The Integration of WEKA and Google Maps in Visualization with a Web Mash-up of Deluge Risks Tracking in E-Tracking System

Kitti Koonsanit[1,2] and Chuleerat Jaruskulchai[1]
[1]*Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand*
[2]*National Electronics and Computer Technology Center, Patumthani, Thailand*
E-mail: g5317400074@ku.ac.th, fscichj@ku.ac.th

Abstract

In the end of the day, meteorological data becomes widely used such as e-monitoring and e-warning. It is used to predict the state of the environmental for a future time. This paper presents the integration of WEKA libraries (Waikato Environment for knowledge analysis) and Google maps API for classification and visualization in e-tracking, which is an environmental tracking. With this software developed as a meteorological data analysis tool, it can analyze resulting and visualization web applications easier and faster. Our software is a front end of WEKA that provides fundamental data mining functions such as classify, clustering, and analysis functions. It also supports pre-processing, analysis, and XML output with exporting result. After that, our software can export and display data result to Google maps API in order to display result and plot labels and visualization in different color labels on Google maps. We created our software using java ,developed on Netbeans IDE version 6.8 and tested the proposed process on meteorological data in Thailand, and the experimental results show that our proposed process can export data result to Google maps and display classification results in order to help visualize its nearly real-time and effectively.

1. Introduction

WEKA and Google maps API have become popular libraries in data mining and visualization on a web, respectively, because WEKA is freely available and contains a large variety of algorithms. The main features of WEKA involve data mining, while the main features of Google Maps API aim toward visualization. Therefore, their integration can make the software more powerful. In this paper, we present a technique for integration WEKA and Google maps API so that advanced data processing techniques in WEKA can be visualized by Google maps API.

2. Related Work

A number of papers and articles have been published regarding the subjects such as e-monitoring and e-warning. One of the first one, introduced by J.G. Goldammer[1] proposed an early warning system for the prediction of an appropriate response to wildfires and related environmental hazards. And many publications have emerged in areas of early warning systems [2][3]. Finally, following the development of visualization as web mash up, Josh Lowensohn[4] proposed a combination of maps and spreadsheets with Google's new mash up tool. However, none of the approaches can be employed WEKA libraries to incorporate with web mash up in visualization. Besides, all the methods, these methods are unsuitable and usable for meteorological data in Thailand that needs to implement software fast and to ease the difficulty of implement software for basic users in the developing country. At the same time, natural hazard should be predicted, controlled and prevented. Therefore, this paper focuses on the development and applications of the e-tracking system for environmental hazards, which was used as one of the essential tools for tracking in Thailand. The environmental hazard tracking was developed based on the environmental characteristics of Thailand.

3. WEKA

WEKA[5] is a widely used toolkit for machine learning and data mining originally developed at the University of Waikato in New Zealand. It is a large collection of state-of-the-art machine learning algorithms written in Java. WEKA contains tools for classification, regression, clustering, association rules, visualization, and data pre-processing. WEKA is open source software under the GNU GPL. It is easily extensible, which allows researchers to contribute new learning algorithms to WEKA, keeping it up-to-date with the latest developments in the field. As a result, WEKA has become very popular with academic and industrial researchers, and is also widely used for teaching purposes. The main focus of WEKA is on classification algorithms. Each data instance is described by its attribute values. For example, clustering which are several different areas based on observations of sky, air temperature, humidity, and wind can be viewed as a clustering task. Each data instance includes values of the observation attributes, such as sunny, warm, humid, strong, and the available classes are {cluster1, cluster2, cluster3, cluster4}. The idea is that a clustering learned on an unlabeled data set can then be used to cluster around unlabeled class. This reason above that we choose and use WEKA to help analysis meteorological data.

4. Google map API

Google maps API[6] is a web mapping service application and technology provided by Google, free (for non-commercial use), that powers many map-based services, including the Google maps website, and maps embedded on third-party websites via the Google maps API. It offers street maps, a route planner for traveling by foot, car, or public transport and an urban business locator for numerous countries around the world. According to one of its creators (Lars Rasmussen), Google map is "a way of organizing the world's information geographically".

