t) \tag{5}$$

Equations (1), (2), (3), (4) and (5) present how the value of each element of the unit is updated. Detail of the explanation is described in [4]. Our forecasting model is a stack of three LSTM layers, each contains three LSTM neurons and an output layer. The input is historical water level data (feet) with time stamps, and output is water level forecasting for the next month. The network is trained in 2000 epochs and a batch size of 16 is used. We use *nadam* optimizer and mean squared error to measure loss while fitting the model. The prediction performance is presented in Table I which proves Stacked LSTM model outperforms traditional prediction models such as Linear regression, Lasso, Ridge, and Arima regression.

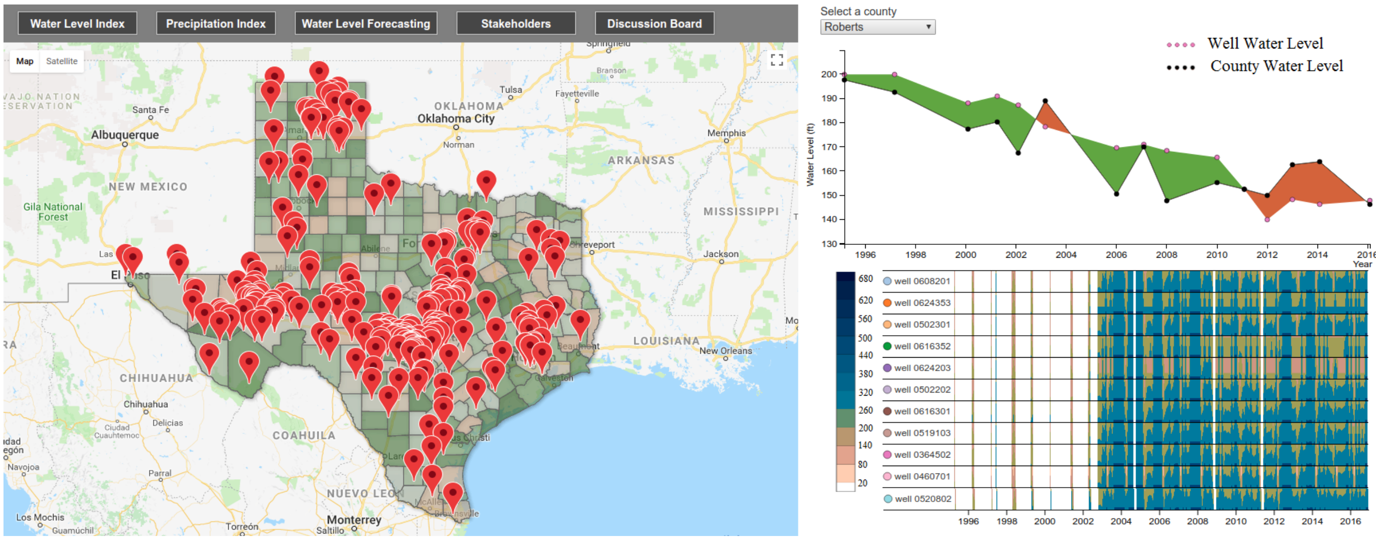



Fig. 3. The WaterScope landing page demonstration.

TABLE I
WATER LEVEL FORECASTING PERFORMANCE COMPARISON. BOLD VALUES REPRESENT BEST PERFORMANCE.

| Method | MAE | MSE | RMSE |
|---|---|---|---|
| Linear Regression | 35.71 | 1869.97 | 43.24 |
| Lasso | 35.78 | 1870.18 | 43.25 |
| Ridge | 37.11 | 1927.56 | 43.90 |
| Arima | **29.90** | 1723.39 | 41.51 |
| Stacked LSTM | 34.06 | **1721.59** | **41.49** |

### C. Water Concern Detecting

*a) Concern Tweet Classification:* To identify water concern related tweets, we developed a supervised classifier using support vector machine (SVM). 3,000 tweets were manually labeled as either concern or non-concern related tweets. Our experiment demonstrated that the SVM classifier performances for accuracy, precision, recall, and f-measure are $0.874, 0.901, 0.893$ and 0.897, respectively.

*b) Concern Term Extraction:* This step will extract the water concern terms from the previously classified tweets. We use a natural-language parser provided from Stanford [1] to determine the part-of-speech (POS) of each tweet data. Doing so, we are able to extract nouns, verbs, etc. that we consider raised concerns among twitters. The concern terms extracted then are parsed through a normalizing process in order to avoid duplicate terms.

### D. Blog Topic Modeling

In the knowledge sharing section, whenever users post some blogs regarding water stories, latest news or questions, the system will extract its topic distributions automatically. It then determines whether its most prevalent topics fall into the set of relevant topics predefined in the system (e.g., water sources, water processing, water usage and water treatment). For the training purpose, we classify a pool of predefined water documents into the four categories. For new posts, we employ latent dirichlet analysis (LDA) to model the similarities [5] and obtain each article's topic distribution.

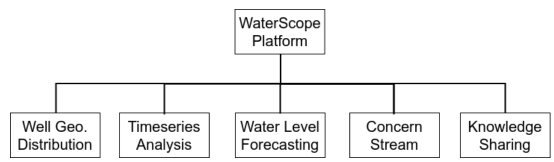
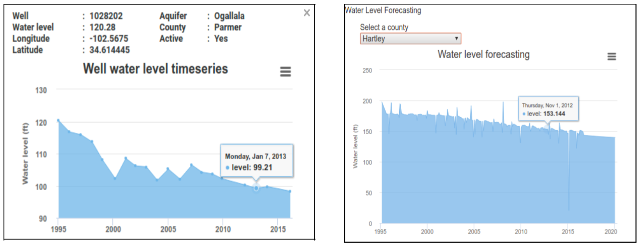



Fig. 4. The WaterScope supported features.

