APPLICATION OF MACHINE LEARNING TO UNCOVERING LINKS BETWEEN DISASTER ALERTS

Mei-Chun Kuo[1,2] and Walter Chen[2]
[1]Information Division, National Science and Technology Center for Disaster Reduction
No. 200, Sec 3, Beisin Rd., Xindian District, New Taipei City 23143, Taiwan
[2]Dept. of Civil Engineering, National Taipei University of Technology
1 Sec. 3 Chung-Hsiao E. Rd., Taipei 10608, Taiwan
Email: bettyk@ncdr.nat.gov.tw; waltchen@ntut.edu.tw

KEYWORDS: Taiwan profile, Common Alerting Protocol, machine learning, association rule, conditional probability, disaster alert

ABSTRACT: Since 2013, the National Science and Technology Center for Disaster Reduction (NCDR) and the rest of the ministries, departments, and agencies of the central government have jointly formulated the Taiwan Profile of the Common Alerting Protocol (CAP-TWP) for use in Taiwan. Nowadays, CAP-TWP has become the standard with the most number of unified data formats in Taiwan. The NCDR built the Alert Public Information Platform (NCDR-APIP) to provide real-time disaster alert data of four primary categories: meteorology, hydrology, transportation, and civic-related information. These data were issued by the central ministries and state-owned enterprises. This research explored the conditional probability of flood alerts given rainfall alerts. Tainan City was selected as an example to mine association rules using rainfall and flood alerts as the antecedent itemset. It is found that the most likely consequent alert type is the reservoir discharge alert, no matter what the alert level is.

1. INTRODUCTION

Taiwan is located in the prevalent path of subtropical typhoons in the western Pacific of the northern hemisphere. It is visited by typhoons, southwesterly flows, plum rains, northeasterly flows, summer heat convection, and torrential rains each year, creating distinctive climatic characteristics in cities of Taiwan year-round. Due to Taiwan's small land area, steep rivers and rapids, the natural conditions are well suited for multiple disasters to occur concurrently and little time for disaster alerts in advance. Currently, alerts for disasters mostly come from devices using the SensorThings API. These devices are usually capable of gathering only a single type of data. There are concerns about whether these devices are suitable and adequately placed at the desired locations.

Because of Taiwan's unique problem and its high occurrence probability of disasters, the present research opts to use the internationally well-perceived Common Alerting Protocol (CAP) instead. CAP is a simple and XML-based standard data format promoted by the Organization for the Advancement of Structured Information Standards (OASIS). It is a standard format for publishing and exchanging various emergency alerts among disaster response systems. Since 2013, the National Science and Technology Center for Disaster Reduction (NCDR) and the rest of the ministries, departments, and central government agencies have jointly formulated the Taiwan Profile of the Common Alerting Protocol (CAP-TWP) for use in Taiwan. The NCDR built the Alert Public Information Platform (NCDR-APIP) to provide real-time disaster alerts of four primary categories: meteorology, hydrology, transportation, and civic-related information. The NCDR-APIP has gathered 47 types of data issued by 31 central ministries and agencies as well as state-owned enterprises. The historical data of 23 past major disasters in Taiwan from 2013 to 2019 have also been collected. Jointly, these data can be used to analyze the occurrence probability of multiple disasters by uncovering the links between different types of disaster alerts. This will improve the ability of local governments (such as cities and counties) to provide a first response tailored to the needs of local residents.

2. LITERATURE REVIEW

There is a growing body of research on the application of machine learning to the commercial sector, which is now shifted to the early warning and monitoring of disasters. Examples include the use of support vector machine to assess the risk of debris flow (Yuan and Zhang, 2006) and the use of random forest or artificial neural network to forecast regional floods (Chen et al., 2020; Puttinaovarat and Horkaew, 2020). The issuing of disaster alerts is subjected to the discretion of the issuing agencies. Therefore, we assume all alerts are useful and justified in this study. The goal is to investigate whether disaster alerts are correlated using machine learning. In particular, association rule learning is used to mine rules among multiple disaster alerts. The results will be displayed in a graphical form for easy understanding and interpretation.

