Semantic Similarity Assessment of Volunteered Geographic Information

Dr. Maythm Al-Bakri
Lecturer
Department of Surveying
College of Engineering
University of Baghdad
Email: m.m.s.albakri@gmail.com

ABSTRACT

The recent development in communication technologies between individuals allows for the establishment of more informal collaborative map data projects which are called volunteered geographic information (VGI). These projects, such as OpenStreetMap (OSM) project, seek to create free alternative maps which let users add or input new materials to the data of others. The information of different VGI data sources is often not compliant to any standard and each organization is producing a dataset at various level of richness. In this research the assessment of semantic data quality provided by web sources, e.g. OSM will depend on a comparison with the information from standard sources. This will include the validity of semantic accuracy as one of the most important parameter of spatial data quality parameters. Semantic similarity testing covered feature classification, in effect comparing possible categories (legend classes) and actual attributes attached to features. This will be achieved by developing a tool, using Matlab programming language, for analysing and examining OSM semantic accuracy. To identify the strength of semantic accuracy assessment strategy, there are many factors should be considered. For instance, the confusion matrix of feature classifications can be assessed, and different statistical tests should be passed. The results revealed good semantic accuracy of OSM datasets.

Key words: OpenStreetMap; Semantic; Classification; Accuracy; Confusion Matrix

تقييم دقة تصنيفات عوارض الخرائط المُنتجة على الشبكة العنكبوتية

م.د. ميثم مطشر شرقي
قسم هندسة المساحة
كلية الهندسة / جامعة بغداد

الخلاصة

سمحت التطورات الأخيرة في تكنلوجيا جمع المعلومات والإتصالات بين الأفراد لتأسيس وإنتاج الخرائط التعاونية الغير رسمية على الشبكة العنكبوتية والتي تسمى بالمعلومات الجغرافية المنتجة من قبل المتطوعين. إن هذه المشاريع، مثل خارطة الشارع المفتوح (OSM) تسعى لانتاج خرائط مجانية بديلة للخرائط التي تنتج من المصادر الحكومية الرسمية. غالبا ما تكون البيانات المنتجة من المصادر المختلفة للبيانات المكانية المجانية غير متوافقة مع أي معيار وكل مجموعة من المتطوعين تنتج بيانات على مستويات مختلفة من الدقة والنوعية. يهدف هذا البحث إلى تقييم جودة التصنيفات الاسمية للعوارض المنتجة للخرائط على الشبكة العنكبوتية. لقد تم مقارنة البيانات الاسمية المنتجة من خارطة الشارع المفتوح (OSM) مع المعلومات من المصادر القياسية مثل المسح الحقلي. إن هذا الفحص يتضمن مقارنة تسميات العوارض الطبيعية والصناعية المعرفة ضمن مفتاح الخارطة او العناصر المرفقة كجداول مع البيانات الجغرافية. لإنجاز ذلك تم بناء برنامج متخصص باستخدام لغة الماتلاب لقياس وتحليل دقة البيانات

الاسمية لخارطة الشارع المفتوح (OSM) ، ولتحديد قوة استراتيجية الاسلوب المتبع في هذه الدراسة هناك عدة عناصر يجب أن تؤخذ بنظر الاعتبار مثل تكوين وتقييم مصفوفة التمييز الإسمي فضلاً عن إجراء فحوصات احصائية متعددة. لقد أظهرت نتائج البحث في هذه الدراسة إن دقة التصنيفات الإسمية لخارطة الشارع المفتوح جيدة الى حدٍ ما.

1. INTRODUCTION

Traditionally, professional surveyors seek to produce geospatial data through different methods such as plane surveying, remote sensing and photogrammetry. The data captured by these methods can usually ordered by users in digital or paper formats. However, recently by web developments, a free geospatial data based on volunteers' efforts have appeared on the Internet, **Esmaili, et al., 2013**. **Goodchild, 2007** termed this phenomenon as volunteered geographic information (VGI). Most VGI data are collected and distributed through the World Wide Web (www) by non-professionals users. Compared to authoritative datasets, the VGI data are freely and dynamically grown with the help of volunteers. The dynamic aspect of VGI service encouraged some mapping agencies and local authorities to utilize VGI to create new datasets and update their data. There are a broad categories of VGI sites are now available on the Internet. Such websites include Google Map, Flickr, Map share, Wikimapia and OpenStreetMap (OSM). The data from OSM project was investigated and analysed in this study because the OSM data can be downloaded easily and freely which makes it ideal for research.

