

A VERSION-SIMILARITY BASED TRUST DEGREE COMPUTATION MODEL FOR CROWDSOURCING GEOGRAPHIC DATA

Xiaoguang Zhou, Yijiang Zhao

School of Geosciences and Info-physics, Central South University, Changsha 410083, China; zxgcsu@foxmail.com

Commission II, WG II/4

KEY WORDS: Crowdsourcing Geographic Data; Trust Degree; Reputation; Version Similarity; OpenStreetMap

ABSTRACT:

Quality evaluation and control has become the main concern of VGI. In this paper, trust is used as a proxy of VGI quality, a version-similarity based trust degree computation model for crowdsourcing geographic data is presented. This model is based on the assumption that the quality of VGI objects mainly determined by the professional skill and integrity (called reputation in this paper), and the reputation of the contributor is movable. The contributor's reputation is calculated using the similarity degree among the multi-versions for the same entity state. The trust degree of VGI object is determined by the trust degree of its previous version, the reputation of the last contributor and the modification proportion. In order to verify this presented model, a prototype system for computing the trust degree of VGI objects is developed by programming with Visual C# 2010. The historical data of Berlin of OpenStreetMap (OSM) are employed for experiments. The experimental results demonstrate that the quality of crowdsourcing geographic data is highly positive correlation with its trustworthiness. As the evaluation is based on version-similarity, not based on the direct subjective evaluation among users, the evaluation result is objective. Furthermore, as the movability property of the contributors' reputation is used in this presented method, our method has a higher assessment coverage than the existing methods.

1. INTRODUCTION

Crowdsourcing geographic data, or Volunteered Geographic Information (VGI), is developing fast in the recent years. It has been proven to be a highly successful means of acquiring timely and detailed global spatial data at very low cost. However, VGI is voluntarily produced by amateurs (or 'neogeographers') without strict regulation or formal training. In VGI database, there is usually a lot of spurious or low-quality data. This limitation impact its fitness for use. In order to clean the spurious or low-quality objects, and use the high-quality objects, it is needed to evaluate the quality of each objects of VGI database. However, unlike professional geographic data, which are collected by trained specialists with specialized standards who guarantee the reliability of the data, the quality of the same professional spatio-temporal data set (or scale) is almost at the same level. VGI is collected by non-professional users without specialized training, the quality of VGI objects is very different, even in the same city, same system. Therefore, before it can be used in scientific analysis, it is necessary to use some reliability measures to clean or filter the spurious or low-quality data. Therefore, quality evaluation and control became the main concern of VGI.

Many efforts have been made to study the quality of VGI, including quality evaluation of VGI data and quality assuring method for VGI data. On the methods for assuring VGI quality, Goodchild & Li (2012) described three approaches for VGI quality assurance, i.e., crowd-sourcing eyes, social hierarchical gate-keeper, and geographic knowledge approaches. Matyas et al (2011) presented an accuracy improvement method in location-Based Games.

Methods for evaluating the quality of VGI can be generally divided into two categories. For the first category, VGI quality

is evaluated by comparing OSM data with corresponding professional data. For example, Haklay (2010) examined the data quality of London and England through a comparison with Ordnance Survey (OS) datasets. Zielstra and Zipf (2010) analyzed the completeness of OSM data in comparison to the navigation data of the TeleAtlas MultiNet datasets in Germany. Girres and Touya (2010) completed a quality assessment of France OSM spatial data using the Large Scale Reference database (RGE) for reference data and a sampling method using the assessment components, i.e., geometric accuracy, attribute accuracy, completeness, logical consistency, semantic accuracy, temporal accuracy, lineage, and usage. Cipeluch et al (2010) described a comparison of the accuracy of Ireland OSM data with Google Maps and Bing Maps. Siebritz and Sithole (2012) evaluated the quality of OSM data in South Africa by comparing them with a reference data set from National Mapping Agencies. Forghani & Delavar (2014) Evaluated the consistency between OpenStreetMap Dataset of Tehran, Iran and the corresponding reference geospatial dataset. Jackson et al (2014) Assessed the completeness and spatial error of features (using school campus extent as example) in the United States (US). Hecht et al (2013) Analyzed the completeness of building footprints in OpenStreetMap by comparing the OSM data with the official data in Germany; Fan et al (2014) have evaluated the quality of OSM building footprint data in terms of completeness, semantic, position and shape accuracy using ATKIS data as reference data. Comber et al (2013) evaluated the reliability of volunteered land cover using GLC-2000, GlobCover and MODIS V5 as control data. For the second category, trust is used as a proxy of VGI quality. For example, Bishr & Mantelas (2008) proposed a formal trust and reputation model using rating value and a distance measure to filter and extract high quality information on urban growth behaviours on the part of actors (people); Keßler & de Groot (2013) used trust as a proxy measure for the quality of VGI and measured VGI

