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 5, May 2019

Extracting Geo-Spatial Information from Open Street Map using Machine Learning

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

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

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

ABSTRACT: Access to spatial data and cartographic products has changed radically over the last decade or so. Customarily, legislative offices, cartographic focuses, and business organizations were the main hotspots for end-clients looking for spatial information. One of the most formidable barriers to more widespread access to these geodata were created by often prohibitive high fees and license charges in combination with time- and purpose-limited copyright restrictions imposed. This business model was rather successful, but made access to high-quality geodata very difficult for all but a small number of end users. Changes in Information and Communication Technology (ICT) brought about by the Internet and social media and the availability of inexpensive portable satellite navigation devices has seen this traditional geodata business model challenged. One of the key driving forces in this change has been the OpenStreetMap (OSM) project. OSM was launched in 2004 with the mission of creating an editable map of the whole world and released with an open content license (http://wiki.openstreetmap.org/wiki/About). When all is said in done, OSM goes for building and keeping up a free editable guide database of the world in a community oriented way so individuals and end-clients are not compelled to purchase geodata in the conventional way and consequently be subjected to prohibitive copyright and permit responsibilities. OSM started initially with a focus on mapping streets and roads. Since then it has moved far beyond these entities and it now contains a very rich variety of geographical objects (e.g., buildings, land use, Points of Interest) from all over the planet being mapped by thousands of volunteer contributors to the project. Aside from the plain industrial advantages offered by OSM, the project has revolutionized the method during which geodata is collected. notar the gathering of geodata and also the development of fashioning product restricted to specialists, geographic surveyors, or cartographers.

KEYWORDS: Geo-Spatial Information ,Open Street Maps (OSM),Crowed Sourcing, Machine Learning, Support Vector Machine (SVM)

I. INTRODUCTION

OSM is regularly alluded to as the Wikipedia guide of the world. As it depends on an extensive number of the same ICT structures as Wikipedia it offers its wander providers the probability of a) moderately fast invigorating of the guide database and furthermore to a great degree visit reviving of related modifying programming and diverse contraptions; b) getting geodata recorded from Global Positioning System (GPS)- enabled devices, phones, and other automated maps gadgets; c) access to the full history of mapping practices in OSM over its lifetime; ultimately d) collaboration with other OSM customers and supporters through various correspondence channels including mailing records, trade exchanges, and physical get-togethers (Mooney and Corcoran, 2013a). The slow development of the OSM biological system has been extremely effective. The undertaking got off to a moderate begin yet since 2007 there has been a consistently expanding rate of individuals joining the venture. In November 2014, OSM had roughly 1.85 million enlisted clients and supporters (http://wiki.openstreetmap.org/wiki/Stats). As said already, the period of omnipresent Internet, online life, open-source programming, and so on has seen numerous national information based ventures for a large group of differing purposes propelled on the Internet throughout the most recent couple of years. OSM has been a one of a kind case. The educational and mechanical systems have seen OSM not only in light of its rising to wind up a basic distributer of geodata yet its more broad accomplishment in growing an overall system of people willing to

