Conference Paper

Map Change Prediction for Quality Assurance

Author(s):
Homburg, Timo; Boochs, Frank; Cruz, Christophe; Roxin, Ana-Maria

ETH Library

Map Change Prediction for Quality Assurance

Timo Homburg*, Frank Boochs*, Christophe Cruz**, Ana-Maria Roxin**

* Mainz University Of Applied Sciences, Lucy-Hillebrand Straße 2, 55212 Mainz, Germany, timo.homburg@hs-mainz.de, frank.boochs@hs-mainz.de
** Université de Bourgogne, 9 avenue Alain Savary, 21000 Dijon, France, ana-maria.roxin@u-bourgogne.fr, christophe.cruz@u-bourgogne.fr

Abstract. Open geospatial datasources like OpenStreetMap are created by a community of mappers of different experience and with different equipment available. It is therefore important to assess the quality of Open-StreetMap-like maps to give recommendations for users in which situations a map is suitable for their needs. In this work we want to use already defined ways to assess the quality of geospatial data and apply them to a Machine Learning algorithm to classify which areas are likely to change in future revisions of the map. In a next step we intend to qualify the changes detected by the algorithm and try to find causes of the changes being tracked.

1.   Introduction

The task of integrating heterogenous geospatial data into the Semantic Web has been getting more attention in the recent years. Homburg et al. (2016) described a process of automatic integration and interpretation of such data for specific kinds of geospatial data. Prudhomme et al. (2017) proposed an automatic way to integrate a subset of geospatial data into the Semantic Web using methods of Natural Language Processing and through the analysis of data column values. In the context of our research project SemanticGIS[1] we create a knowledge base for disaster management purposes using among others the aforementioned approaches to automatically integrate geospatial data. Having created a sufficiently large knowledge base of a certain area, the focus of our research lies on how to ensure that the given geospatial data fulfils quality requirements set by individual tasks to be execut-

---

[1] https://i3mainz.hs-mainz.de/de/projekte/semanticgis

ed. Our particular interest is to verify that open geospatial data is as suited to be used in a disaster management than more traditional data sources e.g. from governments.

1.1. Data Quality in Disaster Management

A typical usecase in our project involves the need for detailed plans of areas in order for rescue forces to navigate to their destinations as soon as possible and to take necessary precautions. Ideal data from the perspective of several rescue forces we talked to is usually governmental data, as these datasets are authored by professionals of the field using professional equipment and often include catasteral data to enrich said datasets. Due to legal problems and/or monetary restrictions, rescue forces cannot always get official governmental data for specific areas and have to rely on open data e.g. OpenStreetMap and/or Google Maps instead. Such maps are of questionable quality because they are edited by a community of individuals with unknown expertise and equipment, but also usually differ in currentness, as official datasets are only updated in yearly intervals, whereas OpenStreetMap data can receive updates at any point of time at any place in the map. The quality assessment task therefore needs to estimate in which areas the current revision of OpenStreetMap is likely to deviate from a future revision of governmental data and which changes are to be expected. In a next step the results of this analysis can be used to indicate which areas on OpenStreetMap are trustworthy enough and meet the quality expectations of rescue forces.

1.2. Related work on Geospatial Data Quality

Geospatial data quality has been defined in various works before. For the sake of this publication we distinguish instrinsic and extrinsic data quality metrics. Intrinsic quality metrics can be applied on geometries without a reference goldstandard geometry present. Examples include the validity of the geometry and metrics which include previous versions of the geometry e.g. to track the frequency of geometry changes over time. Extrinsic data quality metrics rely on a goldstandard dataset of equivalent geometries to compare to. We consider the data of the authorities as our goldstandard data. Examples of such metrics include Positional Accuracy and Shape Similarity algorithms (e.g. Haussdorff distance). Quality Analysis of OpenStreetMap has been done in various aspects by many authors before. Ramm et al. (2011) gives a comprehensive introduction to OpenStreetMap. Haklay and Weber (2008) conducted a first quality analysis of OpenStreetMap roads in England. Mooney et al. (2010) proposed metrics to compare OpenStreetMap data to a reference dataset by applying feature tag analysis, source tag analysis, coverage analysis and a ground-truth comparison with data about Ireland. In a similar fashion Neis et al. (2010) conducted a com-