The OSM project was founded in London UK in 2004. The OSM data are essentially collected using GPS receivers, and then transformed into map using online editing tools. After 2006, Yahoo started supplying free satellite images to OSM community; therefore mapping becomes achieved directly from the images, **Ma, et al., 2015**. The OSM is a collaborative mapping project which creates a free editable map for the whole world. The OSM project can be easily shared under its "Open Data Commons Open Database License (ODbL)", **Dorn, et al., 2015**. Despite the positive aspects of OSM data, the quality assurance of its data is still the major concern of geographical information (GI) users. In recent years, VGI quality has been interesting topic in GIS research. One of the first systematic attempts to assess OSM quality was conducted by **Haklay, 2010**. In addition to Haklay, a large and growing body of literature has investigated the quality of OSM project; see for example **Liu, et al., 2015**; **Sehra, et al., 2014; Koukoletsos et al., 2012**. Although these studies analysed the accuracy of OSM data compared to commercial or authoritative datasets, however the OSM data poses challenges regarding the heterogeneity of classifications or semantic data.

To date, several studies have begun to examine the semantic accuracy of OSM project. For instance, **Vandecasteele** and **Devillers, 2013** suggested a method for reducing the semantic inconsistency and improving the semantic data of OSM project. The method was implemented into a plugin for OSM and different examples illustrate how this plugin can be used to enhance the quality of VGI data. In another major study, **Ramos, et al., 2013** presented a procedure to measure the similarity between the correspondence features in OSM and official datasets. The proposed method was based on using ontologies for handling semantic heterogeneities. In a study conducted by **Ballatore, et al., 2013,** it was described a knowledge-based technique to identify the semantic similarity of lexical definitions. In the same view, **Girres,** and **Touya, 2010** assessed the semantic accuracy of 585 roads in OSM data compared to BD TOPO Large Scale Referential (RGE) from IGN as a reference data. The analysis showed that the percentages of correct classifications are varying for different classifications. For example motorways have 100% semantically correct, whereas the secondary roads have only 49% similarity.

The objective of this study is to evaluate the semantic accuracy of the OSM "tag" also called "features". The main idea is developing a methodological framework based on a confusion matrix approach to determine the classification accuracy using all of the information diversity of OSM project. The reminder of this article is structured as follows: section 2 describes the main characteristics of the features of OSM data. Section 3 presents the approaches for estimating classification accuracy. The discussion of the system prerequisites and code program will be introduced in section 4. In section 5, the results and findings are illustrated and analysed. The last section concludes with a discussion and provides an outlook on future research.

2. THE FEATURES OF OPENSTREETMAP DATA

The OpenStreetmap (OSM) is not the only source of feature classifications data that is available free of charge. Services such as Google Maps, Yahoo Maps, or the Microsoft offering Bing maps have very good mapping available for viewing via the Internet, and they do not require payment. In 2008, Google introduced "Map Maker", an edition that allows users to trace maps data from satellite or aerial imagery and upload it to Google servers. Google is using map maker primarily in countries, such as Iraq, where they cannot buy suitable map data from the traditional geodata supplier. All these offerings only very limited rights to the users for downloading maps with feature classifications. If users would like to add features to web maps, for example, these free

Internet sources are usually not usable. With OpenStreetMap, on the other hand, any form of reproduction or processing is allowed, and users do not have to ask anybody for permission.

The objects of OSM data may be classified into two most important types: nodes (also called points) and ways. A node consists of geographical coordinates (latitude and longitude), while a way consists of an ordered list of at least two nodes. Attributes assigned to these objects in order to describe what they represent are called tags. A tag consists of a key and a value and is usually written with an equals sign between both parts “key=value”. Both can be arbitrary strings of up to 255 characters. The most important tags from OSM map features may be divided into a number of groups such as roads and railways; forests, lakes, and rivers; coastline and islands; buildings and land use areas; villages, cities, and borders, **Ramm, et al., 2011**. An example of OSM tags or classifications can be seen in **Fig 1**.

The OSM data can be exported directly in a variety of data formats such as XML data or Mapnik image (e.g. PNG, JPEG, and PDF). The OSM raw data can be processed with suitable OSM software. This export has a size limit; users can only use it if they are looking at a reasonably small area of the map (approximately 10km x 10km), **Ramm, et al., 2011**. In this research the OSM data was exported as XML format, and it was imported using ArcGIS 9.3 software for processing and manipulating. In order to obtain the required features, a pre-processing (filter) step was adopted. This step was essentially applied for separating the undesired data from attribute table.

3. APPROACHES FOR ESTIMATING CLASSIFICATION ACCURACY

3.1 Confusion (Error) Matrix

An error or confusion matrix evaluates classification accuracy based on comparing actual or reference land class with map data. The matrix has tow dimension with the same number of rows and columns. The rows and columns express the labels of samples assigned to a particular category in one classification relative to the labels of samples assigned to a particular category in another classification **(Fig. 2)**. Each column is assumed to be correct and display the ground reference information, while each row of the matrix represents the map labels obtained from map classifications. The main diagonal of the matrix represents the correct feature classifications, **Congalton and Green, 2009**.