As Google maps is coded almost entirely in JavaScript and XML, some end users have reverse-engineered the tool and produced client-side scripts and server-side hooks, which allowed a user or website to introduce expanded or customized features into the Google maps interface.

Google maps API allowed developers to integrate Google maps into their websites as shown in Fig. 1. It is a free service.

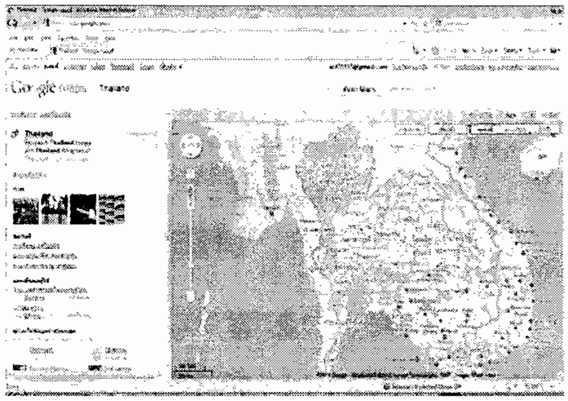

Figure 1. Google map

5. Mash up

A web mash-up[7] is a web application that takes information from one or more others sources and presents it in a new way or with a unique layout that combines data, presentation or functionality from two or more others sources to create new services. The mash-up is easy to implement, fast integration, frequently using an open APIs and data sources to produce enriched results. These features provide a unique way of web interface design. This reason above that we choose and use a web mash-up to visualize classification results nearly real-time and effectively.

6. Design

In the GIS (Geographic Information System) world are graphical representations of point data on a map through the use of colors and icons that indicate the representation of some variable such as the flood, waterless area, or cultivated area. Our software can quickly visualize the representation of locations. Being able to understand the label of point locations makes it much easier to see patterns in result data, especially when using different colors. In this session, we present to show processes for creating maps in our software.

Our software is divided into three distinct functions, including a function for importing meteorological raw dataset, functions for classification dataset using WEKA library, which is a backend, a function for exporting the XML file, an initialization function for creating the Google map from the XML file, and finally a function to plot the map using the store points read from the XML file on top of the Google Map which is a visualization representation as shown in Fig. 2.

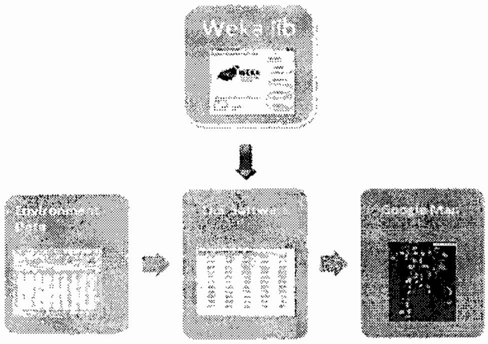

Figure 2. Overall of System

7. Methods

7.1 Dataset

We used the one hundred and eight sets of meteorological raw data from different stations [8]. The dataset are obtained from Thai Meteorological Department (TMD), a member of Ministry of Information and Communication Technology of Thailand[9]. We would like to analyze data, which was registered in Thailand and thus try to forecast the future rainfall forecasting and deluge risk areas in Thailand.

In this paper, the data used to create our attributes include: day, average temperature, average pressure, relative humidity, and accumulated precipitations. The next step of our method focuses on transforming these data in order to be used in Weka.

7.2 Preprocessing

According to our experience, we used evaluate on training data test mode for J48 classifier algorithm, that can only work with nominal variables, so we have to adapt the data to Weka format by a discretize filter, which is an instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. After choosing a discretize filter (weka.filters.unsupervised.attribute.Discretize), it can be applied to the initial dataset. The result of this transformation is shown in Table 1-5.