parative analysis on the German road network with official street atlas data. Fan et al. (2014) examined a ground-truth comparison of OpenStreetMap to catasteral data of the city of Munich in Germany. Positional accuracy, Semantic accuracy for the purpose of identifying a geometry uniquely, completeness in the sense of geometrical differences between the geometries and shape accuracy as a similarity distance measure between the selected geometries. First work in the direction of metadata has been done in 2012 when Neis and Zipf (2012) showed, that it is possible to create comprehensive statistics of OpenStreetMap user behavior, including registration date, number of edits, possible area of expertise and preferred edit behaviors. Further publications on OpenStreetMap data quality include Basiri et al. (2016), who tried to evaluate OpenStreetMap roads by using tracking data of users. Possibilities of an intrinsic analysis of OpenStreetMap without the help of a reference data source have been examined by Barron et al. (2013). In this research automatic validation procedures for OpenStreetMap data are mentioned on which this work will build up. In 2014, Neis developed a comprehensive framework for OpenStreetMap quality evaluation (Barron et al., 2014a), also resulting in a tool for intrinsic analysis of OpenStreetMap data iOSMAnalyzer (Barron et al., 2014b).

2.   Experimental Setup

Our experimental setup considers a test area in Cologne/Germany, for which we extracted building information (building footprints, address data and Semantic attributes) about every building from official data govdata_rev provided by the government of NorthRhine Westphalia/Germany. For every such geometry govdata_geom, our first task is to find one or more corresponding geometry matches in OpenStreetMap. We achieve this by matching a reasonably sized bounding box of the centroid of each gov_geom with corresponding OpenStreetMap geometries. We verify the results by applying several similarity metrics like the intersection area, the geometry type, the area similarity and the Haussdorff distance of the concerned geometries as well as using Semantic matching algorithms proposed in Prudhomme et al. (2017). We end up with a set of geometries osm_geom which we consider as equivalent geometries to the official governmental data. In a first stage, we begin to classify if a geometry is likely to change in the upcoming iteration of governmental data. In our general setup (see Figure 1) we are comparing a snapshot of osm_geom, od_rev_1 to a current revision of governmental data govdata_rev_1 and its previous revisions in order to predict the changes in govdata_rev_2.

To compare matched geometries the following dimensions of data quality are considered both in instrinsic and extrinsic measures if possible:

- Geometry Data Quality (e.g. Validity, Positional Accuracy)

- Attribute Data Quality (e.g. Completeness)

- User Metadata Quality (e.g. NumberOfEdits, Number of LastModifications, Mapping Experience)

- DataSet Metadata Quality (e.g. Data Change Rate)

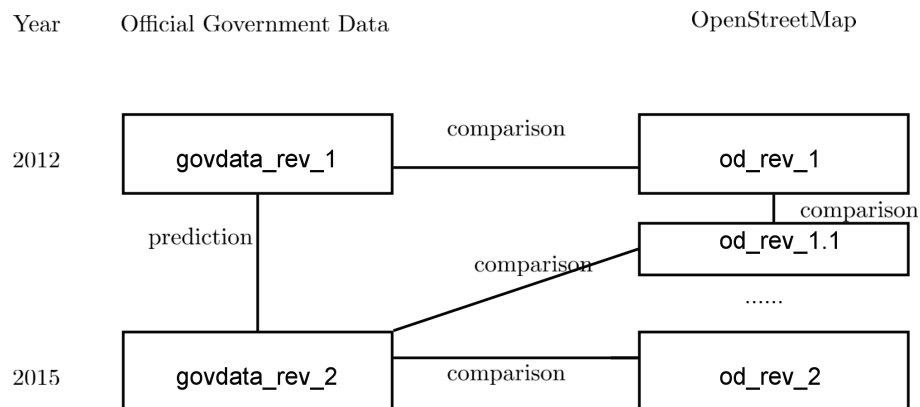- Service Metadata Quality (e.g. Availability of Service, License)

Figure 1. Data revisions

2.1. Preliminary Results

In a first experiment we tried to classify which geometries are likely to change in the next revision of the map in a district of Cologne/Germany consisting of about 1700 geometries. This is a first step into further categorizing changes which are likely to occur in a future work of this project. We applied an SVM classification algorithm and a Random Forest classification algorithm on a feature set of 37 features with the classes ("WillChange"/"WillNotChange") per geometry from the above mentioned data quality dimensions and give preliminary results in table 1.