Confusion matrix can be considered one of the most effective ways to represent classification accuracy. This is because that error matrix can describe the accuracy of each map category based on omission and commission errors. The error of omission refers to the proportion of observed features excluded from map classes, whereas commission error arises when features on map are categorised incorrectly. Besides the omission and commission errors, the overall, producer's and user's accuracy can also be determined by confusion matrix, **Ismail** and **Jusoff, 2008.** The overall accuracy represents the summation of elements on the main diagonal of confusion matrix divided by the total number of samples of confusion matrix. The user's and producer's accuracies are simply the ways of computing individual accuracy rather than calculating overall accuracy, as will be discussed in the following section.

3.2 Mathematical Representation of the Error Matrix

Suppose that there is a square matrix with k^2 cells and n samples. Each sample represents one of k class in a map data, and one of the identical k class in the reference classifications, let n_{ij} symbolize the amount of samples classified into category i (where: i = 1,2,......, k) in the map data and category j (where: j = 1,2,....., k) in the reference classes , **Congalton and Green, 2009**, (as shown in **Fig. 2**).

Let:

$$n_{i+} = \sum_{j=1}^{k} n_{ij} \qquad (1)$$

Be the number of samples classified into category i in the map classification, and

$$n_{+j} = \sum_{i=1}^{k} n_{ij} \qquad (2)$$

Be the number of sampled classified into category j in the reference data set.

Overall accuracy between map classification and the reference data can be computed as follows:

$$overall\ accuracy = \frac{\sum_{i=1}^{k} n_{ij}}{n} \qquad (3)$$

Producer's accuracy can be computed by:

$$producer's accuracy\ j = \frac{n_{jj}}{n_{+j}} \qquad (4)$$

And the user's accuracy can be computed by:

$$user's accuracy = \frac{n_{ii}}{n_{i+}} \qquad (5)$$

In addition to the above models, kappa coefficient can also be used as an index of classification accuracy, as follows:

$$\kappa = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+}.x_{+i}}{N^2 - \sum_{i=1}^{r}\left(x_{i+}.x_{+i}\right)} \qquad (6)$$

Where:

$r = number\ of\ rows\ in\ the\ error\ matrix.$

$x_{ii} = number\ of\ observations\ in\ row\ i\ and\ coloumn\ i\ (on\ the\ major\ diagonal).$

$x_{i+} = total\ of\ obsevations\ in\ row\ i\ (shown\ as\ marginal\ total\ to\ right\ of\ the\ matrix).$

$x_{+i} = total\ of\ observations\ in\ column\ i\ (shown\ as\ marginal\ total\ at\ bottom\ of\ the\ matrix).$

$N = total\ number\ of\ observations\ included\ in\ matrix.$

4. METHODOLOGY IMPLEMENTATON AND PROGRAM STRUCTURE

To achieve the main goal of this research, the classifications quality of OpenStreetMap (OSM) information must be checked carefully. It is indispensable because the information of different Volunteered Geographic Information (VGI) data sources are often not complaint to any standard and each organisation is producing a dataset at various level of richness. In this study the assessment of classifications quality of data provided by web sources will base on comparison with the information from other sources. In other words, utilizing the information from sources with known quality of data to evaluate the quality of data provided by sources with unknown quality of data. As **Thakkar, et al., 2007** observed in relation to VGI data sources and high quality sources, using this technique can produce the most accurate results. Assessment of attribute and feature based accuracy will be undertaken using statistical indices based on, for example, kappa coefficients (as described in previous section). It is proposed that the true or actual classifications by field surveying are to be used for geospatial data collecting of selected location (Baghdad / Iraq). The data set will consist of the self-generated field survey and the open access data from web-based VGI (e.g. OSM).

The intention is to compare classifications of these datasets. Initially this can be done with visual comparison of derived maps, to get a general picture but the methods indicated above will be applied to construct a quantitative approach, identifying the strength of accuracy assessment strategy. This exercise will help in developing methods of evaluating classification quality based on a rule set, coded into the data handling flowline. The result will be a set of operators which can measure data quality, allow for the preparation of datasets prior to successful integration, and actually undertake the integration of data. The proposed methodology was developed using Matlab 7.10.0 programming language, for academic and research purposes. In the first step the program inputs the features name or classifications through read a data file. Then, the input data will be employed to generate the confusion matrix by comparing the feature classes in reference and tested datasets. Subsequently, the program saved the completed confusion matrix and assessed the accuracy of classification data. The assessment processing based on determining the produce, user, an overall accuracy and kappa coefficient values. The report of final outcome can be exported and saved as text (.txt) or excel (.xlsx) format. **Fig. 3** presents the flowchart of the program that used in implementing the methodology of this study.