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 5, May 2019

partake in the get-together and upkeep of geodata. The OSM people group is effectively associated with considerably more than gathering geodata to construct and keep up this worldwide geodatabase. What's more, the network is associated with, for instance, philanthropic work, open source programming improvement to help OSM and the GIS people group, and in building a system of help for those utilizing and adding to the OSM venture. As of late, a few logical controls (e.g. geology, GIScience, spatial masterminding, cartography, programming building, and condition) have comprehended the colossal capacity of OSM and it has transformed into the subject of academic research. OSM offers pros a novel dataset that is worldwide in scale and a gathering of learning made and kept up by a huge shared arrangement of volunteers. Research on OSM has exhibited that its geodata in a couple of segments of the world are more aggregate and locationally and semantically more exact than the contrasting prohibitive datasets (e.g., Zielstra and Zipf, 2010; Neis and Zipf, 2012; Helbich et al., 2012), while being of high spatial heterogeneity. Suspicion among the GIS social order and industry enveloping the idea of the geodata in OSM has seen an essential effort being made on evaluating the idea of the OSM geodata. This has an incited the headway of different programming gadgets and methods of insight for analyzing the quality (Roick et al., 2011; Helbich et al., 2012; Barron et al., 2013; JokarArsanjani et al., 2013 a; JokarArsanjani and Vaz, 2015). Diverse philosophies even undertaking to upgrade the OSM data through computations dedicated to specific inquiry composes, for instance, addresses for geocoding (Amelunxen, 2010). Examination of the progression and improvement of OSM over the globe after some time has moreover ascended as an investigation subject for a few, academic examinations (Mooney et al., 2012; Neis et al. 2012, 2013, JokarArsanjani et al., 2013c; Mooney and Corcoran, 2013, Fan et al. 2014).The relationship between OSM and open data standards, in particular Spatial Data Infrastructures (SDI) and the future direction of the Web 2.0 paradigm, is a question still requiring further discussion (Auer and Zipf, 2009). In particular, the large volumes of data being updated by the minute that are now available pose challenges with regards to their handling and keeping them up to date on a global scale. The discussion in the preceding paragraphs has shown that OSM has now emerged as a new research area. It has the potential to bring disparate research disciplines together and enhances interdisciplinary and multidisciplinary investigations. This interdisciplinary research collaboration can contribute to a more profound and cross-disciplinary understanding of citizens' knowledge-based efforts in projects such as OSM. It also provides an interesting platform for the academic research community to collaborate with these communities towards interactive collection of up-to-date geodata from citizens by means of novel computationally oriented methods such as network analysis, machine learning, and computer simulation models. As the examples above have demonstrated, these practical investigations on OSM provide a rich set of opportunities to discover novel and valuable patterns inherent in the geodata collected by citizens, to better understand the activities of contributors to open knowledge projects, the characteristics of their human-computer interactions, and the potential to tackle classical GIS research questions using this modern and revolutionary approach to the collection and distribution of geographic data.

Open Street Map (OSM) is a group constructed database of geographic information, containing client contributed neighbourhood and state-of-the-art data about historic points everywhere throughout the world. While the primary Programming interface is improved for altering map information, there exists a Programming interface that permit to channel delineate in view of inquiry criteria, for example, area, sort of articles, or highlights with which objects are labelled. In any case, issuing a question that is executable against the OSM database still requires nitty gritty information of database internals. something that can't be normal from a layman client. The objective of our work is the improvement of an interface to OSM that gives a client a chance to make an inquiry in normal dialect, which is then parsed into a database question that is executable against an online separating apparatus and profits OSM information for an intelligent guide. A client with less information of OSM ought to have the capacity to make inquiries that grasp the "fluffiness" of normal dialect, for instance, "What are the areas, names and phone quantities of lodgings in Paris with wheelchair get to that are near the station Gare du Nord?". To discover such data one would need to issue a question that requires point by point information of the database and the inquiry dialect:

" location[place= ' Bombay ']→.a;node(area.a)
[vide= ' cinema ']→.b;node(around.b:2000)
[entertainment='cinema '][wheelchair='yes'];out;".

As a beginning stage for a characteristic dialect interface we assembled a corpus of 2,380 common dialect inquiries combined with Machine Readable Language (MRL) formulae that we used to separate data utilizing a xml parser. We

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 5, May 2019

decide to physically making a corpus of OSMs from which structure and weights of a xml parser can be educated for a few reasons. We will give the OSM people group an arrangement of test addresses that can be executed and whose database question portrayal can be examined, beneath delineates similar information of OSM corpus.

```
<?xml version="1.0" encoding="UTF-8"?>
 <osm version="0.6"/>
        <node id="1" lat="21.4219827" lon="39.8336534">
        <tag k="traffic" v="light"/>
        </node>
        <node id="2" lat="21.4221823" lon="39.8331833">
                <tag k="highway" v="motorway_junction"/>
        </node>
        . . .
        <way id="6">
                <nd ref="1"/>
                <nd ref="2"/>
                <tag k="highway" v="service"/>
        </way>
        <way id="8">
                <nd ref="4"/>
                <nd ref="5"/>
        <tag k="type" v="multipolygon"/>
        </way>
        . . .
        <relation id="2">
                <member type="relation" ref="1" role="inner"/>
                <member type="way" ref="6" role="inner"/>
        <tag k="highway" v="primary"/>
        </relation>
        . . .
</osm>
```

Figure 1: Extensible Mark-up Language OSM Corpus data.

