VOL. 11, NO. 21, NOVEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AN EFFICIENT WEATHER PREDICTION BASED ON LINKED DATA FOR SMART TOUR

N. Senthilselvan, V. Subramaniyaswamy and K. R. Sekar
School of Computing, SASTRA University, India
E-Mail: senthilselvan@cse.sastra.edu

ABSTRACT

In the past one decade linked open data plays an inevitable role in the research arena. Mobility plays a vital role with human fast communication. Location identifiers are an important factor that is incorporated in smart phones. Images and histories are also a part of the same. Now in this paper the proposed methodology is to include the weather condition in that particular location through the well renowned classification methodologies that produces lots of information to the customers for arranging their convenience. Metadata set has been taken from the portal www.indiawaterportal.org; and there it is possible to take 100 years average data for 29 districts in Tamilnadu. For these work 348 records taken and the sample dataset is depicted for example. Ontology, semantic and taxonomy were implied in the same dataset.

Keywords: mobility, classification, linked open data, ontology, semantic and taxonomy.

1. INTRODUCTION

Linked open data provides high value information to the customers in all digital gadgets. Now-a-days mobiles, tablets, and other smart devices are literally having such type of information like locations, directions and particular hot spots. But in the recent days knowing the weather condition in that particular location is an important factor that also a part of the existing issues. Semantic web provides a standard form of metadata model and reduces the complexity in the dataset with standard protocols. Acquiring information is the new philosophy or knowledge is possible through the above said. The “Linked Open Models" as a possible additional step, aims is to enable users to externalize knowledge in the form of diagrammatic models - a type of content that is human-readable, as well as linkable in the way promoted by the Linked Data paradigm [1].

Ontology reveals the interrelation between the columns in a trained dataset and their relationships. If all the metadata is available under the single source of location then it will have a fast retrieval of data for analysis purpose. Nowadays customers need is so high and sophisticated to do their activity. Merging the weather condition in the same will provide a long mileage to the gadgets. Distribution of data over the internet and networks yields a good semantic for scrutinizing the data and to bring-up the hidden features. In tourism open linked data provides numerous information the same evaluates processes of tourism-linked empowerment in four communities outside Cuzco Linkages are evaluated between community-based tourism and empowerment in Peru [2]. So data distribution, semantic, ontology and taxonomy towards the offspring contribute more information to the society for growth in the coming era.

It is possible to get data from different sources that also give a good semantic model for the research. Analyzing the potential of mapping Schema.org terms and the Web of Linked Data Vocabularies collection is one good example for LOD [3]. In this paper the ideology behind the images, locations, road-map, history and weather details are the important factors that all have to be resolved for communication gadgets. Many techniques were available for finding the required semantic with ontology. Different classifications like Bayes, Naïve Bayes, Decision tree, R-part, Prism. etc. are all in the survey to pursue the good precision for getting an accurate answer about the weather condition in the location of that day. This contribution gives a fuel to the mobile vehicle to travel a long while.

The other vital information is discussed in the following sections. Related Works, Proposed Methodology, Results and Discussion and Conclusion.

2. RELATED WORKS

To write this research article some of the conceptual dogma induces in producing good information towards linked open data. The real ideologies are cultural heritage knowledge, personalized search for the info ranking, diverse discipline info, to find the potential about the data, semantic relation between the data, resource description framework and rich content should be formulated with a respective factor affordable to intricate knowledge for the patron. The following research articles adhered in this segment.

A System of cultural heritage knowledge is built from collaboration between open data published by a regional government, enriched with data from other sources of open linked data cloud like DBPedia. Mobile devices are used as e-tourism tools [8]. Personalized concept-based search mechanism for the web of data on results categorization. Results with same concepts are grouped together to form CONCEPTLENS. Within the selected CONCEPTLENS, more relevant results are included using results re-ranking and query expansion [9].The above mentioned work argues that deploying theories from diverse disciplines, and considering using different inquiry systems and research cycles, offers a more disciplined and robust methodological approach allowing to break through the limits of backward induction from the evidence by moving back and forward in exploring the unknown through BOLD [10].

