CHAPTER 9

Tackling the thematic accuracy of areal features in OpenStreetMap

Ahmed Loai Ali*,†

*Bremen Spatial Cognition Center (BSCC), University of Bremen, Germany
†Faculty of Computers and Information, Assiut University, Egypt
loai@informatik.uni-bremen.de, loai.cs@gmail.com

Abstract

With the increasing importance of VGI for GIScience, data quality becomes an issue of high concern. Particularly in collaborative mapping projects, when a group of public participants acts to collect, update and share information about geographic features, aiming to maintain and improve a geo-spatial dataset. OpenStreetMap (OSM) is the most common VGI project that aims to develop free world digital map. Although several studies emphasized the positional accuracy and completeness of the OSM data, particularly in the urban areas, they also highlighted its problematic thematic accuracy. In this chapter, we handle the thematic accuracy quality measure from the facet of classification. This chapter presents an approach for rule-guided classification for VGI projects. The proposed approach exploits the availability of data to learn the distinct characteristics of a set of geographic features. Afterwards, the learned characteristics are used to guide the contributors toward the most appropriate data classes, aiming to improve the data quality. The approach consists of two phases: *Learning* and *Guiding* phases. During the *Learning* phase, data mining algorithms are applied to learn the geographic characteristics of specific features. The learning process results in a set of rules describing these features. The extracted rules are used to develop a classifier. Afterwards, during the *Guiding* phase, the developed classifier is used for several purposes; 1) acts to detect

How to cite this book chapter:
Ali, A L. 2016. Tackling the thematic accuracy of areal features in OpenStreetMap. In: Capineri, C, Haklay, M, Huang, H, Antoniou, V, Kettunen, J, Ostermann, F and Purves, R. (eds.) *European Handbook of Crowdsourced Geographic Information*, Pp. 113–129. London: Ubiquity Press. DOI: http://dx.doi.org/10.5334/bax.i.

problematic classified entities; and 2) guides and aids the contributors during the classification process. An empirical study followed by an implementation is conducted. The results show the feasibility of the proposed approach and highlight some limitations that could be improved in the future studies. The developed tool generates promising results and improves the classification of OSM dataset as well.

Keywords

Volunteered Geographic Information (VGI), Spatial Data Quality, Thematic accuracy, Spatial data mining.

Introduction

Crowd-sourcing, the advance of web technologies and the availability of location sensing devices empower the public to produce contents associated with implicit or explicit spatial references. This form of User Generated Contents (UGC) has been known as *Volunteered Geographic Information*, in which a group of people voluntary acts to collect, update, and share spatial information (Goodchild 2007). VGI changes the conventional way of mapping activities resulting in collaborative mapping. Those activities were exclusively reserved – for a long time – for mapping agencies and specialized organizations. However, in collaborative mapping, participants are eager to collect information about geographic features producing maps (Gillavry 2004). Among others, OpenStreetMap[1] (OSM), Wikimapia[2] and Google MapMaker[3] are examples of collaborative mapping projects. OSM is the most prominent example of a VGI project; it aims to develop a free digital map of the world editable and available to everyone. During the last decade, several applications and services have been developed based on VGI data including – but not limited to – urban planning, environmental monitoring, crises management, map provision, etc.

Despite the increasing utilization of VGI data, its questionable quality still makes it – in some cases – of limited use (Elwood et al. 2012; Flanagin & Metzger 2008). Among other reasons, contributors' diversities and the fixable contribution mechanisms are resulting in data of heterogeneous quality (Mooney & Corcoran, 2012). Several studies assess VGI data by comparison with authoritative data sources. They conclude the promising completeness and positional accuracy of OSM data, particularly in urban areas (Haklay 2010; Neis et al. 2011). In Hecht and Stephens (2014), the authors highlight the declining of

[1] http://www.openstreetmap.org/
[2] http://www.wikimapia.org/
[3] http://http://www.google.com/mapmaker/

data quality with the increasing distance form urban areas. Regarding particular features, Girres and Touya (2010), Haklay and Weber (2008) and Ludwig et al. (2011) emphasize the quality of street networks in France, the UK and Germany respectively. Whereas Arsanjani et al. (2015) and Arsanjani and Vaz (2015) address the promising contributions of land use/land cover features in OSM datasets. The studies highlight the heterogeneous data quality not only regarding the positional accuracy and completeness quality measures, but also regarding the problematic thematic accuracy of data (Haklay 2010; Neis et al. 2011; Vandecasteele & Devillers 2013). Thematic accuracy implies correctness of the assigned classification to a given entity with that entity's characteristics and its geographic context. Hence, this chapter tackles VGI data quality from the *classification* perspective.