However, the determination and facilitation of Open Street Map (OSM) clients and engineers to perceive how the complex geological actualities can be issued as basic normal dialect questions that are parsed into executable channels on Open Street Maps objects.

II. RELATED WORK

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. Rather than the map itself, the data generated by the project is considered its primary output. The creation and growth of OSM has been motivated by restrictions on use or availability of map information across much of the world, and the advent of inexpensive portable satellite navigation devices.[1] OSM is considered a prominent example of volunteered geographic information. Created by Steve Coast in the UK in 2004, it was inspired by the success of Wikipedia[2] and the predominance of proprietary map data in the UK and elsewhere.[38] Since then, it has grown to over two million registered users,[9] who can collect data using manual survey, GPS devices, aerial photography, and other free sources. This crowdsourced data is then made available under the Open Database License. The site is supported by the OpenStreetMap Foundation, a non-profit organisation registered in England and Wales.

The data from OSM is available for use in both traditional applications, like its usage by Facebook, Craigslist, OsmAnd, Geocaching, MapQuest Open, JMP statistical software, and Foursquare to replace Google Maps, and more unusual roles like replacing the default data included with GPS receivers.[3] OpenStreetMap data has been favourably compared with proprietary datasources,[4] although in 2009 data quality varied across the world.[5][6]

Steve Coast founded the project in 2004, initially focusing on mapping the United Kingdom. In the UK and elsewhere, government-run and tax-funded projects like the Ordnance Survey created massive datasets but failed to freely and widely distribute them. The first contribution, made in the British city of London in 2005[7], was thought to be a road by the Directions Mag.[8]

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 5, May 2019

In April 2006, the OpenStreetMap Foundation was established to encourage the growth, development and distribution of free geospatial data and provide geospatial data for anybody to use and share. In December 2006, Yahoo! confirmed that OpenStreetMap could use its aerial photography as a backdrop for map production.[9]

In April 2007, Automotive Navigation Data (AND) donated a complete road data set for the Netherlands and trunk road data for India and China to the project[10] and by July 2007, when the first OSM international The State of the Map conference was held, there were 9,000 registered users. Sponsors of the event included Google, Yahoo! and Multimap. In October 2007, OpenStreetMap completed the import of a US Census TIGER road dataset.[11][12] In December 2007, Oxford University became the first major organisation to use OpenStreetMap data on their main website.[13]

III. PROPOSED SYSTEM

Main idea of the proposed scheme is to inculcate the semantic parsing using SVM. However under the scheme we introduce OSM as a new knowledge base that has not, to the best of our knowledge, been used for question answering, and offer a new corpus to the research community. Work builds the basis of a natural language interface using SVM to OSM that will be enabling for interesting directions of future research, e.g., response-based learning to improve parsing and multilingual database access.The common approach to semantic parsing is a manual annotation of a corpus with natural language utterances and machine readable formulae which are then used to learn the structure and weights of a semantic parser. Corpora that have been utilized for preparing and testing various semantic parsers are GEOQUERY (Zelle and Mooney, 1996; Kate et al., 2005) and FREE917 (Cai and Yates, 2013).)