VOL. 11, NO. 21, NOVEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

A comprehensive overview of existing data mining techniques and related tools are used to illustrate the potential of data mining for different research areas by means of example applications. The linked open data (LOD) approach will be presented as a new possibility in support of complex and inter-disciplinary data mining analysis [11]. The vision of the Semantic Web alleviates the difficulties of finding and retrieving relevant information and tools as well as integrating information from different sources [12]. For applying Linked data technologies to a portion of data found in ERMIS Greek portal for the public focuses on enrichment of the e-GIF ontology, modeling the concepts and relations which are used to organize the information appearing in the ERMIS Greek portal for Public Administration [13].

An open shared RDF-based tour vocabulary is needed to address the problems of using closed format for data repositories, and introduces such a model, Tour RDF and extending the earlier TourML schema into the era of Linked Data[14]. Mobile augmented reality applications for tourism will provide personalized and rich content for the augmented reality applications; Semantic web and Linked open data are used in the process [15].

3. PROPOSED METHODOLOGY

Applying different types of classifications were profound to check the sample data in the dataset model. Different classifiers are entitled here for reference namely Naïve bayes classification, Support Vector Machine, Recursive partitioning and Regression trees. After a thorough test is made with the same sample pattern all the classifiers provide the same result with more or less precision. Every time the samples are tested with all the three classifiers and would like to get the high precision result and infers the weather condition on a particular day and location. The same will be exhibited as a result and incorporated in mobility devices that give a long mileage for the customers to make their trip with a convenience. All the three results have been obtained through the machine learning process because of the huge trained dataset model. In the proposed work R-Tool was deployed to find the inference for the sample data.

In Figure-1 Smart Tour Proposed Architecture is organized in a chronological manner and every component is placed at a right position to get an inference. The components like DBpedia, meant for monumental general information, Monuments.xml expresses naming and coordinates, Flickr provides photos about monuments, MAPS and Routes give the path, Weather exhibits the forecast, Linked Open Data which integrates all the intricate components.

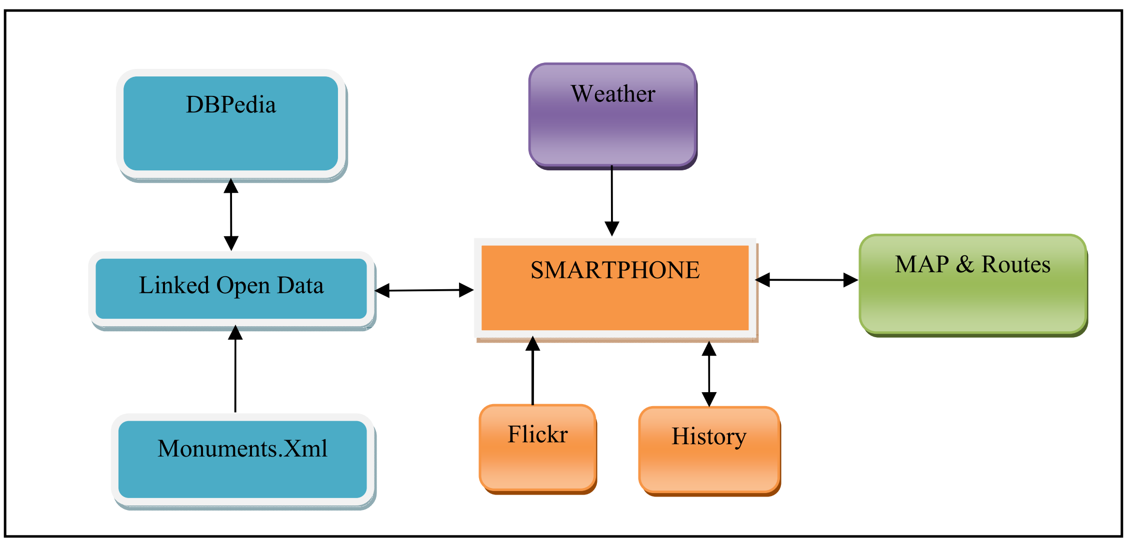

Figure-1. Smart tour proposed architecture.