5. EXPERIMENT AND ANALYSIS

In this research, the obtained confusion matrix constructed from 543 features and 16 classes as shown in **Fig. 4**. In this matrix the x-axis represented the true (reference) class labels, whereas the y-axis listed the tested classes. The correct classifications lay on the main diagonal of the matrix, while the other elements of the matrix showed the misclassifications. The overall number of classifications can be seen in the bottom right cell. The accuracy measurements of confusion matrix have been achieved based on equations 3, 4, 5 and 6 as follows:

$$Overall\ accuracy = \frac{\sum_{i=1}^{k} n_{ij}}{n}$$

$$Overall\ accuracy = \frac{6+10+56+28+168+2+17+141+1+3+11+2+2+6+13+3}{543}$$

$$= \frac{469}{543} = 86\%$$

$$Producer's accuracy\ j = \frac{n_{jj}}{n_{+j}}$$

$$Producer'saccuracy\ (Primary\ road) = \frac{6}{11} = 55\%$$

The procedure's accuracy of the primary road is one example for determining the procedures' accuracy in this article. Figure 5 shows the procedure's accuracy of other features.

$$User'saccuracy = \frac{n_{ii}}{n_{i+}}$$

$$User'saccuracy\ (Primary\ road) = \frac{6}{11} = 55\%$$

The user's accuracy of the primary road is one example for determining the users' accuracy in this article. Figure 6 shows the users' accuracy of other features.

To illustrate the computation of kappa coefficient (k) for the error matrix included in **Fig. 4**:

$$\sum_{i=1}^{r} x_{ii} = 6 + 10 + 56 + 28 + 168 + 2 + 17 + 141 + 1 + 3 + 11 + 2 + 2 + 6 + 13 + 3 = 469$$

$$\begin{aligned}\sum_{i=1}^{r}(x_{i+}.x_{+i}) = {} & (11 \times 11) + (19 \times 13) + (66 \times 77) + (29 \times 28) + (170 \times 181) + (3 \times 2) \\ & + (22 \times 18) + (164 \times 157) + (1 \times 2) + (3 \times 3) + (11 \times 13) + (2 \times 3) + (2 \times 2) \\ & + (19 \times 7) + (17 \times 22) + (4 \times 4) = 121 + 247 + 5082 + 812 + 30770 + 6 + 396 \\ & + 25748 + 2 + 9 + 143 + 6 + 4 + 133 + 374 + 16 = 63869\end{aligned}$$

$$k = \frac{543(469) - 63869}{(543)^2 - 63869} = \frac{190798}{230980} = 0.826$$

The accuracy assessment reports of OSM classifications from 543 reference data are illustrated in Figure 4, 5, and 6. The results showed that the overall accuracy of the tested data was 86%, where the lowest user and procedure accuracy were 32% and 50%, respectively, and the highest user and procedure accuracy were similar with 100%. The classification accuracy is vary and different from one class to another class. For instance, primary road has only 55% accuracy since it was confused with secondary road, service road, and path. The path has only 77% accuracy since it was confused with building. Another example of this is the building which has 90% accuracy since it was confused with secondary road, service road, parking and university. The most likely causes of diverse classification accuracy are because the wrong way for classifications of some of OSM datasets.

The kappa coefficient was also determined from the calculations of tested features and classes. This represents that the kappa statistics value of 0.826 which implies a credible 82% better accuracy than if a random unsupervised classification was adopted. According to **Landis** and **Koch, 1987**, the agreement scale of Kappa value as $k>0.75$ present excellent, $0.4<k<0.7$ present good, and $k<0.4$ present poor. In another major study, **Monserud, 1990**, reported that the Kappa statistic is poor when $k<0.4$, fair when $0.40<k<0.55$, good when $0.55<k<0.70$, very good when $0.70<k<0.85$, and excellent when $k>0.85$. In general, therefore, it seems that the kappa coefficient value of this study demonstrated an excellent to a very good agreement.

6. CONCLUSION

The OpenStreetMap (OSM) is one of the most popular projects of Volunteered Geographic Information (VGI) services. The OSM produced geospatial data by non professional volunteers of varying level of mapping experience. The OSM data does not follow any standard compared to authoritative or official datasets; therefore it's necessary to evaluate its quality continuously. The purpose of the current study was to present a method for assessing the quality of (OSM) semantic data. The methodology was implemented by designing a program using Matlab 7.10.0 programming language. The program was utilised in the assessment of classification accuracy of feature categories of OSM data. This was included the construction of confusion matrix and calculating the overall accuracy, users' accuracy, producers' accuracy and kappa coefficient.

The outcome of this investigation showed that the confusion matrix consisted of 543 elements, which is formed as 16 rows and 16 columns (as illustrated in Fig.4). The number of elements in each row and column are varying and different based on the number of features in each class. For instance there are eleven elements in the first column. These classified as six primary roads, three secondary roads, and two residential roads. Another example of what is meant by different elements of rows and columns of confusion matrix is the fifth row contains 170 elements. These distributed as two primary roads, and one hundred and sixty eight as residential roads. The research has also found that the overall accuracy was 86%; the users' accuracy was between 32% and 100%, while producers' accuracy was between 50% and 100%, and kappa statistics was 0.826. In general, therefore, it seems that the classification accuracy of OSM datasets is acceptable to some extent.