Table 1. Temperature intervals

Temperature intervals	Represented by
-inf-23	1
24-29	2
30-33	3
34-38	4
39-inf	5

The average temperature values will be divided in equal intervals, each ones having attached a number: 1 – (-inf, 23), 2 – (24, 29), 3 – (30, 33), 4 – (34, 38) and 5 – (39, inf).

Table 2. Pressure intervals

Pressure intervals	Represented by
-inf -1004	1
1005-1008	2
1009-1015	3
1016- 1019	4
1020-1027	5
1028- inf	6

The average pressure values will be divided in equal intervals, each ones having attached a number: 1 – (-inf, 1004), 2 – (1005, 1008), 3 – (1009, 1015), 4 – (1016, 1019), 5 – (1020, 1027) and 6 – (1028, inf).

Table 3. Relative humidity intervals

Relative humidity intervals	Represented by
-inf-64	1
65- 68	2
69-74	3
75- 80	4
81- 83	5
84-88	6
89-inf	7

The same codification is used for the relative humidity values: 1 – (-inf, 64), 2 – (65, 68), 3 – (69, 74), 4 – (75, 80), 5 – (81, 83), 6 – (84, 88) and 7 – (89, inf).

Table 4. Accumulated precipitation intervals

Accumulated precipitation intervals	Represented by
-inf-88	1
89-185	2
186-279	3
280-393	4
394-531	5
532-657	6
658-inf	7

The predicted variable, accumulated precipitation is measured in mm. and used 1 for the (-inf, 88), 2 – (89, 185), 3 – (186, 279), 4 – (280, 393), 5 – (394, 531), 6 – (532, 657) and 7 – (658, inf).

Table 5. Example Dataset

Station	LO	LA	AT	AP	RH	ACP	DA
1 (Bangkok)	13.45N	100.35E	2	1	1	1	Flase
2 (ChiangMai)	18.47N	98.59E	3	1	2	1	Flase
3 (Chiang Rai)	19.52N	99.50E	2	1	1	1	Flase
4 (Rayong)	12.40N	101.20E	3	1	1	1	Flase
5 Lop Buri	14.48N	100.37E	3	1	2	1	Flase
6 (Kalasin)	16.26N	103.30E	3	1	1	1	Flase
7 (Loei)	17.29N	101.35E	1	1	2	2	True
8 (Sakaew)	16.26N	103.55F	3	1	1	1	Flase
.....							
......							

where,
LO is longitude,
LA is latitude,
AT is average temperature,
AP is average pressure,
RH is relative humidity,
ACP is accumulated precipitation
and DA= deluge areas. [10]

7.3 Algorithm

After the input dataset is transformed in the format that is suitable for the machine learning scheme, we use J48 to build a classification tree to predict the accumulated precipitation forecasting and deluge risk areas in Thailand because the J48 model format as tree or rules is easy to understand. To predict the state of deluge risk, we need to build decision tree based on the training data set. The results of decision tree as shown in the figure 3.

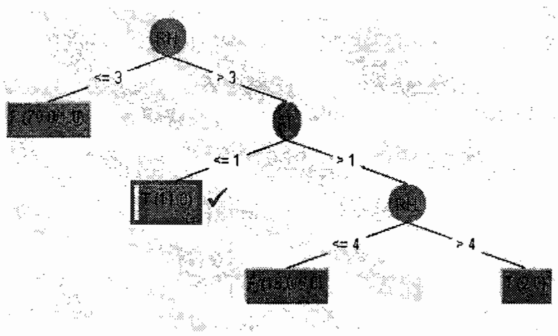

Figure 3. Example decision tree

This tree is interpreted using the If-Then rules:
If (RH > 3) and (AT <= 1) then deluge areas = True.

Furthermore, with the coding rules used for the database, these can be explained: If relative humidity is more 70-74 (3) and the average temperature is 0-25 (1) then deluge risk area is true (T). In this example, out of a total of 108 instances, only 96 have been correctly classified meaning 88.89 %.