Usually IT resources are available in different locations. For easy identification the graphs always gives a clear cut idea for the developers. The same thought is projected in paper to build a relatedness graph among resources in the DBpedia dataset that refer to the IT domain. Our final aim is to create a useful data structure at the basis of an expert system that looks for an IT resource, returns a ranked list of related technologies, languages, tools the user might be interested in [4]. Dictionary formation for vocabulary mapping provides huge information about a particular jargon available in different locations. The potential use of such vocabulary mappings to assist cross search over archaeological datasets from different countries is illustrated in a pilot experiment and the results demonstrate the enhanced opportunities for interoperability and cross searching that the approach offers [5]. Persons from different lands with their linguistic knowledge have to be formulated as a table with semantic offers information to a lot. Addressing the problem of harmonizing different stakeholder's languages and concepts presenting a knowledge base statistical information system and a solution to publish the semantic information on linked open data in the perspective of a shared information system acknowledged by all system actors is the objective [6]. E-based participation knowledge offers a good piece of information and is likely to have easy accessibility for the users. The role of e-participation, mapping practices and crowd source rating are much appreciated in the research area [7].

The Training set has twenty nine districts with seven attributes taken into account for evaluation. The training set literally focuses on hundred years data (1901-2001) and took 348 records for twelve months. Large sampling training set has been taken for work of finding the weather condition in a particular location. The adhered Table-1 exhibits only 10 tuples for example. The remaining records are available in the online portal www.indianwaterportal.org.

In the implementation segment methodologies namely Naïve Bayes, Support Vector Machine and Recursive Partitioning, Regression Tree classifications are presented and mined with the knowledge through the above said. Using R-Tool all the 348 records is evaluated and all the methods applied produce the same result with a less precision gap between them.

Table-1. Sample training dataset.

Districts	**Month**	**Precipitation** (mm)	**Min Temp** (c)	**Max Temp** (c)	**Cloud cover** (%)	**Vapour pressure** (hpa)	**Wet day frequency** (days)
Chennai	Aug	140.44	25.08	34.5	67.62	28.21	6.48
Coimabtore	Jan	14.64	20.06	30.19	31.01	21.85	1.23
Cuddalore	Oct	196.47	23.77	31.49	61.01	28.45	7.96
Dindigul	Jun	215.05	21.37	28.62	74.99	25.91	9.58
Madurai	May	119.49	23.7	31.33	58.15	28.5	5.65
Nagapattinam	Dec	154.81	22.49	28.55	44.26	25.94	5.57
Salem	Mar	11.23	22.71	34.82	26.98	23.87	0.91
Theni	Sep	225.91	22.17	28.87	64.4	26.79	10.22
Tirunelveli	Nov	151.91	18.06	23.81	47.25	21.93	7.39
Virudhunagar	Jul	138.3	22.61	28.97	77.6	27.74	8.29

3.1 Naive Bayes classification

Through the confusion matrix given below it is possible to get inference in a more precise manner. Here any given pattern has been segmented with some numeric value for cloudy, drizzle, heavy showers, partly cloudy, passing showers, rainy and sunny, which is clearly shown in Table-2.

Table-2. Confusion matrix.

	Cloudy	**Drizzle**	**Heavy showers**	**Partly cloudy**	**Passing showers**	**Rainy**	**Sunny**
Cloudy	15	10	00	08	02	00	04
Drizzle	07	47	00	00	09	04	00
Heavyshowers	00	00	28	00	00	05	00
Partlycloudy	11	00	00	12	00	00	05
Passingshowers	00	10	02	00	43	08	00
Rainy	00	05	03	00	12	26	00
Sunny	00	00	00	08	00	00	63

Test pattern

Precipitation	**Min temperature**	**Max temperature**	**Cloud cover**	**Vapour pressure**	**Wet day frequency**	**Class**
56.45	26.18	35.45	51.51	30.85	3.02	?