trustworthiness based on five indicators, i.e., VGI contributors, versions, confirmations, corrections and revisions. Inspired by the work of Keller & de Groot (2013), D’Antonio et al. (2014) presented a data trustworthiness and user reputation computation model whereby the trust of VGI was evaluated based on direct, indirect and temporal effects and whereby user reputation could be computed based on the trust value of objects created by a user. However, the above trust assessment methods are preliminary realization of the “many eyes principle”, and suffer from highly depending on the revision versions. But from statistics, more than 60% objects in OSM just have one version, therefore the trust of these one-version-objects can not be computed by the above methods.

According to our observations, the quality of the crowdsourcing geographic objects is determined by the professional skill and integrity (called reputation together in this paper) of its contributor. The reputation of the contributor can be improved in the contribution process. Thus the reputation of the contributor is dynamically and movable, i.e., it is reasonable to think that the objects contributed by a same contributor in a not long term with the same quality (called trust degree in this paper). The reputation of the contributors cannot be computed directly. But the quality of the crowdsourcing objects yielded by the contributors can be validated or evaluated by the other users. Therefore, the reputation of the contributor can be yielded from his (or her) contributed objects, the similarity degree among the different versions of a same entity state implicit the reputation evaluation among the contributors, and the reputation value can be move to the other objects contributed by the same user. Therefore, the trust degree of the most VGI objects can be evaluated in this way. Based on this observation, a version-similarity based trust degree computation model for crowdsourcing geographic data is presented in this paper.

The remainder of the paper is structured as follows. In section 2, the forming mechanism of the VGI object version is analysed, and the concept of version-similarity is introduced. In Section 3, the computation of contributor’s reputation is discussed. The version-similarity-based trust degree computation model for crowdsourcing geographic data is described in Section 4. Experimental tests of this study are presented in Section 5. Finally, Section 6 provides a summary and concludes the discussion.

2. INTRODUCTION OF VERSION-SIMILARITY

In order to find the impact mechanism of VGI objects’ quality, we analyzed the formation process of VGI object version. In traditional geographic information system, one version is corresponding to one state, updates are made only due to state change; However in a VGI context, one state of an entity may have several versions, updates can be made due to two main reasons (Zhao & Zhou, 2016), i.e., a real entity’s state change; or to improve the quality of VGI object (the state of a corresponding geographic entity in the real world does not change), as Figure 1 shows. In Figure 1, object A has m states, i.e., S_1, S_2, S_j, S_m ; S_j has five versions (i.e., $V_{i-2}, V_{i-1}, V_i, V_{i+1}, V_{i+2}$) contributed by five contributors (i.e., $C_{j-3}, C_{j-2}, C_{j-1}, C_j, C_{j+1}$). As $V_{i-2}, V_{i-1}, V_i, V_{i+1}, V_{i+2}$ are corresponding to the same state, in fact there are four repeated observations. The repeated data can be used as VGI data quality evaluation, i.e., there are

implicit evaluations among the contributors, i.e., $C_{j-3}, C_{j-2}, C_{j-1}, C_j, C_{j+1}$. While, as V_1, V_m , and $V_{i-2}, V_{i-1}, V_i, V_{i+1}, V_{i+2}$ are corresponding to different states, the states are not comparable, there is no evaluation relation among C_g, C_n , and $C_{j-3}, C_{j-2}, C_{j-1}, C_j, C_{j+1}$.