For future work, it is recommended that the further studies need to be carried out in order to apply this method with different data sources such as governmental agency data, Google map, and Wikimapia. Testing semantic accuracy of several geospatial data source can give an idea about the possibility of integrating different geospatial datasets to improve and enhance its quality. It is also suggested that the classification of OSM data can be investigated and assessed based on ontologies and definitions of online dictionary.

REFERENCES

- Ballatore, A., Wilson, D., and Bertolotto, M., 2013, *Computing the Semantic Similarity of Geographic Terms using Volunteered Lexical Definitions*, International Journal of Geographical Information Science, Vol. 27(10), PP. 2099–2118.

- Congalton, R. G. and Green, K., 2009, 'Thematic accuracy ', in *Assessing the Accuracy of Remotely Sensed Data:Priciples and Practices.* USA: CRC Press, PP. 55-59.

- Dorn, H., Törnros, T. and Zipf, A., 2015, *Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany*, ISPRS International Journal of Geo-Information, 4, PP. 1657-1671.

- Esmaili, R., Naseri, F., and Esmaili, A., 2013, *Quality Assessment of Volunteered Geographic Information*, American Journal of Geographic Information System, 2(2), PP. 19-26.

- Girres, J., and Touya, G., 2010, *Quality Assessment of the French OpenStreetMap Dataset*, Transactions in GIS, 14(4): PP. 435–459.

- Goodchild, M.F., 2007, *Citizens as Sensors: The World of Volunteered Geography*, Geojournal, 69 (4), PP. 211-221.

- Haklay, M., 2010, *How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets*, Environ. Plan. B Plan., 37, PP. 682–703.

- Ismail, M., H., and Jusoff, K., 2008, *Satellite Data Classification Accuracy Assessment Based from Reference Dataset*, International Journal of Computer and Information Science and Engineering, 2 (2). PP. 96-102.

- Koukoletsos, T., Haklay, M., Ellul, C., 2012, *Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data*, Transactions in GIS, 16(4), PP. 477–498.

- Landis,J.R. and Koch,G.C., 1987, *The Measurement of Observer Agreement for Categorical Data*, Biometric, 33: PP. 159-179.

- Liu, C., Xiong, L., Hu, X., and Shan, J., 2015, *A Progressive Buffering Method for Road Map Update Using OpenStreetMap Data*, ISPRS International Journal of Geo-Information, 4, PP. 1246-1264.

- Ma, D., Sandberg, M., and Jiang, B., 2015, *Characterizing the Heterogeneity of the OpenStreetMap Data and Community*, ISPRS International Journal of Geo-Information, 4, PP. 535-550.

- Monserud,R.A., 1990, *Method for Comparing Global Vegetation Map*, WP- 90-40 II ASA,Laxenburg.

- OpenStreetMap, 2015, accessed on 27-12-2015, available at https://www.openstreetmap.org/note/new#map=15/33.2996/44.4606&layers=D

- Ramm, F., Topf, J. and Chilton, S., 2011, *OpenStreetMap - Using and Enhancing the Free Map of the World*, Cambridge, England: UIT Cambridge Ltd.

- Ramos J, M., Vandecasteele A., and Devilers, R., 2013, *Semantic Integration of Authoritative and Volunteered Geographic Information (VGI) using Ontologies*, Paper presented at 16th AGILE Conference on Geographic Information Science, 14–17 May.

- Sehra, S., Singh, J., and Rai, H., 2014, *A Systematic Study of OpenStreetMap Data Quality*
- *Assessment*, 11th International Conference on Information Technology: New Generations. PP. 377-381.

- Thakkar, S., K. Craig, and A. Jose , 2007, *Quality-Driven Geospatial Data Integration*, Proceeding of the 15th International Symposium on Advances in Geographic Information Systems.

- Vandecasteele, A., and Devillers, R., 2013, *Improving Volunteered Geographic Data Quality using Semantic Similarity Measurements*, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 8th International Symposium on Spatial Data Quality, Hong Kong, PP. 1-6