3. MATERIAL AND METHOD

The present study employed the CAP-TWP data collected over the period from 2013 to 2019 from the NCDR-APIP. However, close inspection of the data indicates that the data in the first two years were partial and mostly test alerts. Therefore, they were excluded from the following analysis. The remaining data used are in the period from January 2015 to October 2019. The entire dataset contains 140 thousand records spanning 20 categories, which include rainfall, reservoir discharge, strong wind, flood, road closure, high river water level, typhoon, air quality, low temperature, debris flow, earthquake, workplace and school closure, thunderstorm, high-speed rail accident, railway accident, dense fog, mobile phone interruption, power outage, high temperature, and landline interruption alerts.

Further analysis of the data reveals that there were rainfall alerts issued every month in Taiwan. In fact, the rainfall alert is the most common type of alert among the 20 categories of alert data (Figure 1), accounting for 36% of the total number of alerts. We thus chose the rainfall alerts along with the significantly correlated flood alerts to explore the relationship between floods alerts and different levels of rainfall alerts. We first identified the highly probable flood risk areas using conditional probability (equation 1) and then estimated the likelihood of floods in cities/counties under moderate, severe, and extremely severe rainfall alerts.

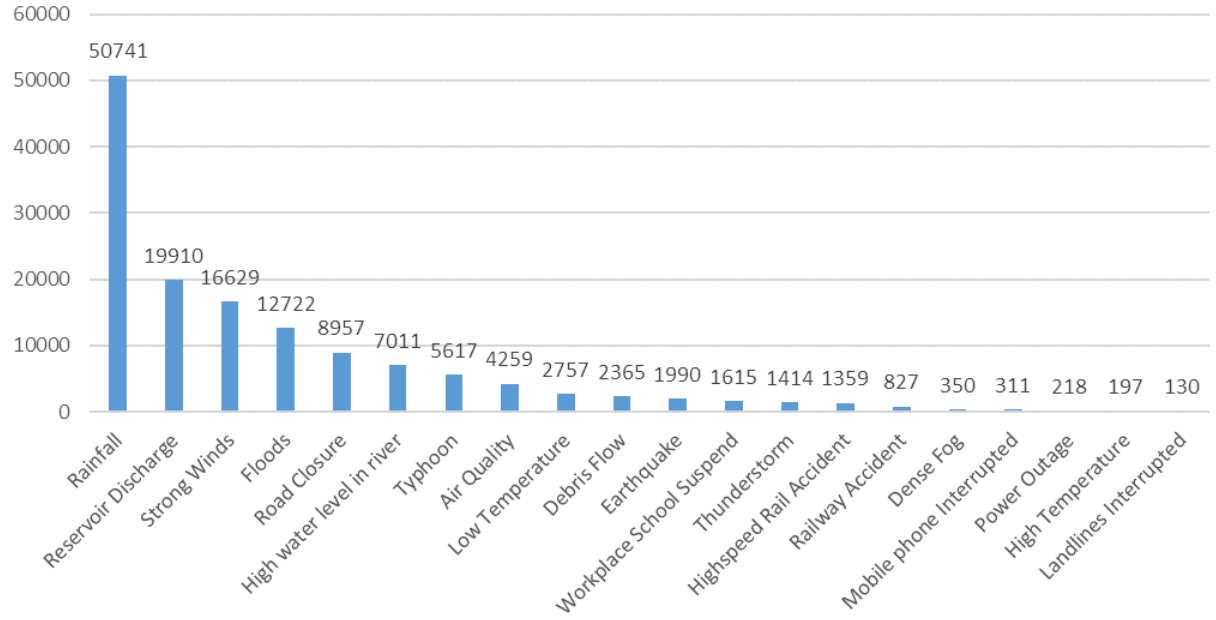

Figure 1 Distribution of the number of disaster alerts among 20 categories

The conditional probability equation for calculating the probability of flood alerts given rainfall alerts is as follows:

$$P(F|R) = \frac{P(R \cap F)}{P(R)} \quad (1)$$

where

$P(F|R)$ = probability of issuing flood alerts given rainfall alerts
$P(R \cap F)$ = probability of concurrent rainfall and flood alerts
$P(R)$ = probability of issuing rainfall alerts