Result for the test pattern: Passing Showers

ARPN Journal of Engineering and Applied Sciences

3.2 Support vector machines

Test pattern

Precipitation	Min temperature	Max temperature	Cloud cover	Vapour pressure	Wet day frequency	Class
56.45	26.18	35.45	51.51	30.85	3.02	?

Parameters: SVM-Type: C-classification, SVM-Kernel: radial, cost: 1, gamma: 0.1666667
Number of Support Vectors: 267
Result for the test pattern: Passing Showers

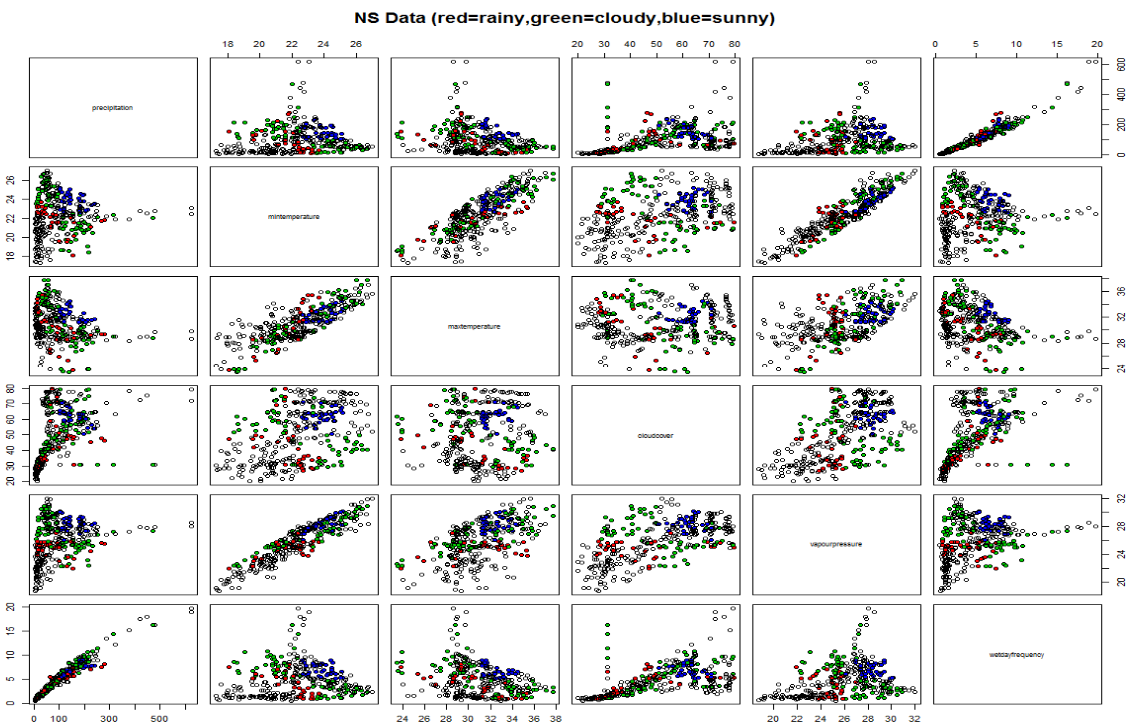

Figure-2. Scatter plot matrix.

3.3 Recursive partitioning and regression trees

Test pattern

Precipitation	Min temperature	Max temperature	Cloud cover	Vapour pressure	Wetday frequency	Class
56.45	26.18	35.45	51.51	30.85	3.02	?