We found that for the purpose of classifying changes, the following features were most significant (using a CorrelationAttributeEvaluation algorithm):

- HaussdorffSimilarity to Reference Geometry (47%)

- Average Amount of LastGeometry Edits of Editing Users (44%)

- Distance to Reference Geometry (43%)

- Average UserExperienceScore (comprised of MappingDays,User RegistrationTime,

- UserNumberOfEdits) (45%)

Algorithm	Precision	Recall	F-Measure	Correctly Classified
Random Forest	92.0%	90.0%	95.2%	95.23%
SVM	87.6%	87.1%	86.3%	87.1%

Table 1. Preliminary Results

2.2. Interpretation of preliminary results

The goal of our preliminary experiment was to find out features which would help us to identify possible future changes in OpenStreetMap, possibly leading to better or worse quality of the map data for the tasks we need to solve in our project. Firstly we predicted any change which is likely to happen on OpenStreetMap during the next iteration of governmental data and found out that using the significant features listed above we could predict a change behavior in about 90% of cases on the limited dataset we worked on in the city of Cologne. We also analyzed the results to exclude cases of overfitting. Our observation was that in the case given, a Random-Forest classification performed better than an SVM classification which we think might also provide better results when predicting the nature of the changes using a more sophisticated approach. In our opinion our findings indicate, that our approach to predict changes in OpenStreetMap data can lead to promising results in the area of building footprints, but needs to be thouroughly tested in the case of other geometries like LineStrings (e.g. roads, railways and canals) as well as in different regions and using different time frames. We intend to diversify our featureset to include more Semantic features to classify with. Such features may include the usage of the building represented as a Semantic Web concept and its purposes in the context of disaster management we are targeting to solve in our research project. In addition we would like to consider interrelations between geometries on the map leading to a variety of metrics comparing the currently observed geometry to its neighbouring geometries in terms of e.g. freshness, classification. We will continue to explore other useful features in future work in order to focus on more precise prediction cases and welcome the opportunity to initiate a discussion among the research community on what they believe to be a good prediction algorithm for map data quality.

3.   Conclusion and Future Work

In this submission we introduced our approach to predict data quality developments on OpenStreetMap data using machine learning algorithms. In this early stage our goal was to classify changes of geometries and features in order to achieve a change prediction map of a future revision of governmental data. For disaster management this can give us a hint to know which geometries in OpenStreetMap are likely to be unchanged and therefore keep their quality level. We found that in our experiment a change prediction for a geometry in OpenStreetMap results in about 90% correctly classified instances can be made and could identify features that will help us to further refine our prediction algorithm. The next step in our research will be to qualify the changes we are predicting using the Machine Learning algorithms. We would like to classify if the quality of the changing geometries in OpenStreetMap is likely to increase or decrease based on previous classifications on training data of the algorithm and welcome ideas and feedback on this matter by the research community. Classifying map changes will therefore help us to identify areas in a disaster management area which can be considered as safe for the operations required.

4.   Acknowledgements

References

Barron, C., Neis, P., Zipf, A.: Towards intrinsic Quality Analysis of OpenStreetMap Datasets. AGILE 2013 – Workshop: Action and Interaction in Volunteered Geographic Information (ACTIVITY), 14.-17. May 2013, Leuven, Belgium (2013)

Barron, C., Neis, P., Zipf, A.: A comprehensive framework for intrinsic openstreetmap quality analysis. Transactions in GIS 18(6), 877–895 (2014)

Barron, C., Neis, P., Zipf, A.: iosmanalyzer-ein werkzeug für intrinsische osm-qualitätsuntersuchungen. (none) (2014)

Basiri, A., Jackson, M., Amirian, P., Pourabdollah, A., Sester, M., Winstanley, A., Moore, T., Zhang, L.: Quality assessment of openstreetmap data using trajectory mining. Geo-spatial information science 19(1), 56–68 (2016)

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

Fan, H., Zipf, A., Fu, Q., Neis, P.: Quality assessment for building footprints data on open-streetmap. International Journal of Geographical Information Science 28(4), 700–719 (2014)

Homburg, T., Prudhomme, C., Würriehausen, F., Karmacharya, A., Boochs, F., Roxin, A., Cruz, C.: Interpreting heterogeneous geospatial data using semantic web technologies. In: International Conference on Computational Science and Its Applications. pp. 240–255. Springer (2016)

Mooney, P., Corcoran, P., Winstanley, A.C.: Towards quality metrics for open- streetmap. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. pp. 514–517. ACM (2010)

Neis, P., Zielstra, D., Zipf, A., Strunck, A.: Empirische untersuchungen zur datenqualität von openstreetmap-erfahrungen aus zwei jahren betrieb mehrerer osm-online-dienste. Angewandte Geoinformatik 2010 (2010)

Neis, P., Zipf, A.: Analyzing the contributor activity of a volunteered geographic information project—the case of openstreetmap. ISPRS International Journal of Geo-Information 1(2), 146–165 (2012)

Prudhomme, C., Homburg, T., Jean-Jacques, P., Boochs, F., Roxin, A., Cruz, C.: Automatic integration of spatial data into the semantic web. In: WebIST 2017 (2017)

Ramm, F., Topf, J., Chilton, S.: OpenStreetMap: using and enhancing the free map of the world. UIT Cambridge Cambridge (2011)