In OSM projects, the loose classification mechanisms lead to inappropriate classification of data. Whether a piece of land covered by grass is classified as *park*, *meadow* or *forest*, if a water body is classified as *pond* or *lake*, whether an area is classified from the land use perspective as *residential* or *industrial*, etc. All these classifications mainly depend on contributors' perception (*subjective classification*).

Otherwise, the appropriate classification should reflect the inherent geographic characteristics of an entity (*objective classification*). For example, *park* and *garden* are likely used for entertainment and should contain amusement facilities like a playground, sport area, etc. and a *lake* is likely surrounded by a natural landscape and some facilities like tracks or benches, and is larger in size than a *pond*; whereas *residential* areas mostly cover residential buildings and likely contain some residential services, whereas *industrial* areas usually have industrial properties like a company, factory, etc.

In this chapter, we propose a rule-guided classification approach. The approach aims to improve the data classification; it works to develop a recommendation system able to guide the contributors towards appropriate classification. The approach works to extract the distinct geographic characteristics of a specific feature and encode them in the form of rules. The rules are encoded together into a classifier. Afterwards, the developed classifier is applied to guide the contributors towards appropriate classifications.

As an empirical study, we address the classification of some grass-related features; where a piece of land covered by grass could be classified as *forest*, *garden*, *grass*, *meadow* or *park*. The classification of these features generates a challenge; they are commonly covered by grass, however each class has its distinct characteristics. For example, the *park* and *garden* classes have entertainment characteristics, the *forest* class are usually covered with trees or other woody vegetation and the *meadow* class has agricultural characteristics, etc. The findings are promising and show the feasibility of the approach.

This chapter is organized as follows: the 2nd section gives insights into the classification challenges in an OSM project, while the 3rd section presents the proposed approach of guided classification for VGI. An empirical study

is presented in the 4th section. The last section summarizes the findings and points out the future research directions.

Classification Challenges at OpenStreetMap

The OSM project is the most prominent collaborative mapping project: it covers most of the world, has more than 2 million registered users on October 2015[4] and the OSM data is utilized in various services and applications. However, its problematic classifications make its data of limited use (Devillers et al. 2010). In particular, the problematic classification results in inaccurate results and/or incomplete answers. The uncertainty, poor definitions and various individual conceptualizations of geographic features are other reasons behind the problematic classification of data (Fisher 1999; Grira et al. 2010). However, regarding the OSM project, the problematic data classification might come back to the following:

- **Contributors' heterogeneity**: the project harnesses the contributors' diversities to produce rich datasets. However, these diversities influence the resulting data quality (Coleman et al. 2009); contributors have various geographic and cartographic knowledge; this fact results in heterogeneous perceptions of the geographic features and consequently problematic classifications; what is perceived by a contributor as a *park* could be considered by another as a *grass* or *garden* type area.
- **Contribution methodologies**: OSM supports the contributors' heterogeneity by providing different methods of contribution. The most popular contribution methods are either by uploading GPS tracks directly or by editing geographic features over satellite images. The later method is the most common and is known as remote contribution. Figure 1 illustrates, a remote contribution (armchair contribution) process, in which a contributor uses an editor (e.g. JOSM) to contribute information about a specific feature by tracking the feature on satellite images. The contribution method itself generates a challenge during the classification process (Mooney & Corcoran 2012). For example, in Figure 1, the pieces of grass-covered land look similar and their classifications, whether *park, garden, grass* or *meadow*, mainly depend on the contributors perception and need some sense of locality. Moreover, the loose tagging mechanism of OSM also results in problematic classifications. There is no restriction on the number of tags associated with a certain entity; an entity could be associated with no tags or several tags with endless combinations without any integrity checking mechanism. For example, an entity could plausibly be tagged with *leisure=park, natural=grass, landuse=meadow* and *place=garden*.

[4] http://osmstats.neis-one.org/.