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

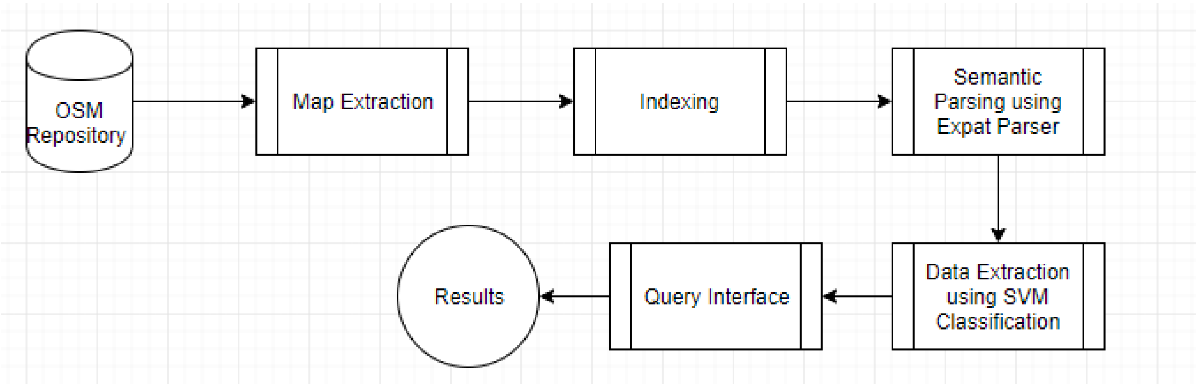

Figure 2. Proposed Architecture of the Amplified Semantic Parsing of Open Street Map using SVM

IV. ALGORITHM

In this we present the formulation for SVM classification representation as under.

SV classification:

$$\min_{f,\xi_i} \|f\|_K^2 + C\sum_{i=1}^{l}\xi_i \qquad y_i f(\mathbf{x}_i) \geq 1 - \xi_i, \text{ for all } i\ \xi_i \geq 0$$

SVM classification, Dual formulation:

$$\min_{\alpha_i} \sum_{i=1}^{1}\alpha_i - \frac{1}{2}\sum_{i=1}^{1}\sum_{j=1}^{1}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \qquad 0 \leq \alpha_i \leq C, \text{ for all } i; \qquad \sum_{i=1}^{l}\alpha_i y_i = 0$$

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 5, May 2019

Variables ξ_i are called slack variables and they measure the error made at point ($\mathbf{x}_i$,y_i). Training SVM becomes quite challenging when the number of ways, nodes, change sets, id points to latitude and longitude based on tag-nodes and dimensions . A number of methods for fast SVM training have been proposed with semantic parsing which is pseudonym derived as under:-

```
Algorithm using python herewith for explanation :-
importxml.etree.ElementTree as ET  # Use cElementTree or lxml if too slow
OSM_FILE = "/ML/new-delhi_india.osm"
SAMPLE_FILE = "/ML/new-delhi_sample.osm"
k = 100 # Parameter: take every k-th top level element
defget_element(osm_file, tags=('node', 'way', 'relation')):
context = iter(ET.iterparse(osm_file, events=('start', 'end')))
    _, root = next(context)
for event, elem in context:
if event == 'end' and elem.tag in tags:
yieldelem
root.clear()
with open(SAMPLE_FILE, 'wb') as output:
output.write('<?xml version="1.0" encoding="UTF-8"?>\n')
output.write('<osm>\n  ')
    # Write every kth top level element
fori, element in enumerate(get_element(OSM_FILE)):
ifi % k == 0:
output.write(ET.tostring(element, encoding='utf-8'))
output.write('</osm>')
defcount_tags(filename):
tags = {}
for event, elem in ET.iterparse(filename):
if elem.tag in tags:
tags[elem.tag] += 1
else:
tags[elem.tag] = 1
return tags
pprint.pprint(count_tags(OSMFILE))
```

V. Results

The common approach to semantic parsing is a manual annotation of a corpus with natural language utterances and machine readable formulae which are then used to learn the structure and weights of a semantic parser using SVM.