For implementation, this research used the apriori() command in the R language to mine association rules. There are three parameters associated with each discovered rule, namely support, confidence, and lift. Graphically, the confidence is expressed by the size of a red circle, whereas the darkness of the red color represents the lift. Rainfall alerts are categorized into three levels--moderate, severe, and extremely severe. We selected the association rules with a minimum support of 0.01 and sorted the significant rules (lift > 1) by support. Both the rainfall and flood alerts were used as the antecedent itemset for association rule mining to determine the consequent itemset. Specifically, support represents the probability of issuing both rainfall and flood alerts at the same time. Confidence signifies the likelihood of issuing flood alerts given rainfall alerts. Finally, lift is the ratio of the probability of issuing both alerts

to the probabilities of issuing the two alerts individually. If the lift is greater than 1, then the resulting rule will be better than a random guess. The mathematical representations of support, confidence, and lift are shown in equations (2) to (4).

$$Support(A \rightarrow B) = P(A \cap B) \tag{2}$$

$$\text{Confidence}(\text{A} \rightarrow \text{B}) = \text{P}(\text{B}|\text{A}) \tag{3}$$

$$\text{lift}(\text{A} \rightarrow \text{B}) = \frac{P(A \cap B)}{P(A)P(B)} \tag{4}$$

where

A = antecedent itemset
B = consequent itemset

4. RESULTS

The results of the conditional probability calculation are shown in Table 1. The table shows that under moderate rainfall alerts, the top three places to anticipate flood alerts are Changhua County, Kaohsiung City, and Tainan City in the order of probability. If severe rainfall alerts are issued, then the top three places that will anticipate flood alerts are Pingtung County, Tainan City, and New Taipei City. Finally, if extremely severe rainfall alerts are announced, then Pingtung County, Tainan City, and Kaohsiung City will be the top three places to receive flood alerts. Overall speaking, Tainan City is the only place that appears in all three rankings, no matter what the alert level is. This shows that flood alerts should be of particular concern to Tainan City. Hence, Tainan City is chosen as the study area for the rest of this study.

Table 1 Probability of flood alerts in cities/counties under different levels of rainfall alerts

City/County	Flood_Condition_Probability_When_ModerateRainAlert	Flood_Condition_Probability_When_SevereRainAlert	Flood_Condition_Probability_When_ExtremeRainAlert
Yilan County	3.23%	15.65%	27.03%
Hualien County	2.44%	6.33%	26.92%
Kinmen County	0.00%	0.00%	0.00%
Nantou County	3.64%	19.05%	20.00%
Pingtung County	8.99%	28.19%	45.83%
Miaoli County	2.78%	5.08%	18.18%
Taoyuan City	3.92%	8.47%	8.33%
Kaohsiung City	10.40%	22.82%	38.46%
Keelung City	0.69%	7.23%	16.13%
Lienchiang County	0.00%	0.00%	0.00%
Yunlin County	2.34%	6.62%	33.33%
New TaipeiCity	5.18%	24.32%	31.43%
Hsinchu City	0.00%	6.25%	0.00%
Hsinchu County	2.78%	5.13%	12.50%
Chiayi City	1.47%	4.11%	8.33%
Chiayi County	4.71%	12.58%	18.18%
Changhua County	22.86%	14.58%	27.27%
Taichung City	4.13%	8.53%	27.27%
Taipei City	3.81%	10.99%	17.86%
Taitung County	3.13%	9.46%	36.00%
Tainan City	10.29%	24.62%	43.75%
Penghu County	0.00%	11.11%	11.11%

Using rainfall and flood alerts as the antecedent itemset for association rule mining in Tainan City, it is found that under moderate rainfall alerts, the most likely consequent itemset is a reservoir discharge alert, as shown in Table 2. The support, confidence, and lift of the rule is 6.36%, 78.57%, and 1.71%, respectively. The next likely consequent itemset is a thunderstorm alert. This rule has lower support and confidence, but a much higher lift than the first rule. The two rules are illustrated in Figure 2. The size of the red circle represents the confidence, and the darkness of the red color denotes the lift. It is evident that a reservoir discharge alert is more likely to be issued than a thunderstorm alert. However, the local government needs to take precautions to ensure that emergency preparedness is up-to-date, and the first response is fast and effective.