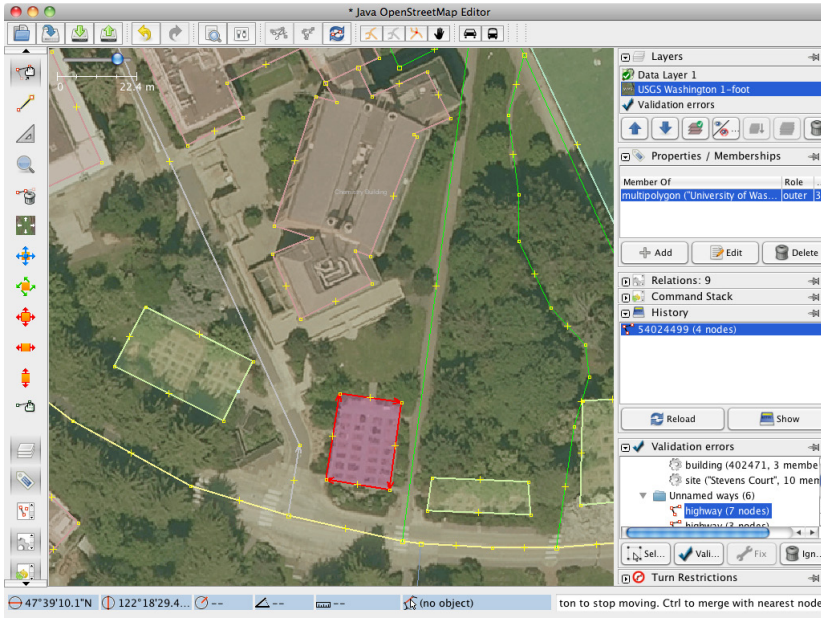


Figure 1: Remote contribution using JSOM editor.

- **Ambiguous recommendations**: the OSM project provides only recommendations for contributors through its Wiki[5] pages. These recommendations resulted from discussions between mappers. However, it is probable that most of the contributors do not spend enough time checking these recommendations. Furthermore, due to ambiguous terminologies (e.g. *wood* or *forest*, *landuse* or *landcover*, etc.), some recommendations might be conceptually misinterpreted, particularly by non-experts. For example, the unclear distinction between *lake* and *pond* classes results in a new class of *lake; pond*.

The previous points summarize the fundamental reasons behind the problematic data classification of OSM. These challenges come up due to the nature of VGI and the OSM project in particular. There exist other reasons due to the nature of geographic data as well. Most geographic features are not well defined; the fact that results in crisp boundaries between classes. In some cases, an identical feature could plausibly belong to multiple classes. However, small details usually exist and distinguish between conceptually overlapping classes. In the case of remote contribution, these details are hardly recognized

[5] http://wiki.openstreetmap.org/wiki/Map_Features

by armchair contributors, and consequently, they contribute either imprecise or incomplete data.

Rule-Based Guided Classification Approach

To tackle the classification challenges, we propose a rule-based guided classification approach. Through guiding and recommendations, the approach aims to produce data with appropriate and consistent classifications. The approach consists of two phases: *Leaning* and *Guiding* phases.

Learning Phase

During the *Learning* phase, the approach employs the increasing availability of OSM data in learning the characteristics that distinguish between similar classes. Figure 2 shows a summary of the learning phase. In this phase, the task is to develop a classifier able to distinguish between related classes. Data mining algorithms are used to find the distinct topological characteristics that distinguish between classes. The extracted characteristics have the form of predictive rules. Afterwards, the rules are integrated into a classifier. During the mining process, we depend on qualitative spatial analysis to find the characteristics of a specific class. Topology, direction and distance are the common qualitative spatial relations. In this work, we particularly investigate the topological relations to understand the geographic context of the given classes.

Topological Analysis Based on the first law of geography (Tobler 1970), nearby geographic features are related to each other. For example, the existence of sport's areas and playgrounds inside the *park* and *garden* features, the location of gas stations in a direct access to roads area, etc. Hence, in this work we investigate the topological relations between pairs of entities to find the characteristics that identify each class. Each entity is characterized by its interior and exterior context. At the same time, the *appropriate classification* should reflect the entities characteristics and matches its geographic context.

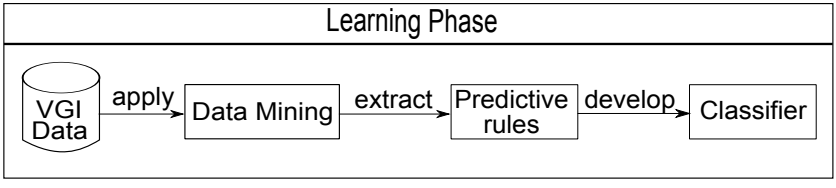


Figure 2: Learning phase of the proposed guided classification approach.