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 5, May 2019

```
[root@localhost anita]# python 4_audit.py
{'15th': set(['15th Road']),
 '16th': set(['16th Road']),
 '45,': set(['45, Tarun Bharat Co Op Society']),
 '4th': set(['4th Pasta Lane']),
 'Dinshaw': set(['Dinshaw Wachha Road']),
 'Dr': set(['Dr Ambedkar Road, Bandra West']),
 'Garden': set(['Garden Road']),
 'Hanuman': set(['Hanuman Road']),
 'Hindcycle': set(['Hindcycle Road, Worli, Mumbai']),
 'Hiranandani': set(['Hiranandani Estate']),
 'K': set(['K C Marg']),
 'Mathew': set(['Mathew Road']),
 'Mathurdas': set(['Mathurdas Road']),
 'Navghar': set(['Navghar Marg']),
 'Navi': set(['Navi Wadi']),
 'New': set(['New Link Road, Andheri West']),
 'Sector-48': set(['Sector-48']),
 'Shanti': set(['Shanti Vidya Nagri Road']),
 'St.': set(['St. Andrew Road', "St. Martin's Road"]),
 'Vidhan': set(['Vidhan Bhavan Road']),
 'Waman': set(['Waman Tukaram Patil Marg'])}
Sector-48 => Sector-48
Navi Wadi => Navi Wadi
Mathurdas Road => Mathurdas Road
Hiranandani Estate => Hiranandani Estate
St. Andrew Road => St. Andrew Road
St. Martin's Road => St. Martin'S Road
Dinshaw Wachha Road => Dinshaw Wachha Road
Waman Tukaram Patil Marg => Waman Tukaram Patil Marg
```

Figure 3.Result Extracted using Proposed Amplified Semantic Parsing of Open Street Map and SVM

VI. CONCLUSION AND FUTURE SCOPE

Under the proposed scheme ee 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 for prompt and accurate information anytime and anywhere scenario. 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 globally.

For future scope though we can extract the map of continent, country, state or city and then download to machine for further process and procedures that we have inculcated in the proposed scheme, However, the same solution using python with CGI can be integrated on server models to extract the information online itself

REFERENCES

1. 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.
2. P. Mooney and P. Corcoran. Characteristics of heavily edited objects in openstreetmap. Future Internet, 2012.
3. OpenStreetMap Wiki - History of OpenStreetMap
4. Anderson, Mark (18 October 2006). "Global Positioning Tech Inspires Do-It-Yourself Mapping Project". National Geographic News. Retrieved 25 February 2012.
5. Lardinois, Frederic (9 August 2014). "For the Love of Mapping Data". TechCrunch.

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 5, May 2019

6. Frederick Ramm; JochenTopf; Steve Chilton (2011). OpenStreetMap: Using and Enhancing the Free Map of the World. UIT Cambridge.
7. Neis, Pascal; Zipf, Alexander (2012), "Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap", ISPRS Int. J. Geo-Inf., 1 (2): 146–165, Bibcode:2012IJGI....1..146N, doi:10.3390/ijgi1020146
8. OSM Maps on Garmin. OpenStreetMap Wiki. Retrieved 25 July 2014.
9. Zielstra, Dennis. "Comparing Shortest Paths Lengths of Free and Proprietary Data for Effective Pedestrian Routing in Street Networks" (PDF). University of Florida, Geomatics Program. Archived from the original (PDF) on 15 December 2012. Retrieved 14 November 2012.
10. Haklay, M. (2010). "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets" (PDF). Environment and Planning B: Planning and Design. 37 (4): 682–703. doi:10.1068/b35097.
11. Coleman, D. (2013). "Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs". In Sui, D.; Elwood, S; Goodchild, M. (eds.). Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. New York, London: Springer Science+Business Media Dordrecht. pp. 245–264. doi:10.1007/978-94-007-4587-2. ISBN 978-94-007-4586-5.
12. Coast, Steve. "Changest #1 on OpenStreetMap". OpenStreetMap. Retrieved 19 September 2018.
13. Sinton, Diana (6 April 2016). "OSM: The simple map that became a global movement". The Directions Mag. Retrieved 19 September 2018.
14. Coast, Steve (4 December 2006). "Yahoo! aerial imagery in OSM". OpenGeoData. Retrieved 15 April 2011.
15. Coast, Steve (4 July 2007). "AND donate entire Netherlands to OpenStreetMap". OpenGeoData. Retrieved 15 April 2011.