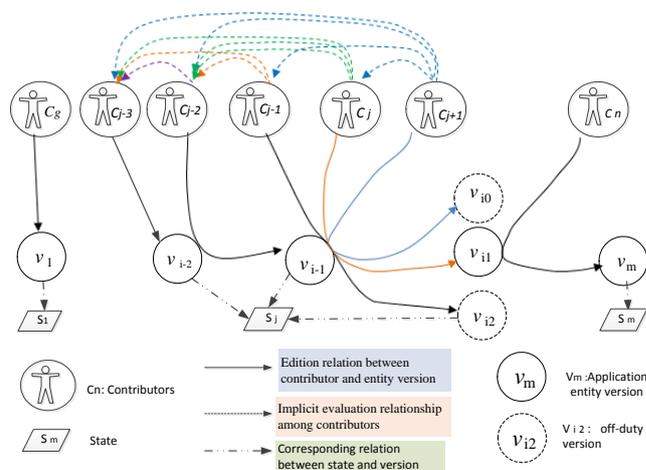


Figure 1. the forming mechanism of the VGI object version

From Figure 1, we can conclude that there are implicit evaluations among the contributors who yield the different versions for the same entity state. In order to explore the method for computing the implicit evaluations among the contributors, we analyzed a set of VGI objects edited from the same image (they can be seen as the different versions of the same state). For example, a swimming pool in OSM with id=4839528 includes three versions, i.e., A_1, A_2 and A_3 (Figure 2). A_1, A_2 and A_3 were formed at January 15, January 18 and March 17, 2011 respectively. The three versions were yielded in two months, and the online image did not change during this period, therefore, the three versions can be seen as the different versions for the same entity state. From Figure 2, it can be concluded that the quality of the object is improving from A_1, A_2 to A_3 , the detail is enhanced from A_1, A_2 to A_3 . Thus, the A_2 and A_3 can be seen as the edition used to improve the quality of the same entity state by the succeeding contributors using “many eyes principle”. In fact, when a succeeding contributor editing the preceding version, he (or she) evaluated the corresponding versions using his (or her) mouse. The evaluation value is implicit in the extent of the modification to the preceding versions.

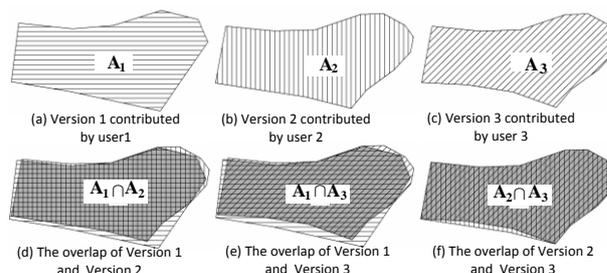


Figure 2 the similarity among different versions of the same entity state

In Figure 2, the area of A_2 intersects A_3 is larger than that of A_1 intersects A_3 , i.e., the spatial similarity degree between A_2 and A_3 is larger than that between A_1 and A_3 . It usually can be seen

as that the support degree of user 3 to user 2 is larger than that of user 3 to user 1. In another words, the spatial similarity degree among the different versions for the same entity state can be used to compute the support degree among the contributors. Furthermore, the support degree among the contributors usually can be seen as the reputation evaluation among the contributors.

3. COMPUTATION MODEL OF CONTRIBUTOR'S REPUTATION

From the above analysis, we can conclude that the reputation of a contributor can be calculated using the support degree from the users given to the objects he (or she) contributed. the support degree from the users can be computed by the similarity degree among the different versions for the same entity state yielded by the different users. The similarity degree among the different versions for the same entity state can be defined as version similarity degree. In GIS, spatial object usually include two basic properties, i.e., spatial property and thematic property. The version similarity degree also can be calculated using the spatial similarity and thematic similarity among the versions.