We utilize the 9-Intersection Model (9IM) (Egenhofer 1995), in which the topological relations between pairs of entities are defined as follows: *disjoint, meet, overlap, covers, coveredBy, contains, inside* and *equal*. In this work, *disjoin, meet, overlap, contains* and *coverdBy* relations are considered. While *inside, covers* and *equals* relations are neglected due to two reasons: (a) *inside* and *covers* are the inverse of *contains* and *coveredBy* relations respectively; and (b) the *equal* relation rarely occurs and does not add information to this analysis.

Mining Process The topological analysis aims to find frequent patterns (topological relations) involved between target classes and other geographic features, e.g. *park contains* playground, sport center, etc. Each tag is considered as a new feature. For example, *leisure = playground* and *leisure = sport* are treated differently. We encode them as leisure playground and leisure sport respectively and associate each one with a unique identifier (ID), to facilitate the mining process. The processing is computationally exhaustive and should be done in advance during the preparation for the mining task. Afterwards, the mining process works to extract atomic rules in the following form:

$$Class\ (E, C) \leftarrow R(E, F) \qquad (1)$$

where E represents an entity, $C \in$ {'*park*', '*meadow*', etc.}, R is one of the topological relations where $R \in$ {'*contains*', '*meet*', *etc.*} and F represents the set of frequent features that mostly involved in a relation R with entities of C.

We apply the Apriori algorithm (Agrawal et al. 1994) to extract the rules. In particular, we use the class association rules mining task, when rules have a predefined class (e.g. '*park*') as their outcome. Appropriate constraint parameters like *support* and *confidence* should be adjusted to extract and filter the interesting patterns. Afterwards, the extracted rules are integrated into a classifier.

Basically, developing a classifier based on a set of predictive rules consists of the following steps: (1) find all the interesting class association rules from a dataset; (2) filter the extracted rules into a set of predictive association rules; (3) encode the rules into a classifier; and (4) evaluate the classifier on a test dataset.

Guiding Phase

During the *Guiding* phase, the developed classifier could be used in many different scenarios. In this approach, we present three scenarios: the contributing, checking and enriching scenarios. Figure 3 gives brief illustrations of these scenarios as follows:

Contributing Scenario In this scenario, the classifier is embedded into an editing tool. At the contribution time, the classifier checks the validity of a given classification. In case of a problematic classification, the classifier informs the contributor of some recommendations. Then, according to

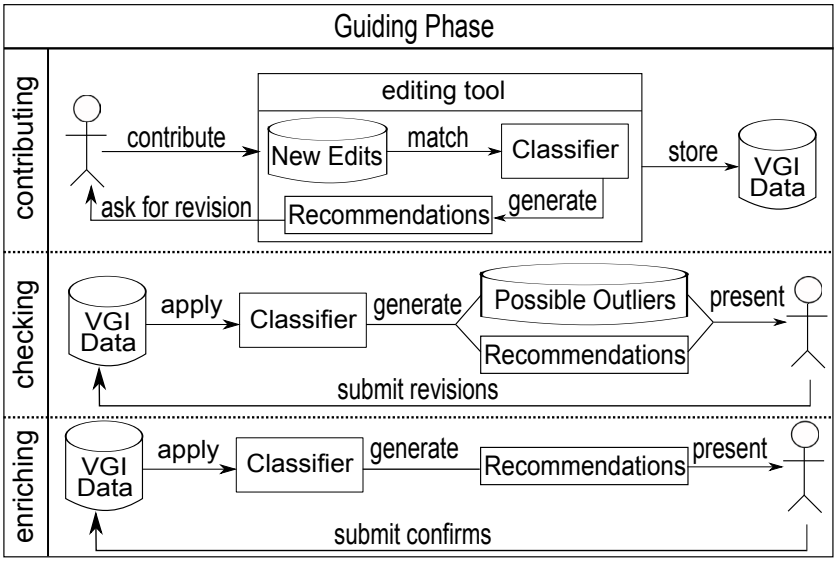


Figure 3: Guiding phase of the proposed guided classification approach.

the given recommendations, the contributor reacts with a correction (if required). The real challenge in this scenario is the computational complexity. During our study, the entities are processed in advance to investigate their geographic context. In contract, this scenario requires on-line processing of contributions at the time of editing.

Checking Scenario This scenario could be directly applied when the developed classifier is applied to an existing dataset generating the potential problematically classified entities. Afterwards, the outliers are presented – associated with recommendations – to crowd-sourcing revision. The contributors act to correct the problematic entities (if required).

Enriching Scenario In which the classifier is applied to a set of unclassified entities. The classifier predicts classifications for these entities and presents them for crowd-sourcing confirmation. The contributors' role here is to confirm the given classification and make corrections (if required). Another enrichment scenario could also be achieved, when the contributor reacts to add more information to satisfy the given recommendations.