7.4 Display

To display deluge risk on the Google maps, the result of deluge risk prediction is stored in XML form. We plot red markers for deluge risk areas. Our XML file, called Landmarks.xml, contains several landmarks in a format as seen in the figure 4.


```
<markers>

<!-- E-tracking : Loading the data from an XML file -->

<marker lat="43.65654" lng="-79.90138" label="Green"/>
<marker lat="42.31892" lng="-78.89231" label="Red"/>
<marker lat="42.42589" lng="-79.10040" label="Green "/>
<marker lat="43.44654" lng="-79.90138" label="Green"/>
<marker lat="43.78892" lng="-78.89231" label="Green "/>
<marker lat="43.11589" lng="-79.10040" label="Green "/>
<marker lat="42.39892" lng="-78.89231" label="Red"/>
<marker lat="41.44589" lng="-79.10040" label="Green "/>

......

</markers>
```


Figure 4. XML file contains several dozen landmarks

Finally, this file contains elements for latitude and longitude which we will use to plot the marker along with one attribute elements, including the clustering name such as green and red. We are going to do in this simple application is read this file, plot out a new instance of GMarker using the <Longitude> and <Latitude> elements for each site, and displayed in Google maps as markers.

8. Results

We have tried to highlight the way the stored data about past events can be used in the forecast of the future ones. If we use partial auto-correlation methods for each parameter to determine time dependencies we see that it is sufficient to consider the data on the current day and the day before in order to predict the state of risk areas for the next day.

8.1 Experimental Result

Original aspect rule from Thai Meteorological Department [11] is that if accumulated precipitation in a day is approximately more 90 mm, it was assumed a state of "deluge risk areas" and if otherwise, it was assumed a state of "normal areas".

Besides above aspect, our research can show a new rule as seen through the If-then rules: "If

(Relative Humidity > 3) and (Average Temperature <= 1) then deluge areas = True". This rule from our experiments can be used in prediction work. Thus, if we know the relative humidity and the average temperature value, using this decision tree, we can estimate the future probability for the deluge risk.

Finally, we are displaying and exporting an auto-updating XML file that is clustered around results containing 108 points provided by e-tracking. This e-tracking integrates WEKA and Google maps to deliver interesting area and locations to deluge risk areas around Thailand. It is built on a web mash up, a web mapping interface that displays interesting locations delivers near real-time information to publish users as seen in the figure 5.

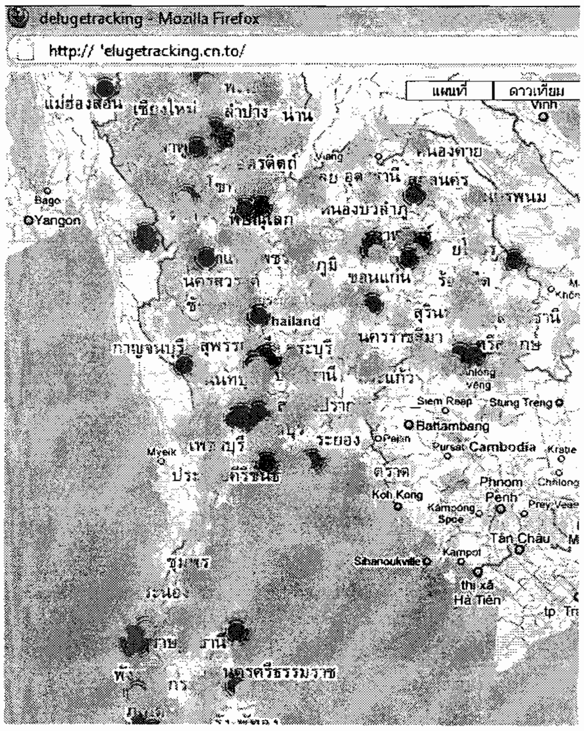

Figure 5. Result image (http://delugetracking.cn.to)