## IV. THE WATERSCOPE PLATFORM

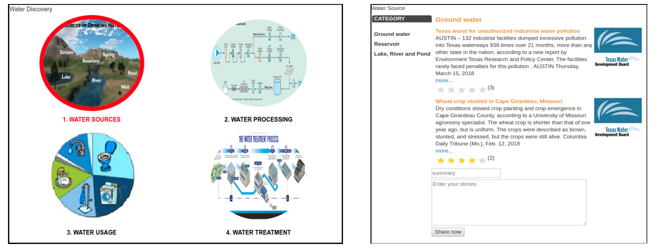
The prototype offers five main functions as shown in Figure 4. The geographical well distribution helps water users locate the wells. Time series analysis and water level forecasting provide trends of water level, while concern stream analysis stresses on bigger scope of problems and solutions to water matters with time stamps. Lastly, knowledge sharing saves water stakeholders time and resources in finding solutions to similar problems. These analytic and multidisciplinary water perspectives help connect water stakeholders to meet their common interest and share their experiences.

As shown in Figure 3, the platform starts with a map and wells distribution across Texas. When users select a county, a list of wells in that county is presented at the right corner. Upon selection, a well's water level is compared with the county's average water level on the top right. The green area represents the period that the well's water level is higher than the county's average water level; whereas, the dark brown area identifies the period of time that the well's water level is below average. At the bottom right, a list of well's water level time series in each county is displayed as a horizon graph. The color band represents water level of the wells from lighter colors to darker colors. The darker color means the deeper water level it was in that period.




(a) Well water level time series. (b) County water level forecast.

Fig. 5. Water level timeseries and forecasting.




(a) Discussion topics. (b) Subtopics.

Fig. 7. Information shared platform.

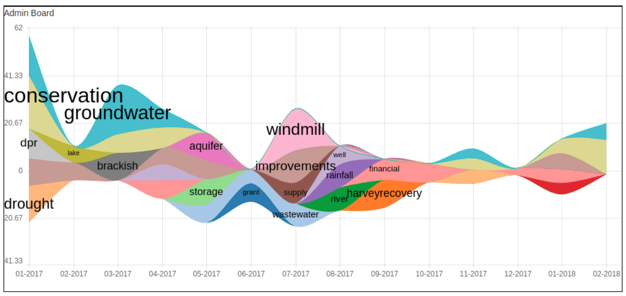



Fig. 6. The concern stream.

### A. Water level forecasting

Wells are positioned with respect to their latitude and longitude in the Google map. Users are also able to mouseover each well to see its time series water level as illustrated in Figure 5(a). In the homepage (Figure 3), after clicking the “Water Level Forecasting” button, a pop-up window will appear presenting the water level prediction of the selected county. As shown in Figure 5(b), a county is required to be selected prior to forecasting. The presented water level forecasting chart is a combination of the sensor data and the predicted data plots. WaterScope forecasts the county’s water level with lead time of one month. It is recurred until year 2020. This will let water users foresee the water trend so that they may plan for water conservation or vegetation irrigation.

### B. Concern analysis

This feature provides water stakeholders a way to understand concerns from their peer stakeholders. In this paper, we present concerns of the Texas Water Development Board (TWDB) by mining tweets from their official Twitter account in 2017, as shown in Figure 6. We can see TWDB was concerned more about *drought, conservation* and *groundwater* in the first quarter of the year. Later, *waste water, water supply, water storage, windmill* and *grant* were mostly discussed. In the fall of 2017, they were concerned about *rainfall, river, Harvey recovery* and *financial*, as Hurricane Harvey dumped trillion gallons of water into Texas during this period. Throughout the year, *improvement* is always the repeated concern to sustain overall water related conservation solutions.

### C. Knowledge Sharing

This feature is designed to connect water stakeholders (farmers, water researchers, water technicians) to discuss questions or share their best practices, or propose technical solutions to certain problems. There are four categories which are *water sources, water processing, water usage* and *water treatment*. Each topic has a set of articles shared by users automatically classified into subtopics with a rating mechanism for quality control. Figure 7(b) demonstrates the articles, rating, and classification of subtopics of *water sources* including *ground water, reservoir* and natural water containers such as *lake, river* and *pond*. The WaterScope helps create an interconnected virtual community that is working towards the common goal of improving water supply resilience in otherwise disconnected groundwater-dependent rural economies.

## V. Summary

We build WaterScope, an information-sharing platform, and demonstrate its novel application of state of the art data mining techniques as the solution to water awareness and management. The platform incorporates a variety of dynamic and heterogeneous water-related datasets, and provides insight for water level forecasting, water concern flow, individual well distributions, and county average water level trends. It is not only a smart reference tool for water managers but also a place where water stakeholders are connected and can share best practices. Even though the platform is built for water resource management, its conceptual solution can be applied in other domains that helps bring together different stakeholders and promote community engagement through multiple data sources, especially social media.

## VI. Acknowledgement

This work was supported by the U.S. National Science Foundation under Grant CNS-1737634.

## References

[1] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, “The stanford corenlp natural language processing toolkit,” in *Proc. ACL’14.*, 2014, pp. 55–60.
[2] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.
[3] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with lstm,” 1999.
[4] A. Graves, “Generating sequences with recurrent neural networks,” *arXiv preprint arXiv:1308.0850*, 2013.
[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.