In all of the proposed scenarios, the classifier cannot do automatic classification or automatic correction directly on the data source. However, it provides recommendations for directing contributors towards data of appropriate classification. At the same time, developing a global classifier might also be inaccurate. Therefore, the proposed approach is to maintain locality during both the *Learning* and *Guiding* phases. We assumed that a geographic feature should be classified identically, at least on the country level.

Empirical Study: Grass & Green

To evaluate the proposed approach, an empirical study and an implementation are conducted. The study aims to develop a classifier to distinguish between grass-related classes: *forest*, *garden*, *grass*, *meadow* and *park* classes. The classes are the most common grass-related classes within the boundaries of urban cities (the geographic scope of the research). The classification of these features represents a challenge due to the following: (1) in satellite images, they appear similar as a green area; (2) in some cases, a feature could plausibly belong to multiple classes (e.g. *park* and *garden*); and (3) for non-experts, they are all *grass*. Thus, a contributor might be unfamiliar with the characteristics that distinguish be- tween classes. Table 1 shows the OSM Wiki recommendations for these classes. The given recommendations are based on discussion between mapper communities. The given recommendations at Table 1 indicate that there exist unique characteristics that distinguish between classes.

Data Processing

We use an OSM dataset from Germany dated to December 2013. The choice of Germany comes from the following reasons: i) the existence of a large group of active mappers; and 2) several researchers have emphasized the quality of

Class	Recommendations
forest	Some use this tag for land primarily managed for timber production, others use it for woodland that is in some way maintained by humans.
garden	A distinguishable planned space, usually outdoors, set aside for the display, cultivation and enjoyment of plants and other forms of nature. It incorporates both natural and man-made materials. The most common form is known as a residential garden, it is a form of garden and is generally found in proximity to a residence, such as the front or back garden. Residential gardens are usually of human scale, as they are most often intended for private use.
grass	A tag for a smaller areas of mown and managed grass, for example in the middle of a roundabout or verges beside a road. Should not be used where a more specific tag is available.
meadow	Used to tag an area of meadow, which is an area of land primarily vegetated by grass plus other non-woody plants.
park	An area of open space provided for recreational use, usually designed and in a semi-natural state with grassy areas, trees and bushes. Parks are often but not always municipal.

Table 1: OSM recommendations for the target classes.

the data. We extract the entities from the 10 most densely populated cities[6] to ensure active mappers and hence a certain level of quality. The dataset consists of 3,724 *forest*, 3,030 *garden*, 7,336 *grass*, 4,277 *meadow* and 4,445 *park* entities. About 50% of the extracted entities have only one version (edits), which indicates the lower attraction of these entities to mappers. According to Mooney & Corcoran (2012), an increasing number of edits does not usually imply high quality. However, it reflects the heavy collaboration/competition among contributors to improve the data quality. The extracted entities are processed individually by checking the topological relations between each entity and other entities nearby.

Learning Process

During the learning process, the objective is to develop atomic rules per class per topological relation. Due to the uncertainty of spatial context, we take into account that everything is possible. Thus, a 1% *support* threshold is considered sufficient to extract the interesting patterns (frequent topological relations). Each topological relation is processed individually with a given class producing a set of predictive rules of that class. We extracted 8,504 rules: 4,100 describe *forest*, 215 describe *garden*, 745 describe *grass*, 506 describe *meadow* and 2,938 describe *park*. Although a large number of rules have a *confidence* threshold greater than 50%, the rules themselves represent some difficulties in the classification process due to: (1) they have a wide range of *confidence* threshold from 100% to 0.7 %; (2) due to the similarity between some classes, there exist duplicated rules pointing to different classes; and (3) regarding the topological relations, some relations have higher *confidence* thresholds than the others.

Classification Process

During the classification process, each entity is checked against all extracted rules. For example, Figure 4 shows an entity[7] with a *meadow* classification which has *osm_id* = 96279661. The entity matches 46 rules: 26 *park*, 6 *meadow*, 5 *forest*, 5 *garden* and 4 *grass*. Table 2 presents a sample of the matched rules for this entity. The figure illustrates that the entity contains a playground, sport areas and planned footways, which reflect the characteristics of the *park* class.