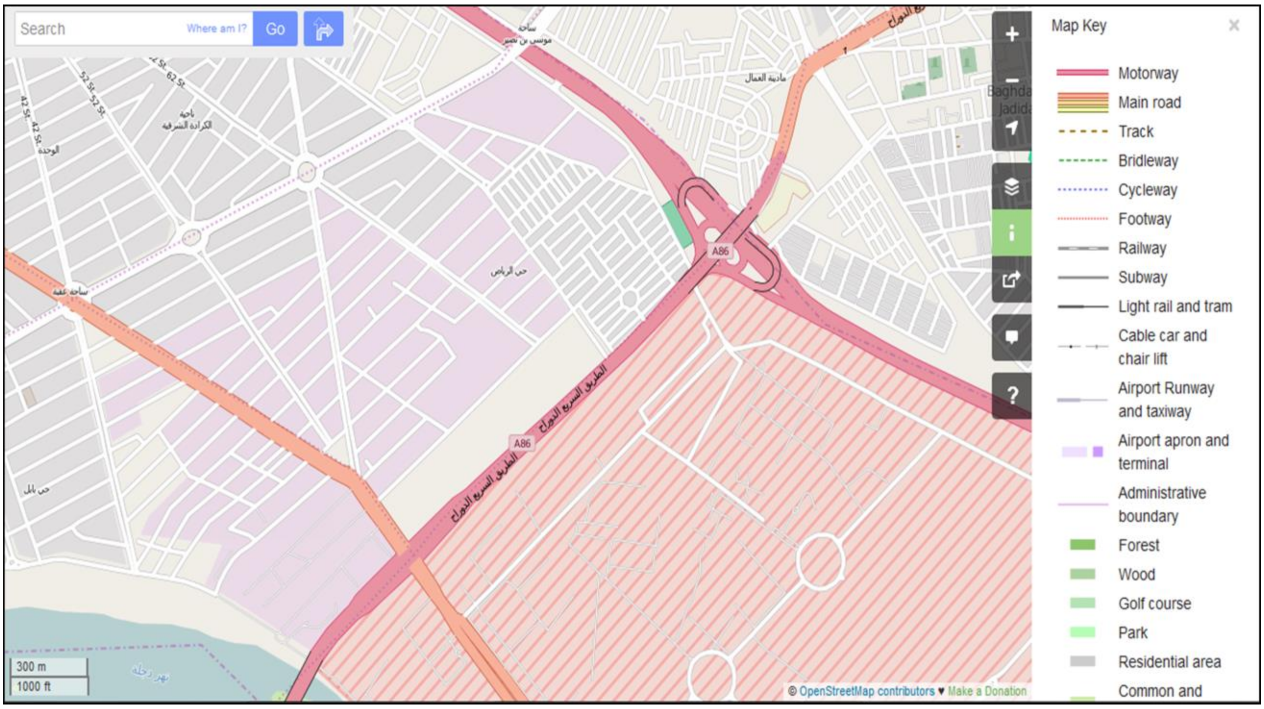

Figure 1. An example of feature classifications visualised on OSM project, **OpenStreetMap, 2015.**

		Reference Data			Row Total
		1	2	k	n_{i+}
Classified Data	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{k1}	n_{k2}	n_{kk}	n_{k+}
Column Total	n_{+j}	n_{+1}	n_{+2}	n_{+k}	n

Figure 2. Example of Error Matrix.

