ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

Review on Extracting Geo-Spatial Information From Open Street Map using Expat Semantic Parser and Support Vector Machine

Anita Janmeda [1], Dr.Dinesh Kumar [2]

P.G. Student, Department of Computer Science & Engineering, SRCEM at Palwal, Haryana, India[1]

Assistant Professor, Department of Computer Science & Engineering, SRCEM at Palwal, Haryana, India[2]

ABSTRACT: The OpenStreetMap project has developed into one of the largest VGI (Volunteered Geographic Information) datasets today. Over the years, its quality was analyzed and compared to commercial or authoritative datasets by various researchers and in different scenarios. Time and again, it proved to be a reliable alternative to commercial and authoritative datasets, despite being maintained mostly by amateurs and enthusiasts with little or no background in cartography or spatial sciences. After a decade of admirable results, a keen interest has arisen in the evolution of the project, most notably in the activities of its members and the changes they leave behind them. The aim of this research is to develop a augmented model that would aid masses and the OSM community in determining the nature of changes present in OSM objects, as well as their currency and credibility. The model described in this paper is able to analyze all OSM objects, regardless of their geometry type, location and semantic attributes to find out appropriate information from mammoth resource using Support Vector Machine classification.

KEYWORDS: Geo-Spatial Information , Open Street Maps (OSM), Expat Semantic Parsing, Crowed Sourcing, Support Vector Machine (SVM).

I. INTRODUCTION

One of the most interesting aspects of Web 2.0 was the emergence of crowd sourced information, which also represents one of the most significant and potentially controversial developments in Web 2.0. Crowd sourcing refers to large groups of users performing functions that are either difficult to automate or expensive to implement. It can be used in large-scale community activities focused on the development of software or on the collection and sharing of information. These activities are carried out by large groups of volunteers who work independently and without much coordination. Similar efforts are the foundation of the OpenStreetMap (OSM) project. Unlike other platforms that rely on user contributions in form of collected information about a particular subject, the OSM project contains more specific details about spatial elements (e.g. streets, buildings, etc.) which always include a geographic reference. This type of data is often described as Volunteered Geographic Information (VGI), while the whole process is described as crowd sourcing geospatial data. The OSM project has developed into one of the largest sources of VGI in recent years, and with the change of the licensing model by Google Maps in early 2012, more and more businesses are moving toward the free option offered by the OSM project. The location-based social network FourSquare and the Nestoria Property Search are the two major examples. Furthermore, professional spatial data providers and companies have created their own platforms which allow users to edit their own data on the provided maps (e.g. Google Map Maker, TomTom Map Share). These developments show that the success of the VGI approach to data collaboration and sharing is undeniable. On the other hand, most of the VGI projects (including OSM) rely on volunteers that do not necessarily have professional qualifications and background in geo data collection or surveying. Contribution to the project largely depends on the technical aspects (e.g. PC, Internet connection, GPS receiver, Smartphone, etc.) as well as the population density of specific areas. However, the local knowledge of most participants should make them local experts. In light of the data collection by amateurs, the distributed nature of data collection and the loose coordination in terms of standards, one of the significant questions about VGI is the quality of information collected through such activities. Recently, the quality of the OSM project has been analyzed in several studies by comparing the OSM dataset to authoritative or commercial dataset and analysed the quality of the OSM street and road network by comparing it with the Ordnance Survey (OS) dataset. The results of this research indicate that OSM information can be fairly accurate: on average within 6 m of the position recorded by the Ordnance Survey and with approximately 80% of overlap of motorway objects.