According to Table 2, considering the maximum *confidence* of the matched rules, the top 20 rules have *confidence* thresholds ranging form 92% to 80% and all of them have the result *Class*(*E,* '*park*'). At the same time, when considering the maximum *confidence* per class, this entity matches with *park*, *meadow*,

[6] http://www.citymayors.com/gratis/german_topcities.html

[7] http://www.openstreetmap.org/way/96279661, last accessed April 2015

Figure 4: A entity with osm_id=96279661 classified as 'meadow' (last visit at April 2015).

Rule — Confidence
Class (*E*, *'park'*) ← *contains* (*E*, [1,22,156])) – 92%
Class (*E*, *'park'*) ← *contains* (*E*, [1,15,22, 156])) – 91%
Class (*E*, *'park'*) ← *contains* (*E*, [15,21])) – 89%
Class (*E*, *'park'*) ← *contains* (*E*, [1,15])) – 88%
…
Class (*E*, *'park'*) ← *contains* (*E*, [22])) – 76%
Class (*E*, *'park'*) ← *contains* (*E*, [15])) – 66%
Class (*E*, *'meadow'*) ← *containsBy* (*E*, [128])) – 46%
Class (*E*, *'park'*) ← *meet* (*E*, [15])) – 34%
Where, 1=leisure_playground, 15=highway_footway, 21=sport_soccer, 22=leisure_pitch, 128=landuse_forest, 156=sport_basketball

Table 2: A sample of matched rules for the entity with osm_id=96279661.

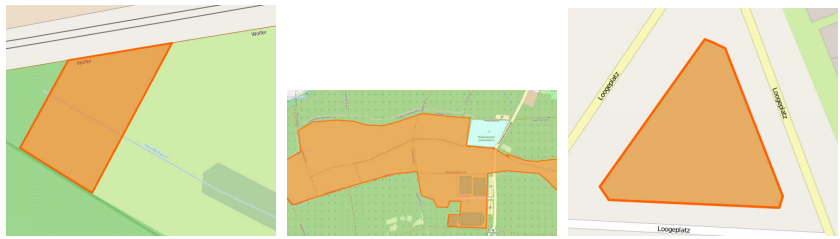
grass, *forest* and *garden* classes in descending *confidences* of 92%, 46%, 32%, 13% and 12%, respectively. Although the entity is currently classified as *meadow*, its characteristics make it more appropriate to be classified as a *park*. Hence, our recommendation works to guide contributors towards the most appropriate classification.

Evaluation Process

To evaluate the classifier, we do not have a ground-truth dataset for these entities. The available ground-truth datasets cover a higher classification level of land use or land cover. Thus, we depended on manual visual investigation to evaluate the results. Figure 5 presents examples of appropriate and inappropriate classifications, based on the developed classifier and recommendations.

Figure 5(a) gives examples of appropriate classifications. From left to right, the first entity is adjacent to residential houses and other gardens and does not contain much infrastructure. The entity is appropriately classified as *garden*. The second one, located between highways and containing nothing, is most likely to be classified as *grass*. The last entity contains a water body, sports centers, footways and other infrastructure. It is correctly classified as a *park*.

(a) Appropriate Classification of *garden*, *grass* and *park* classes



(b) Inappropriate Classification of *garden*, *grass* and *park* classes

Figure 5: Example of appropriate and inappropriate classifications.

In Figure 5(b), the classifier detects these entities as problematically classified entities. From left to right, the first entity is classified as *garden*. The entity meets *meadow* and is located near to a farmland. It does not inherit any plant or decoration characteristics. The classifier recommends *meadow* as an appropriate class. Whereas the middle entity is classified as *grass*, despite the fact that it seems too large, contains sports centers, is surrounded by forest areas and is adjacent to a playground. The classifier recommends *park* class for this entity. The entity on the right shows a clear example of inappropriate classification of *park*. The entity is located between roundabouts and does not contain any infrastructure at all. The classifier recommends it to be classified as *grass*.

Grass&Green: a quality assurance web tool

As another way to evaluate the proposed approach, we developed a web tool as a recommendation system called Grass&Green[8] as indicated in Figure 6. The tool presents the generated recommendations for crowd revisions as proposed in the checking scenario (see section 3.2). We created social media pages to attract the contributors for revisions: Facebook and Twitter. Moreover, we wrote OSM diaries to announce the tool to the OSM community. In this tool, the user logs in via his/her OSM account and contributes directly to the project. The tool presents entity by entity, combined with the recommended classes and