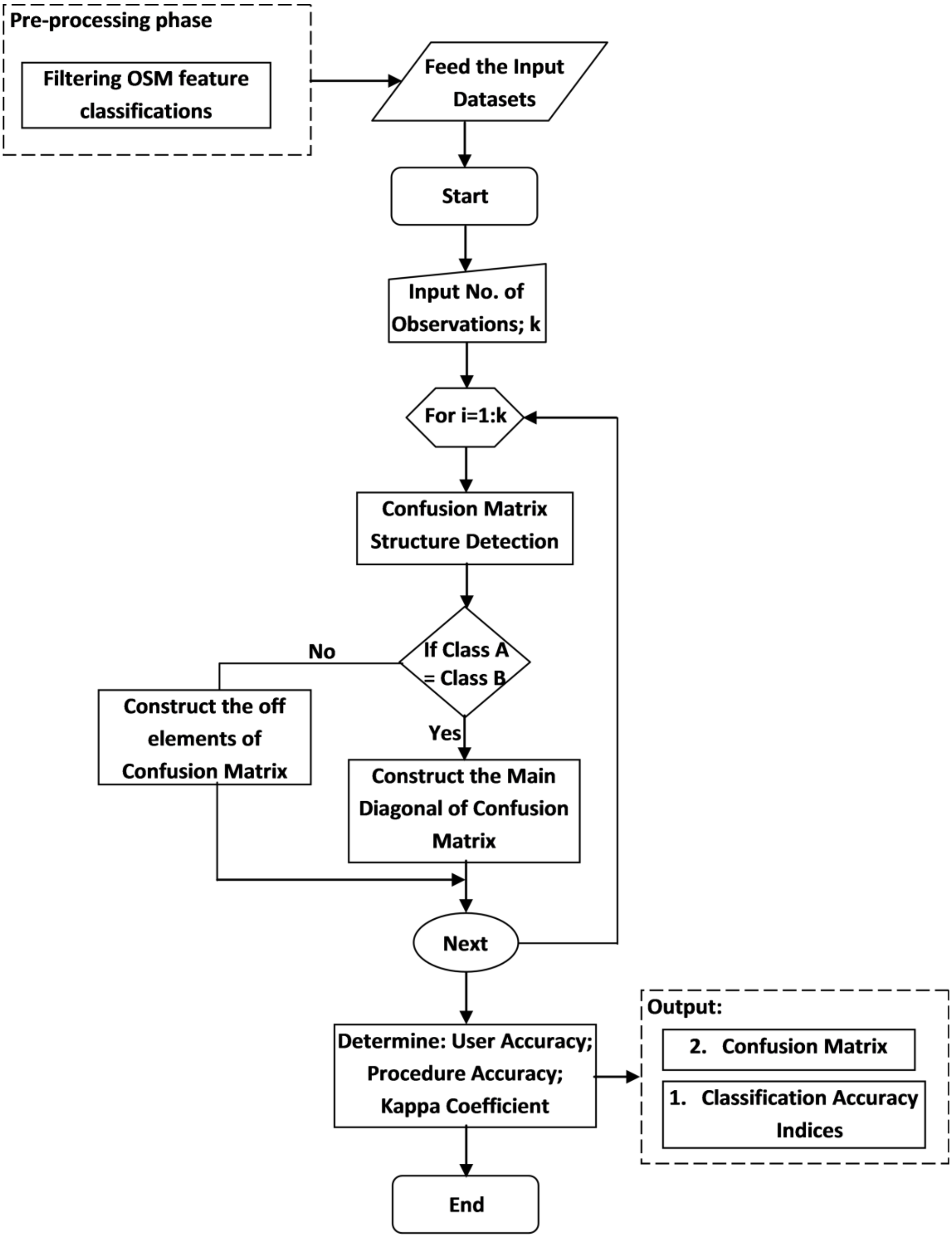

Figure 3. The Flowchart of the Designed Program.

		Reference Data																
		Primary road	**Secondary road**	**Service road**	**Track**	**Residential road**	**Bridge**	**Path**	**Building**	**River**	**Lake**	**Parking**	**Green area**	**Sports pitch**	**Residential area**	**University**	**Orchard**	**Row total**
OpenStreetMap Data	**Primary road**	**6**	2	2	0	0	0	1	0	0	0	0	0	0	0	0	0	**11**
	Secondary road	3	**10**	4	0	0	0	0	0	1	0	0	0	0	0	1	0	**19**
	Service road	0	0	**56**	0	0	0	0	7	0	0	0	0	0	0	2	1	**66**
	Track	0	0	0	**28**	0	0	0	0	0	0	0	1	0	0	0	0	**29**
	Residential road	2	0	0	0	**168**	0	0	0	0	0	0	0	0	0	0	0	**170**
	Bridge	0	0	1	0	0	**2**	0	0	0	0	0	0	0	0	0	0	**3**
	Path	0	0	0	0	0	0	**17**	5	0	0	0	0	0	0	0	0	**22**
	Building	0	1	14	0	0	0	0	**141**	0	0	2	0	0	0	6	0	**164**
	River	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	**1**
	Lake	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0	0	0	**3**
	Parking	0	0	0	0	0	0	0	0	0	0	**11**	0	0	0	0	0	**11**
	Green area	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	**2**
	Sports pitch	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	**2**
	Residential area	0	0	0	0	13	0	0	0	0	0	0	0	0	**6**	0	0	**19**
	University	0	0	0	0	0	0	0	4	0	0	0	0	0	0	**13**	0	**17**
	Orchard	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	**3**	**4**
	Column total	**11**	**13**	**77**	**28**	**181**	**2**	**18**	**157**	**2**	**3**	**13**	**3**	**2**	**7**	**22**	**4**	**543**

Figure 4. The Confusion Matrix.