II. RELATED WORK

We discovered the optimal location query problem based on road networks. Specifically, we have a road network on which some clients and servers are located. Each client finds the server that is closest to her for service and her cost of getting served is equal to the (network) distance between the client and the server serving her multiplied by her weight or importance. The optimal location query problem is to find a location for setting up a new server such that the maximum cost of clients being served by the servers (including the new server) is minimized. This problem has been studied before, but the state-of-the-art is still not efficient enough. In this paper, we propose an efficient algorithm for the optimal location query problem, which is based on a novel idea of \emph{nearest location component}. We also discuss three extensions of the optimal location query problem, namely the optimal multiple-location query problem, the optimal location query problem on 3D road networks, and the optimal location query problem with another objective. Extensive experiments were conducted which showed that our algorithms are faster than the state-of-the-art by at least an order of magnitude on large real benchmark datasets. For example, on our largest real datasets, the state-of-the-art ran for more than 10 hours but our algorithm ran within 3 minutes only (i.e., >200 times faster). [1]

In research we presents Spatial Hadoop as the first full-fledged MapReduce framework with native support for spatial data. Spatial Hadoop is a comprehensive extension to Hadoop that pushes spatial data inside the core functionality of Hadoop. Spatial Hadoop runs existing Hadoop programs as is, yet, it achieves order(s) of magnitude better performance than Hadoop when dealing with spatial data. Spatial Hadoop employs a simple spatial high level language, a two-level spatial index structure, basic spatial components built inside the MapReduce layer, and three basic spatial operations: range queries, k-NN queries, and spatial join. Other spatial operations can be similarly deployed in SpatialHadoop. We demonstrate a real system prototype of SpatialHadoop running on an Amazon EC2 cluster against two sets of real spatial data obtained from Tiger Files and OpenStreetMap with sizes 60GB and 300GB, respectively.[2]

In paper describes the results of an analysis of the OpenStreetMap (OSM) database for the United Kingdom (UK) and Ireland (correct to April 2011). 15, 640 OSM ways (polygons and polylines), resulting in 316, 949 unique versions of these objects, were extracted and analysed from the OSM database for the UK and Ireland. In our analysis we only considered “heavily edited” objects in OSM: objects which have been edited 15 or more times. Our results show that there is no strong relationship between increasing numbers of contributors to a given object and the number of tags (metadata) assigned to it. 87% of contributions/edits to these objects are performed by 11% of the total 4128 contributors. In 79% of edits additional spatial data (nodes) are added to objects. The results in this paper do not attempt to evaluate the OSM data as good/poor quality but rather informs potential consumers of OSM data that the data itself is changing over time. In developing a better understanding of the characteristics of “heavily edited” objects there may be opportunities to use historical analysis in working towards quality indicators for OSM in the future. [3]

Real spatial data, e.g., detailed road networks, rivers, buildings, parks, are not easily available for most of the world. This hinders the practicality of many research ideas that need a real spatial data for testing and experiments. Such data is often available for governmental use, or at major software companies, but it is prohibitively expensive to build or buy for academia or individual researchers. This paper presents TAREEG; a web-service that makes real spatial data, from anywhere in the world, available at the fingertips of every researcher or individual. TAREEG gets all its data by leveraging the richness of OpenStreetMap data set; the most comprehensive available spatial data of the world. Yet, it is still challenging to obtain OpenStreetMap data due to the size limitations, special data format, and the noisy nature of spatial data. TAREEG employs MapReduce-based techniques to make it efficient and easy to extract OpenStreetMap

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

data in a standard form with minimal effort. Experimental results show that TAREEG is highly accurate and efficient.[4]

Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data Classification etc. Most of the existing supervised classification methods are based on traditional statistics, which can provide ideal results when sample size is tending to infinity. However, only finite samples can be acquired in practice. In this paper, a novel learning method, Support Vector Machine (SVM), is applied on different data (Diabetes data, Heart Data, Satellite Data and Shuttle data) which have two or multi class. SVM, a powerful machine method developed from statistical learning and has made significant achievement in some field. Introduced in the early 90's, they led to an explosion of interest in machine learning. The foundations of SVM have been developed by Vapnik and are gaining popularity in field of machine learning due to many attractive features and promising empirical performance. SVM method does not suffer the limitations of data dimensionality and limited samples. In our experiment, the support vectors, which are critical for classification, are obtained by learning from the training samples. In this paper we have shown the comparative results using different kernel functions for all data samples. [9]