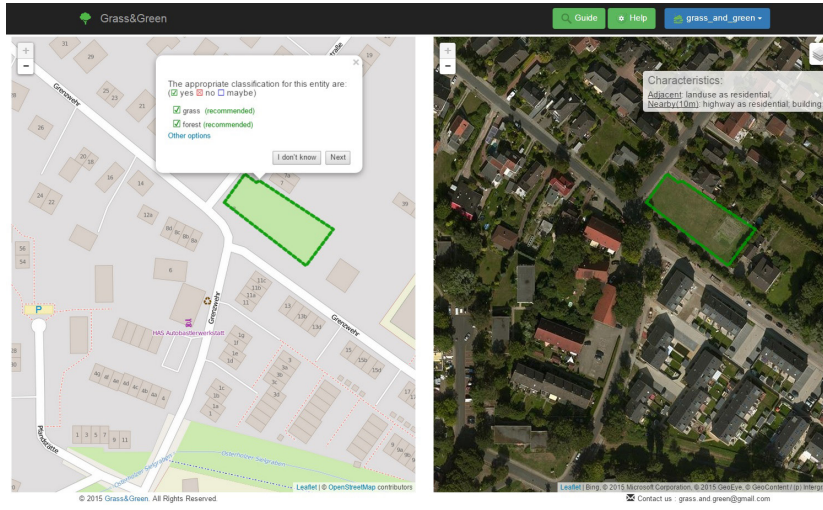


Figure 6: Grass&Green: the main contribution interface (last visit at September 2015).

[8] http://opensciencemap.org/quality

other potential classes. Due to the ambiguity of grass-related features, we provide the users with the two most recommended classes. The user has the ability to press an "I don't know" (inactive participation) in unclear cases. Moreover, the user could also change the recommendations or choose between (yes, no or maybe) in case of high ambiguity. The tool provides users with detailed textual and visual descriptions about the target classes and other grass-related classes.

Eleven days after launching the tool, we obtained promising results. We had around 80 users from various countries. They checked 560 entities: 485 active and 75 inactive participations. They agree with the generated recommendations as follows: 30.10% full agree, 60.84% partial agree and 9.05% disagree. The findings indicate the feasibility of the approach and the tool acts perfectly to improve the classification of OSM data.

Discussion and Conclusions

Conceptualization of geographic features has long been a topic of debate (Frank 1997). However, with the increasing role of public participants in collecting geospatial data, it becomes a crucial issue. GIS applications exploit VGI as an auxiliary data source. That means the data developed by public participants is used to provide services for others. A fact that raises more attention to the resulting data quality.

In particular, how do the participants perceive the space? How do they group and categorize the geographic features? How do they find the commonalities and differences between conceptually overlapping classes? All these questions might be addressed by utilizing the developed geospatial ontologies. Frank (1997) discussed the vital role of ontology in GIS applications, to achieve a better understanding of the space and to build more efficient information systems. For example, the OWL2 ontology that has been developed for structuring city information modeling with respect to land use mapping (Montenegro et al. 2012). OSMonto is another ontology, which has been developed to enrich the semantics of OSM tags, without correcting or modifying any conceptual mistakes in the taxonomy of OSM tags (Codescu et al. 2011). However, the link between ontologies' producers and consumers, in the GIS domain, still needs more research.

In VGI, the data is classified following the bottom-up approach; where the participants contribute data based on their local knowledge. They translate their observations into classes and categories. While in professional methods, the data is classified based on a top-down approach; where a pre-defined model is developed based on strict measures defining the classes. The difference of VGI approach leads to questionable data classification. Therefore, guiding amateur participants is needed for enhanced data classification. For example, designing intelligent data capturing interfaces is one possibility, among others, to support the contribution of enhanced data classification. This chapter

calls for the development of intuitive interfaces for VGI projects; negotiation, exemplifications and comparisons are human-centered approaches that could be used to support the VGI participants during the contribution process.

In this chapter, we addressed the VGI quality from a classification perspective. We investigated the classification correctness of an entity with respect to its inherent geographic characteristics. We proposed an approach for guided classification. The approach tackles the classification challenges in the OSM project by guiding the contributors during the classification process. The approach has two phases: the *Learning* and *Guiding* phases. During the *Learning* phase, the approach utilizes the OSM dataset to learn the distinct topological characteristics that distinguish between similar classes. Data mining algorithms have been used to develop a classifier. Afterwards, the developed classifier is used in different scenarios (contributing, checking and enriching) during the *Guiding* phase. The approach aims not only to improve the classification of data, but it could be used to enrich the data source as well.