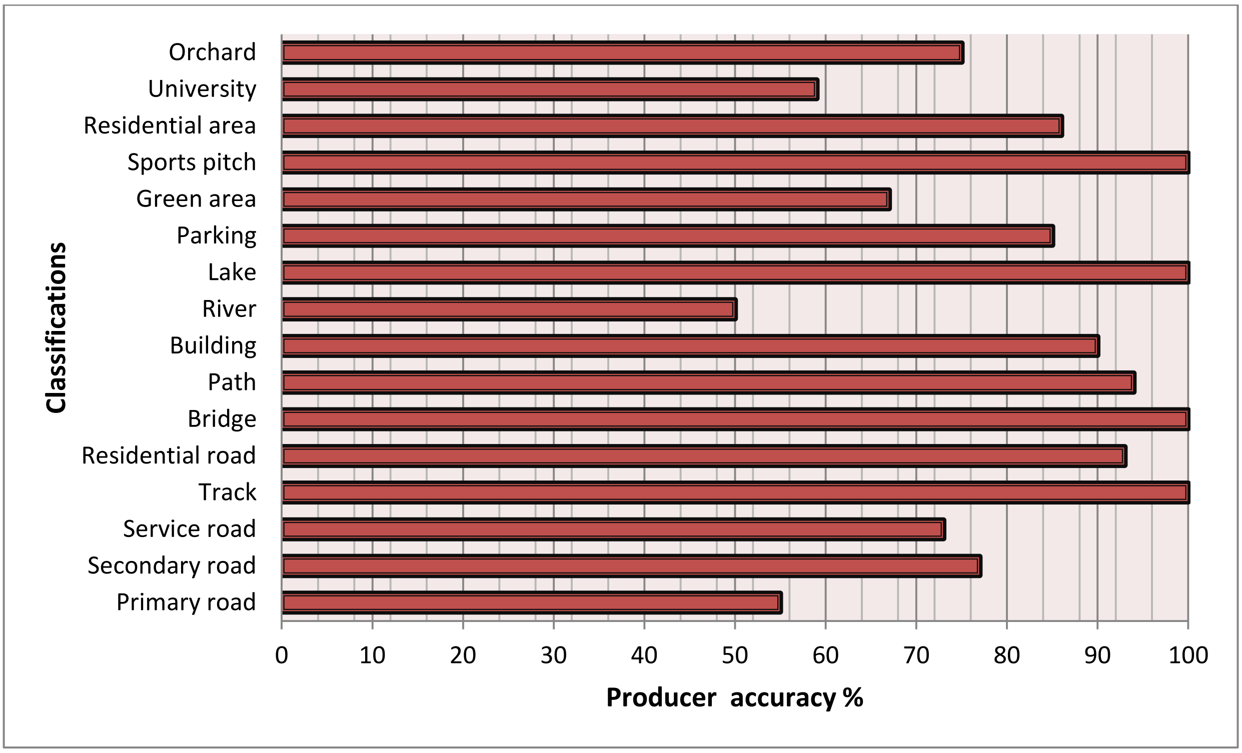

Figure 5. The Procedure Accuracy of OSM Classifications from 534 Reference Data.

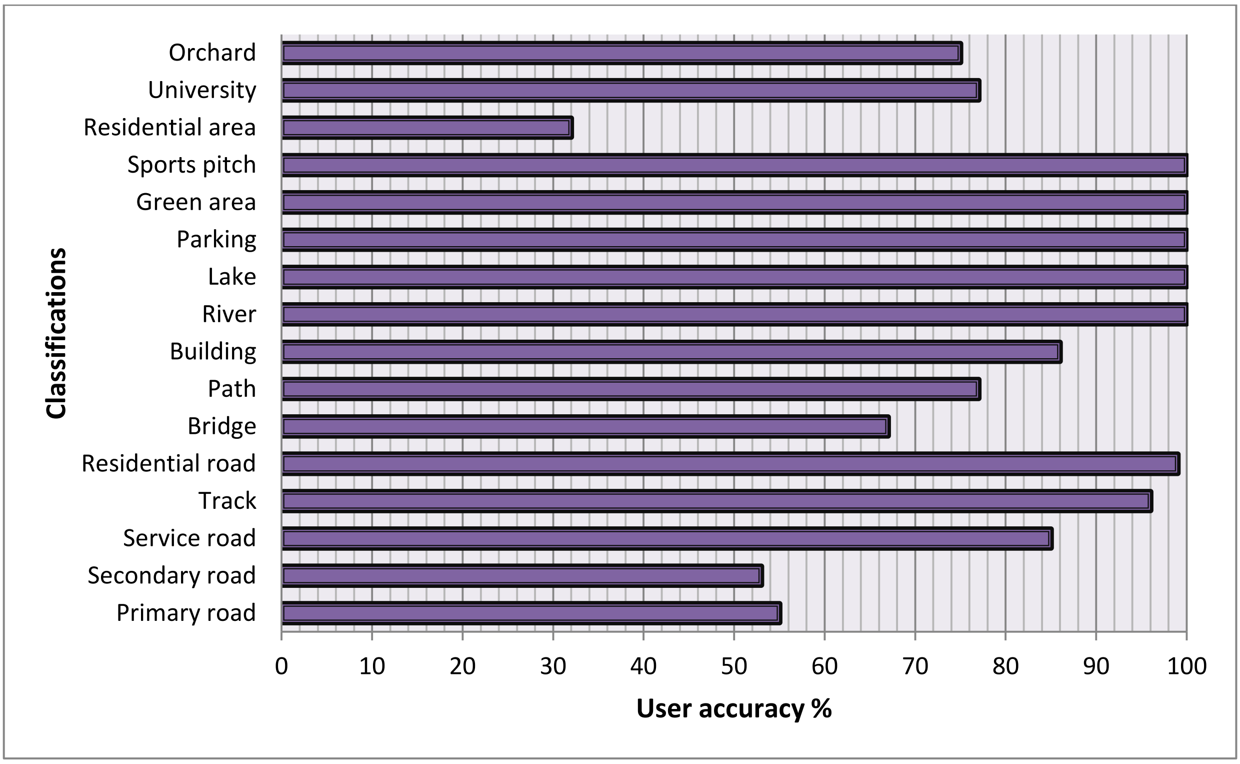

Figure 6. The User Accuracy of OSM Classifications from 534 Reference Data.