III. PROPOSED SYSTEM

Extracting Geo-Spatial Information From OpenStreetMap using Expat Semantic Parser and Support Vector Machine the following series of steps will be followed to achieve the accurate and prompt information on the fly the following is amalgamated scenario of semantic expat parsing and SVM therein :-

1. **Map Extraction:** The respective Territory, City, State, Country corpus will be downloaded from openstreetmap.com termed as map extraction on the fly.
2. **Indexing:** This module runs immediately after the Data or Map Extraction module to index the extracted data. Hence, it is also a background process runs on a weekly basis to index the new downloaded data. Given the large size of the extracted data, proposed solution will leverages to partition and index the extracted data over a set of computing nodes in an R-tree-like way. It is important to note that each type of spatial data (e.g., parks, roads, and lakes) are partitioned and indexed separately. Hence, there will be one index designated for road network data over all available computing nodes, while another completely separate index will be designated for lakes or rivers data, and so on.
3. **Semantic Parsing:** This module receives the user requests for obtaining spatial data, converts the request to a range query with the user specified area and a predicate filter for the spatial feature (e.g., road network, lakes), and finally exploits the R-tree-like index structure to retrieve the requested data in an efficient way. In this module, The proposed solution will leverage the facts that vide Indexing module has partitioned the extracted data into multiple computing nodes to execute its range query over multiple nodes in parallel. Hence, an efficient query processing can be achieved using SVM.
4. **Data Extraction:** This module runs as a background process that infuses the SVM load spatial files from OpenStreetMap, extracts its spatial features, and cleans its noisy data. This module also classifies the extracted data into separate files, each represents one kind of spatial data, e.g., parks, road, or lakes. This module faces two main challenges: (1) The large volume of the dataset and (2) the noisy dataset coming from using non-standard tags.
5. **Front-End Visualizer**. This module is basically a console or web interface that allows users to express their data requests, calls the Data Extraction module for execution, sends an connecters to the users with links to the requested data, and finally visualize the extracted data on various consoles. This module is also responsible on producing the output data in various formats that include Comma Separated Values (CSV) further the csv will be bundled with and DBMS for query processing and result retrieval.

Below the diagrammatic representation is depicted to understand extracting Geo-Spatial Information From OpenStreetMap using Expat Semantic Parser and Support Vector Machine and workflow process :-

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

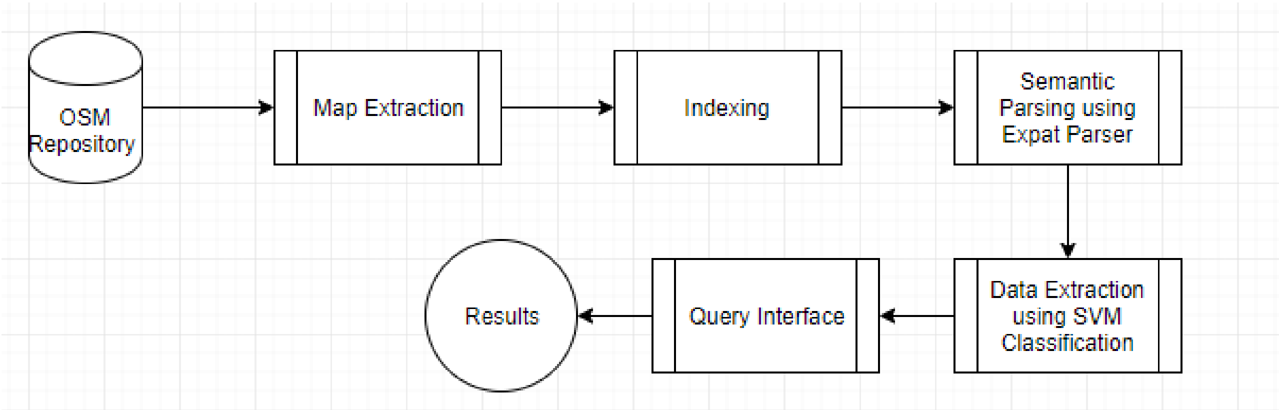

Figure 1: Work Flow of Proposed Scheme depicting Map Extraction, Indexing, Semantic Parsing, SVM Classification and Resultant Query Interface.