It is assumed that A is an object in VGI context with n versions, v_i is the i^{th} version edited by u_i , v_j is the j^{th} version edited by u_j , $0 < i < j \leq n$. The version similarity degree between v_i and v_j ($VerSim(v_i, v_j)$) can be computed using Equation (1).

$$VerSim(v_i, v_j) = w \cdot VSS(v_i, v_j) + (1 - w) \cdot VTS(v_i, v_j) \quad (1)$$

In Equation (1), $VSS(v_i, v_j)$ and $VTS(v_i, v_j)$ denote the spatial and thematic similarity respectively, w ($w \in [0, 1]$) is the weight of spatial similarity. The spatial and thematic similarity degrees can be calculated using Equation (2) and (3) respectively.

$$VSS(v_i, v_j) = \frac{\sum_{k=1}^K w_k \cdot SS_k(v_i, v_j)}{\sum_{k=1}^K w_k} \quad (2)$$

$$VTS(v_i, v_j) = \frac{\sum_{l=1}^L w_l \cdot TS_l(v_i, v_j)}{\sum_{l=1}^L w_l} \quad (3)$$

In Equation (2) and Equation (3), K and L denote the number of the parameters involved in the spatial and thematic similarity degrees computation respectively; $SS_k((v_i, v_j))$ ($SS_k((v_i, v_j)) \in [0, 1], 1 \leq k \leq K$) denotes the spatial similarity degree between the k^{th} spatial factor, $TS_l((v_i, v_j))$ ($TS_l((v_i, v_j)) \in [0, 1], 1 \leq l \leq L$) denotes the thematic similarity degree between the l^{th} thematic factor; w_k and w_l denote the weight of the corresponding factors. The spatial factors usually include the position, shape, area for region objects, length for line objects, etc.; the thematic factors usually include type, name, etc.

According to our analysis to OSM historical data, in VGI system a user usually contributed a set of objects, even an entity state may modified several times by the same user. For example, as Figure 3 shows, an entity state has 6 versions, i.e., $v_1, v_2, v_3, v_4, v_5, v_6$. v_1 and v_2 are contributed by User A; v_3 and v_5 are contributed by User B. The phenomenon of multi-versions contributed by a same user for an entity state usually is caused by the edition event broken for some reason or the user want to store the edition result momentarily. Therefore, for the multi-versions contributed by a user for an entity state, it is reasonable to think that the same user does not evaluate the reputation of himself, and just the last version is effective for the reputation evaluation among users, e.g., in Figure 3, just v_2 ,

v_4, v_5 , and v_6 are used to take part in the reputation evaluation among User A, User B, User C, and User D.

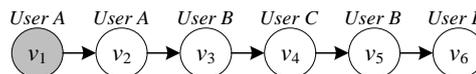


Figure 3 Successive versions of an entity state

As mentioned above, just the versions for the same entity state can be used for reputation computation. While usually, the reason for user modification is not available in VGI system (e.g., in OSM), in this paper, it is assumed that the versions in a short time period are corresponding to the same state, i.e., a time threshold τ is used to distinguish the versions are belong to a same state or not. In addition, as in social communication, the comment from a high reputation person usually is more important than the others, the reputation value of the former user getting from the subsequent user and the reputation degree of the subsequent user are in the direct ratio too.

Thus, it is assumed that: A is an object in VGI context with N versions, v_i is the i^{th} version edited by u_i at time t_i , v_j is the j^{th} version edited by u_j at time t_j , $0 < i < j \leq N$. If the reason for modification v_i to v_j is that the entity state has changed, then u_i does not get reputation evaluation from u_j ; if the reason is for improving the quality clearly, then the reputation of u_i getting from u_j ($rep_{u_i}(v_i, v_j)$) can be calculated using Equation (4); if the reason is not clear, a period threshold τ is used to determine if the two versions are belong to the same entity state, i.e., if $t_j - t_i \leq \tau$, the $rep_{u_i}(v_i, v_j)$ still can be computed using Equation (4); else the probability of the two versions are not belong to the same entity state, $rep_{u_i}(v_i, v_j)$ will be not calculated. The total reputation of u_i getting from his (her) contributed version v_i ($rep_{u_i}^{v_i}$) can be computed using Equation (5)

$$rep_{u_i}(v_i, v_j) = VerSim(v_i, v_j) \cdot R(u_j) \quad (4)$$

$$rep_{u_i}^{v_i} = (rep_{u_i}(v_i, v_{i+1}) + \dots + rep_{u_i}(v_i, v_j) + \dots + rep_{u_i}(v_i, v_N)) / (N - i) \quad (5)$$