9. Conclusion

This paper presents the integration of WEKA libraries (Waikato Environment for Knowledge Analysis) and Google maps API for classification and visualization in e-tracking, which is an environmental tracking. With this software developed as a meteorological data analysis tool, it can analyze resulting and visualization web applications easier and faster. Our software is a front end of WEKA that provides fundamental data mining functions such as classify, clustering, and analysis functions. It also supports pre-processing, analysis, and XML output with exporting result. After that, our software can export and display data result to Google maps API in order to display result and plot labels and visualization in different color labels on Google maps. We created our software using java ,developed on Netbeans IDE version 6.8 and tested the proposed process on meteorological data in Thailand, and the experimental results show that our proposed process can export data result to Google maps and display clustering results in order to help visualize its nearly real-time and effectively. The outcome of this research will be used in further steps for visualization tools that help visualize meteorological data such as e-monitoring, e-tracking, natural resources, agricultural, environment and climate change.

A result of this research was developed to provide users process to visual classification result on Google maps. Many data mining techniques in WEKA libraries can be included within our software in order to make the software more useful and valuable. We hope that it can be used as visualization tool and help develop research in meteorological data mining software in the future.

10. Future work

Future work, we will include the extension of the database with other important weather parameters like wind speed, wind direction or radiation. Beside this aspect, we can enlarge our database with records from other years not only from this year. Having all this improvements in mind, we think that it can increase the precision in building the decision tree and the weather prediction based on it.

11. Acknowledgments

The authors would like to thank the Thai Meteorological Department (TMD), a member of Ministry of Information and Communication Technology of Thailand for providing dataset on meteorological and Thailand Graduate Institute of Science and Technology (TGIST), a member of National Science and Technology Development Agency (NSTDA) for their financial support. The scholar student ID is TG-22-11-53-005D and the grant number is TGIST 01-53-005.

12. References

[1] J.G. Goldammer, "Early Warning Systems for the Prediction of an Appropriate Response To Wildfires And Related Environmental Hazards," *Health Guidelines for Vegetation Fire Events*, Lima, Peru, 6-9 October 1998.

[2] Yi Zhou; Shirong Chen; Weiqi Zhou; Litao Wang; , "Early Warning and Monitoring System for Forest and Grassland Fires by remote sensing data," *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International* , vol.7, no., pp.4799-4802 vol.7, 20-24 Sept. 2004

[3] Geowebinfo, "*The World's Most Earthquake - Vulnerable Cities,*" Retrieved: Sep 2, 2010, Available to: http://geowebinfo.com/disasters/EQVulnCities.html

[4] Josh Lowensohn, "*Combine maps and spreadsheets with Google's new mashup tool,*" Retrieved: Sep 2, 2010, Available to: http://news.cnet.com /8301-17939_109-9704486-2.html

[5] Remco R. Bouckaert, "*WEKA Manual,*" Retrieved: Sep 2, 2010, WAIKATO, Available from University, pp.1-303, January 2010.

[6] "*Google Map API,*" Retrieved: Sep 2, 2010, Available to: http://code.Google.com/apis/maps/index.html

[7] LIN Hui and YE Lei, "A Mash-up of the Tracking System and Real-time Monitoring System Based on MAS," *IJCSNS International Journal of Computer Science and Network Security 2006*, vol.6 No.12, December 2006

[8] "*Meteorological raw data,*" Retrieved: Sep 2, 2010, Available to: http://www.tmd.go.th/climate/climate.php

[9] "*Image map 3 hrs of rainfall, temperature, temperature, relative humidity and MSL air pressure,*" Retrieved: Sep 2, 2010, Available from http://www.arcims.tmd.go.th/dailydata/Map3hrs.php

[10] "*Warning List Data,*" Retrieved: Sep 2, 2010, Available from http://www.tmd.go.th/list_warning.php

[11] "*Accumulated Precipitation Data,*" Retrieved: Sep 2, 2010, Available from http://hydromet.tmd.go.th/Monitor/Forecast.aspx