Table 2 Association rules found under moderate rainfall alerts in Tainan City

rules	support	confidence	lift
{Rainfall,Floods} => {Reservoir.Discharge}	6.36%	78.57%	1.710
{Rainfall,Floods} => {Thunderstorm}	2.31%	28.57%	4.707

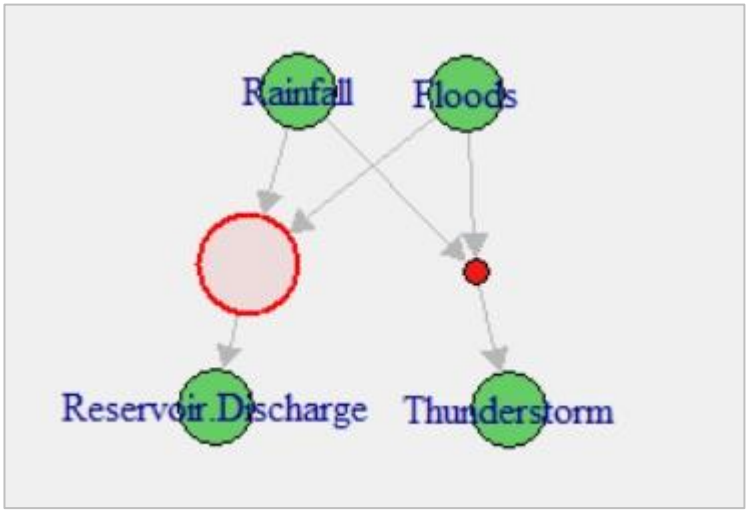

Figure 2 Illustration of rule confidence and lift under moderate rainfall alerts in Tainan City

Also, if Tainan City is under severe rainfall alerts, then the association rule mining shows that the most likely consequent itemset is also a reservoir discharge alert (Table 3). The support, confidence, and lift of the rule is 15.51%, 67.44%, and 1.47%, respectively. The next likely consequent itemset is a high river water level alert. This rule has lower support and confidence, but a slightly higher lift than the first rule. The two rules are illustrated in Figure 3. The size of the red circle represents the confidence, and the darkness of the red color denotes the lift. It is evident that a reservoir discharge alert is more likely to be issued than a high river water level alert in Tainan City.

Table 3 Association rules found under severe rainfall alerts in Tainan City

rules	support	confidence	lift
{Rainfall,Floods} => {Reservoir.Discharge}	15.51%	67.44%	1.466
{Rainfall,Floods} => {High.water.level.in.river}	2.14%	9.30%	1.581

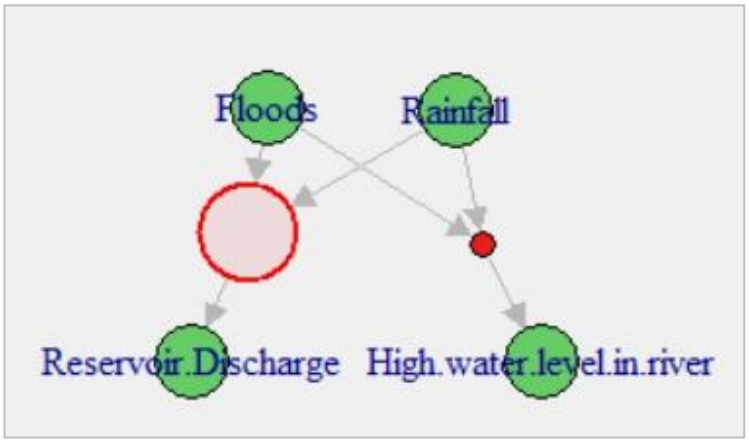

Figure 3 Illustration of rule confidence and lift under severe rainfall alerts in Tainan City