IV. CONCLUSION

We expect that this will deliver significant scientific outcomes, which will stimulate international research networking and collaboration. As outlined above, the inherent cross-disciplinary essence of OSM research combined with the emerging data quality, data mining, and patterns determination approaches to analysis of OSM using machine learning techniques i.e. SVM and Semantic Expat Parsing . We believe that, this inter-disciplinary contributions permit a deeper understanding of how the OSM works and will become the phenomenal success for future. Last, but not least, this scheme will strive to bring OSM into the core of GIScience where the diverse world of new and classical geography and cartography will meet requirements of users and customers. The above proposed system will form the effective and accurate geo-spatial extraction mechanism and will evaluated with existing schemes thereinafter to propose the better ecosystem for OSM.

REFERENCES

1. Z. Chen, Y. Liy, R. C.-W. Wong, J. Xiong, G. Mai, and C. Long. Efficient algorithms for optimal location queries in road networks. In SIGMOD, 2014.
2. A. Eldawy and M. F. Mokbel. A Demonstration of SpatialHadoop: An Efficient MapReduce Framework for Spatial Data (System Demo). In VLDB, Riva del Garda, Italy, Aug. 2013.
3. P. Mooney and P. Corcoran. Characteristics of heavily edited objects in openstreetmap. Future Internet, 2012.
4. L. Alarabi, A. Eldawy, R. Alghamdi, and M. F. Mokbel. TAREEG: A MapReduce-Based Web Service for Extracting Spatial Data from OpenStreetMap (System Demonstration). In SIGMOD, pages 897–900, Snowbird, UT, June 2014
5. X. Ma, S. Shekhar, and H. Xiong. Multi-type nearest neighbor queries in road networks with time window constraints. In SIGSPATIAL GIS, pages 484–487, 2009.
6. L. Zhu, Y. Jing, W. Sun, D. Mao, and P. Liu. Voronoi-based aggregate nearest neighbor query processing in road networks. In SIGSPATIAL GIS, pages 518–521, 2010.
7. L. Wu, X. Xiao, D. Deng, G. Cong, A. D. Zhu, and S. Zhou. Shortest path and distance queries on road networks: An experimental evaluation. PVLDB, 5(5):406–417, 2012.
8. S. Vanhove and V. Fack. An effective heuristic for computing many shortest path alternatives in road networks. International Journal of Geographical Information Science, 26(6):1031–1050, 2012.
9. Durgesh K. Srivastava, Lekha Bhambhu Data Classification Using Support Vector Machine, Journal of Theoretical and Applied Information Technology 2005 - 2009 JATIT.
10. Graham, M.; Hale, S.; Stephens, M. Digital Divide: “The Geography of Internet Access. Environ. Plan”. A 2012, 44, 1009–1010.
11. Jokar Arsanjani, J.; Helbich, M.; Bakillah, M.; Loos, L. “The Emergence and Evolution of OpenStreetMap: A Cellular Automata Approach”. Int. J. Digit. Earth 2013 b, 00, 1–15
12. Weiwei Sun, Chong Chen, “Merged aggregate nearest neighbor query processing in road network”, Singapore Management University Institutional Knowledge at Singapore Management University, 10-2013.
13. Zitong Chen, Yubao Liu, Raymond Chi-Wing Wong, “Efficient Algorithms for Optimal Location Queries in Road Networks”, The Hong Kong University of Science and Technology, Hong Kong, China.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

14. Aye Su Yee Win, “Fast Algorithm for Multi-type Nearest Neighbor Quer”, Graduate School of Science and Engineering, Saitama University, D-002.
15. Lingkun Wu, Xiaokui Xiao, “Shortest Path and Distance Queries on Road Networks: An Experimental Evaluation”, School of Computer Engineering , Nanyang Technological University, Singapore.
16. A. Eldawy and M. F. Mokbel. Pigeon: A spatial mapreduce language. In ICDE, pages 1242–1245, 2014