In VGI context, one user usually contributed a set of object, the reputation for these prolific users can be computed using the average reputation value of each single object he (or she) contributed. It is assumed that user u_i contributed M objects O_l ($1 \leq l \leq M$), $rep_{u_i}^l$ denotes the reputation user u_i getting from O_l , $R(u_i)$ denotes the reputation of user u_i . $R(u_i)$ can be computed using Equation (6).

$$R(u_i) = \frac{\sum_{l=1}^M rep_{u_i}^l}{M} \quad (6)$$

4. TRUST DEGREE COMPUTATION MODEL FOR CROWDSOURCING GEOGRAPHIC DATA

As mentioned above, the quality of the crowdsourcing geographic objects is determined by the reputation of its contributor. The reputation of the contributor is dynamically and movable. The quality (or trust degree) of the objects contributed by a same contributor in a not long term usually is at the same level. The reputation of the contributors can be computed using Equation (7). According to the contributor-

reputation- movable property, the trust degree of VGI object can be evaluated using the reputation of its contributor.

However as mentioned in section 2, in a VGI context, one entity state may have several versions yielded by several contributors with different reputation, a non-first versions usually are the result of collective intelligence. It is clear that the reputation of the last contributor can not represent the trust degree of this object version. As in the production line of many commodities, there are many worker in the production line, while the reputation of each worker just can effect the quality of the part he contributed; in VGI objects' production, the reputation of a contributor only can represent the trust of the part he contributed (or he confirmed) too. For example, it is assumed that one entity state L has 5 versions, i.e., v_0, v_1, v_2, v_3, v_4 as Table 1 shows, v_0 is NULL, v_1 is contributed by u_1 , v_2, v_3 and v_4 are modification from the former version, e.g., v_2 is modification from v_1 by u_2 , v_3 is modification from v_2 by u_3 , and so on. The reputation of u_1, u_2, u_3 and u_4 are 0.7, 0.9, 0.4, and 0.85 respectively, the modification proportion of each version is shown in Table 1. From usual cognition perspective, the trust degree of the involved (i.e., v_0, v_1, v_2, v_3, v_4) versions should be equal to the values in Table 1.

Table 1. Editing process and trustworthiness for the versions of an entity state

	versions				
	v_0	v_1	v_2	v_3	v_4
Contributor	Null	u_1	u_2	u_3	u_4
Reputation of contributor	Null	0.7	0.9	0.4	0.85
Modification proportion	0%	100%	10%	50%	40%
Trust degree	Null	0.7	0.9	0.65	0.85

From the above analysis, we concluded the following rules:

- 1) The trust degree of the first version can be equal to the reputation of its contributor.
- 2) For the other versions, the trust degree of the subsequent versions is the composition of the trust degree of the former version and the reputation of the new contributor. The trust degree of the new modification proportion is equal to the reputation of the new contributor. As the remainder proportion (i.e., the overlapping part with the former version) is confirmed by the new contributor too, the trust degree of this proportion can be the max value between the trust degree of the former version and the reputation of the new contributor. In fact the new version is composed of the modification part and the overlapping part, the overlapping proportion can be equal to the version similarity degree. Therefore the trust degree of the subsequent versions can be calculated using version similarity degree too.

Based on the above observation, we construct the trust degree computation model based on version similarity degree. It is assumed that $T(v_i)$ denotes the trust degree of version v_i , $R(c_i)$ denotes the reputation of its contributor c_i , $VerSim(v_i, v_j)$ denotes the version similarity degree between v_i and v_j . Thus $T(v_i)$ can be calculated by Equation (7).

$$T(v_i) = \begin{cases} R(c_i) & i=1 \\ (1-VerSim(v_i, v_k)) \cdot R(c_i) + VerSim(v_i, v_k) \cdot Max(T(v_k), R(c_i)) & i > 1 \end{cases} \quad (7)$$

In Equation (7), v_k is the previous version of v_i (k is less than i). Therefore, for the entity state just has one version, i.e., $i=1$, the trust degree of the object version equals the contributor's reputation. Otherwise, for the object state with multi-versions, i.e., $i > 1$, the trust degree of the object version v_i is determined by the trust degree of its previous version $T(v_k)$, the contributor reputation of v_i , i.e., $R(c_i)$, and the modification proportion between versions v_i and v_k (or the similarity degree between v_i and v_k).