We conducted visual investigations and an implementation to evaluate the proposed approach. We developed a classifier to distinguish among a set of grass-related classes. The selected classes have some similarity, but each one has its unique characteristics. The findings emphasized the feasibility of the proposed approach. The developed tool shows the positive response of the crowds towards the data quality. The presented results are preliminary indicators of an enhanced data classification. We will keep investigating the tool results and check the enhanced data classification in more details. In future work, the research would investigate how to generalize the developed classifier. In addition, the intuitive user interface would be studied to develop human-centered guided classification that corresponds to the nature of VGI.

Acknowledgments

We gratefully acknowledge support provided by the German Academic Exchange Service (DAAD), as well as the hosted research group in Bremen Spatial Cognition Center (BSCC) at University of Bremen. Furthermore, acknowledge for ICT COST Action IC1203 for Short Term Scientific Mission (STSM) support. We also thank all anonymous users of the developed tool (Grass&Green).

References

Agrawal, R., Srikant, R., et al. 1994. Fast algorithms for mining association rules. In: *Proc. 20th int. conf. very large data bases*, VLDB, Volume 1215: pp. 487–499.

Arsanjani, J. J., & Vaz, E. 2015. An assessment of a collaborative mapping approach for exploring land use patterns for several european metropo-

lises. *International Journal of Applied Earth Observation and Geoinformation, 35*(Part B): 329–337.

Arsanjani, J. J., Mooney, P., Zipf, A., & Schauss, A. 2015. Quality assessment of the contributed land use information from openstreetmap versus authoritative datasets. In OpenStreetMap in GIScience,. Springer: pp. 37–58.

Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., & Rau, R. 2011. OSMonto-an ontology of OpenStreetMap tags. State of the map Europe (SOTM-EU) 2011.

Coleman, D. J., Georgiadou, Y., Labonte, J., et al. 2009. Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research, 4*(1): 332–358.

Devillers, R., Stein, A., B′edard, Y., Chrisman, N., Fisher, P., & Shi, W. 2010. Thirty years of research on spatial data quality: achievements, failures, and opportunities. *Transactions in GIS, 14*(4): 387–400.

Egenhofer, M. J. 1995. On the equivalence of topological relations. *International Journal of Geographical Information Systems, 9*: 133–152.

Elwood, S., Goodchild, M. F., & Sui, D. Z. 2012. Researching Volunteered Geographic Information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers, 102*(3): 571–590.

Fisher, P. F. 1999. Models of uncertainty in spatial data. *Geographical information systems, 1*: 191–205.

Flanagin, A. J., & Metzger, M. J. 2008. The credibility of Volunteered Geographic Information. *GeoJournal, 72*(3–4): 137–148.

Frank, A. U. 1997. Spatial ontology: A geographical information point of view. In: *Spatial and temporal reasoning*. Springer: pp. 135–153.

Gillavry, E. M. 2004. Collaborative Mapping: By the People, for the People. *Society of Cartographers Bulletin, 37*(2): 43–45.

Girres, J.-F., & Touya, G. 2010. Quality assessment of the french Open- StreetMap dataset. *Transactions in GIS, 14*(4): 435–459.

Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*(4): 211–221.

Grira, J., B′edard, Y., & Roche, S. 2010. Spatial data uncertainty in the VGI world: Going from consumer to producer. *Geomatica, 64*(1): 61–72.

Haklay, M. 2010. How good is Volunteered Geographic Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning. B, Planning & design, 37*(4): 682.

Haklay, M., & Weber, P. 2008. OpenStreetMap: User-generated street maps. *Pervasive Computing, IEEE, 7*(4): 12–18.

Hecht, B., & Stephens, M. 2014. A tale of cities: Urban biases in Volunteered Geographic Information.

Ludwig, I., Voss, A., & Krause-Traudes, M. 2011. A comparison of the street networks of Navteq and OSM in Germany. In: *Advancing Geoinformation Science for a Changing World*. Springer: pp. 65–84.

Montenegro, N., Gomes, J. C., Urbano, P., & Duarte, J. P. 2012. A land use planning ontology: Lbcs. *Future Internet, 4*(1): 65–82.

Mooney, P., & Corcoran, P. 2012. The annotation process in OpenStreetMap. *Transactions in GIS, 16*(4): 561–579.

Neis, P., Zielstra, D., & Zipf, A. 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet, 4*(1): 1–21.

Tobler, W.R. 1970. A computer movie simulating urban growth in the detroit region. *Economic geography, 46*: 234–240.

Vandecasteele, A., & Devillers, R. 2013. Improving volunteered geographic data quality using semantic similarity measurements. ISPRS-International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences, 1*(1): 143–148.