Result for the test pattern: Passing Showers
Root node error: 275/347 = 0.79251
n= 347

	CP	nsplit	rel error	xerror	xstd
1	0.221818	0	1.00000	1.03636	0.025949
2	0.090909	1	0.77818	0.80364	0.032575
3	0.083636	3	0.59636	0.71636	0.033557
4	0.060000	4	0.51273	0.61818	0.033862
5	0.054545	6	0.39273	0.53818	0.033501
6	0.050909	7	0.33818	0.46545	0.032684
7	0.043636	8	0.28727	0.41455	0.031815
8	0.032727	9	0.24364	0.33455	0.029900

VOL. 11, NO. 21, NOVEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

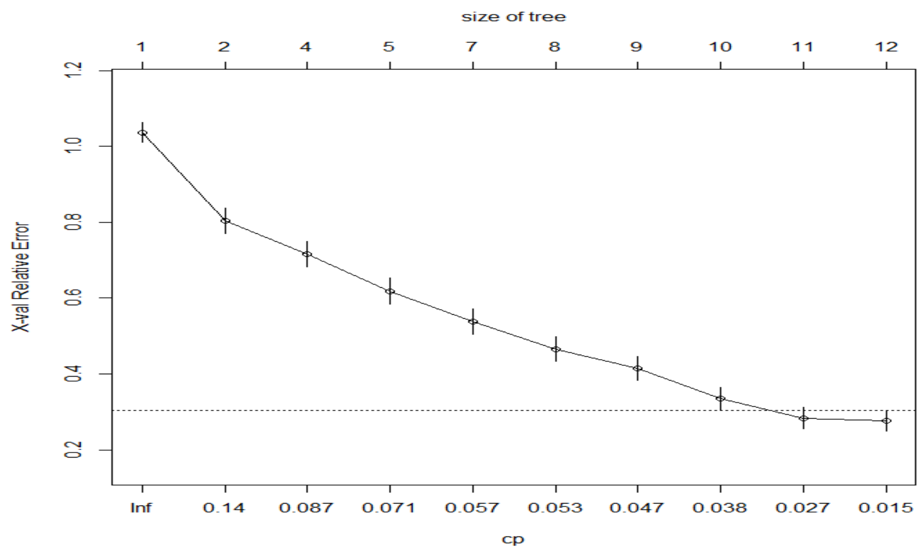

Figure-3. Regression tree.

Variable importance

Precipitation	Min temperature	Max temperature	Cloud cover	Vapour pressure	Wet day frequency
19	16	15	16	14	19

Node number 1: 347 observations, complexity param=0.2218182
predicted class=Drizzle expected loss=0.7925072 P(node) =1
class counts: 33 72 33 28 66 43 72
probabilities: 0.095 0.207 0.095 0.081 0.190 0.124 0.207
left son=2 (241 obs) right son=3 (106 obs)

Primary splits
vapour pressure < 24.865 to the right, improve=40.54612, (0 missing)
cloud cover < 47.42 to the right, improve=35.10961, (0 missing)
min temperature < 22.115 to the right, improve=30.18276, (0 missing)
wet day frequency < 4.535 to the right, improve=24.78694, (0 missing)
precipitation < 37.16 to the right, improve=24.56138, (0 missing)

Surrogate splits
min temperature < 21.49 to the right, agree=0.899, adj=0.670, (0 split)
precipitation < 36.195 to the right, agree=0.810, adj=0.377, (0 split)
cloud cover < 39.86 to the right, agree=0.810, adj=0.377, (0 split)
wet day frequency < 1.71 to the right, agree=0.801, adj=0.349, (0 split)
max temperature < 28.445 to the right, agree=0.755, adj=0.198, (0 split)

In Figure-2 scatter plot matrix shown for the classifier Support Vector Machine that provides the prediction towards the class. Recursive Partitioning and Regression Tree classifier is employed for the same dataset which appears in Table-1. Using the classifier it is also possible to substantiate the outcome of the result that is exhibited in Figure-3. By using all the three classifiers like Naïve Bayes, Support Vector Machine, Recursive Partitioning and Regression Trees show the same outcome that proves a decision towards the dataset which is literally trivial for the research work.