Finally, under extremely severe rainfall alerts, the association rule mining uncovers six rules when the antecedent itemset is composed of rainfall and flood alerts. The most likely consequent itemset is still a reservoir discharge alert (Table 4). The support, confidence, and lift of the rule is 28.26%, 86.67%, and 1.08%, respectively. The next likely consequent itemset is a workplace and school closure alert. The support, confidence, and lift of the rule is 17.39%, 53.33%, and 2.45%, respectively. The support and confidence are lower than the first rule, but the lift is significantly higher. The remaining four consequent itemsets are a typhoon alert, a high river water alert, a debris flow alert, and a power outage alert. Coincidentally, they all have higher values of lift than the first rule. All six rules are visualized in Figure 4. Again, the size of the red circle and the darkness of the red color represent the confidence and the lift, respectively. It is practically impossible to overemphasize the importance of proper preparedness by having properly trained emergency personnel on staff and emergency procedures in place to meet the needs of citizens in the city.

Table 4 Association rules found under extremely severe rainfall alerts in Tainan City

rules	support	confidence	lift
{Rainfall,Floods} => {Reservoir.Discharge}	28.26%	86.67%	1.077
{Rainfall,Floods} => {Workplace.School.Suspend}	17.39%	53.33%	2.453
{Rainfall,Floods} => {Typhoon}	17.39%	53.33%	2.044
{Rainfall,Floods} => {High.water.level.in.river}	15.22%	46.67%	1.789
{Rainfall,Floods} => {Debris.Flow}	8.70%	26.67%	2.453
{Rainfall,Floods} => {Power.Outage}	4.35%	13.33%	1.533

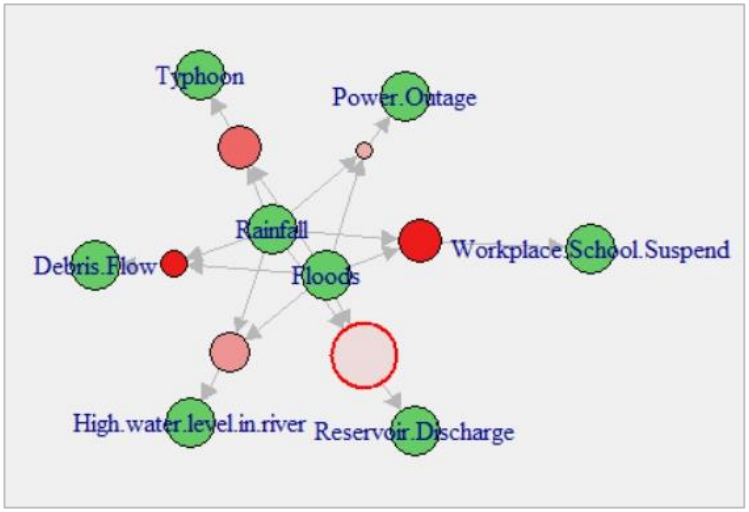

Figure 4 Illustration of rule confidence and lift under extremely severe rainfall alerts in Tainan City

5. DISCUSSION AND CONCLUSIONS

This study used disaster alert data from the NCDR-APIP collected from January 2015 to October 2019 to study the association between disaster alerts. It is found that Tainan City consistently ranks high to receive flood alerts after a rainfall alert. Furthermore, using association rule mining, it is determined that the most likely consequent alert in Tainan City is a reservoir discharge alert given both rainfall and flood alerts, no matter what the alert level is. As the alert level increases from moderate to extremely severe, the number of consequent alerts also increases from two to six. Therefore, it is advised that attention should be paid to the preparedness for possible emergency events and effective response. Critical infrastructures such as hospitals should be ready to provide emergency care in the event of a disaster. The local government needs to take precautions to ensure that emergency personnel are on staff and emergency procedures are in place to meet the needs of citizens.

Acknowledgments

The assistance of Sheng-Xiang Fan in the preparation of this manuscript is gratefully acknowledged.

References

Chen, J., Li, Q., Wang, H., Deng, M., 2020. A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the Yangtze river delta, China. International Journal of Environmental Research and Public Health, 17 (1), 49.

Puttinaovarat, S., Horkaew, P., 2020. Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques. IEEE Access, 8, 8949502, pp. 5885-5905.

Yuan, L., Zhang, Y., 2006. Debris flow hazard assessment based on support vector machine. Wuhan University Journal of Natural Sciences, 11 (4), pp. 897-900.