4. RESULTS AND DISCUSSIONS

In the eventual world mobile communication device is literally inevitable and in-disposable to the living society. Providing 'N' number of amenities to the mobile commodity gives a long mileage to the strong and sturdy community. In the proposed work spatial understanding is the main philosophy for 'N' number of components which constitute and provide great information. Through the mobile device it is possible to know the tourist spots and monuments right from the instant location. In the paper weather condition component is also assimilated to know the spot and its weather condition. Hence three classifiers are engaged to substantiate the class result. For any coming pattern, the classifiers able to afford an optimum result.

5. CONCLUSIONS

Dataset and the classifiers in data-mining provide greater prediction than an optimum result that reduce the unforeseen and uncertainty to the core. The above discussed three classifiers were engaged to analyze the incoming patterns to a great extent. Researchers have got a great intuition in applying different other classifiers which is also highly pronounced in the mining arena. Data-mining and Big-data is literally a flourishing area for prediction in the coming decades.

REFERENCES

[1] Karagiannis D. and Buchmann R. A. 2016. Linked Open Models: Extending Linked Open Data with conceptual model information. Information Systems. 56, 174-197.

[2] Knight D. W. and Cottrell S. P. 2016. Evaluating tourism-linked empowerment in Cuzco, Peru. Annals of Tourism Research. 56, 32-47.

[3] Nogales A., Sicilia M. A., Sánchez-Alonso S. and Garcia-Barriocanal E. 2016. Linking from Schema. org microdata to the Web of Linked Data: An empirical assessment. Computer Standards and Interfaces. 45, 90-99.

[4] Di Noia T., Ostuni V. C., Rosati J., Tomeo P., Di Sciascio E., Mirizzi R. and Bartolini C. 2016. Building a relatedness graph from Linked Open Data: a case study in the IT domain. Expert Systems with Applications. 44, 354-366.

[5] Binding C. and Tudhope D. 2015. Improving interoperability using vocabulary linked data. International Journal on Digital Libraries. 1-17.

[6] Martelli C., Bellini E. and Salvatori M. F. 2015. Knowledge management and reuse in tourism destination observatories. International Journal of Knowledge and Learning. 10(1): 1-15.

[7] Villa D. 2015. Crowdsourced Heritage Tourism Open-Data, Small-Data and e-Participatory Practices as Innovative Tools in Alps Cultural Heritage Topic: Information Technology and e-Tourism. In Cultural Tourism in a Digital Era (pp. 229-230). Springer International Publishing.

[8] Fermoso A. M., Mateos M., Beato M. E. and Berjón R. 2015. Open linked data and mobile devices as e-tourism tools. A practical approach to collaborative e-learning. Computers in Human Behavior. 51: 618-626.

[9] Sah M. and Wade V. 2015. Personalized concept-based search on the Linked Open Data. Web Semantics: Science, Services and Agents on the World Wide Web.

[10] Janssen M. and Kuk G. 2015. Big and Open Linked Data (BOLD) in Research, Policy and Practice. Journal of Organizational Computing and Electronic Commerce.

[11] Lausch A., Schmidt A. and Tischendorf L. 2015. Data mining and linked open data–New perspectives for data analysis in environmental research. Ecological Modelling. 295, 5-17.

[12] Lambrix P. 2014. Semantic Web, Ontologies and Linked Data.

[13] Fragkou P., Galiotou E. and Matsakas M. 2014. Enriching the e-GIF Ontology for an Improved Application of Linking Data Technologies to Greek Open Government Data. Procedia-Social and Behavioral Sciences. 147, 167-174.

[14] Ikkala E., Mäkelä E. and Hyvönen E. 2014. TourRDF: Representing, Enriching, and Publishing Curated Tours Based on Linked Data. In Knowledge Engineering and Knowledge Management (pp. 145-149). Springer International Publishing.

[15] Vert S. and Vasiu R. 2014. Relevant Aspects for the Integration of Linked Data in Mobile Augmented Reality Applications for Tourism. In Information and Software Technologies (pp. 334-345). Springer International Publishing.