5. EXPERIMENTAL TEST AND ANALYSIS

In order to verify this presented model, a prototype system for computing the trust degree of VGI objects is developed by programming with Visual C# 2010. The historical data of Berlin of OpenStreetMap (OSM) from 28 January 2006 to 15 February 2013 are employed for experiments. The file format of the dataset is XML, and total data is about 1.96G and it contains editing history of all objects in OSM. The OSM objects are converted to points, lines and polygons of the Chinese national fundamental data model using the Rule-based model transformation method (Zhou et al, 2015). The line objects are used as the testing dataset, including about 557 thousand line object versions and about 2.7 thousand contributors.

In fact, in the trust degree computation model presented in this paper, the initial value of the contributors' reputations is a fundamental for the trust degree computation of VGI versions. The initial reputation can be calculated using the contributor's registration information, e.g., identity information, major or education, e-mail, cell phone, etc. (Zhao et al, 2016). However, OSM does not release the related information. Therefore, we use the following rules for our experiments in this paper.

- 1) The initial reputations of all contributors are set as NULL if they have never been assessed.
- 2) When a contributor with NULL reputation assesses another contributor, the assessment value would be computed by the version similarity merely.
- 3) When a VGI object is contributed by volunteers with NULL reputations, the rules for this case is shown as following.
 - a) If the version number of the object is 1, its trust degree is NULL, i.e., discard to evaluate this object;
 - b) Else if the version number is greater than 1, the average reputation of all non-NULL value contributors is used as the trust degree of this object.

Furthermore, version similarity plays an important role in our trust degree computation model. As mentioned above, version similarity include spatial similarity and thematic similarity. However, from cognition perspective, the spatial similarity is widely accepted by people, and there are a set of methods for the computation of spatial similarity for line and area objects in the existing contexts, the computation methods are objective and representative for version similarity; while for thematic similarity, especially thematic similarity evaluation among OSM objects contributed by non-professional volunteers is uncertain. Therefore according to the clear criterion rules (Chang, et al, 2006), only spatial similarity is used in the trust degree computation experiments in this paper.

Several researchers have studied the spatial similarity computation methods for line and area objects, e.g., (Arkin & Chew et al., 1991; Veltkamp & Hagedoorn, 2001; Masuyama, 2006; Shi & Cheung, 2006; Fan & Zipf et al., 2014). According to the characters of the line objects and the requirements of trust assessment for VGI, version similarities of line objects are

computed from length similarity, overlapping rate of the buffer area and shape similarity between versions. Among them, length similarity of line L_1 and L_2 ($SimLen(L_1, L_2)$) is calculated using equation (8). In equation (8), the function $Len(l)$ denotes the length of the line l ; the function $Max(x, y)$ denotes the maximum value of x and y parameters.

$$SimLen(L_1, L_2) = 1 - \frac{|Len(L_1) - Len(L_2)|}{Max(Len(L_1), Len(L_2))} \quad (8)$$

The overlapping rate of buffer area of line L_1 and L_2 ($SimBuff(L_1, L_2, R_{bf})$) is calculated using equation (9). In equation (9), the function $Buff(l, R_{bf})$ denotes the buff area of the line l ; R_{bf} is the distance of the buffer.

$$SimBuff(L_1, L_2, R_{bf}) = 1 - \frac{|Buff(L_1, R_{bf}) - Buff(L_2, R_{bf})|}{Max(Buff(L_1, R_{bf}), Buff(L_2, R_{bf}))} \quad (9)$$

Shape similarity of line L_1 and L_2 ($SimShape(L_1, L_2)$) was calculated using equation (10).

$$SimShape(L_1, L_2) = 1 - \frac{TFD(L_1, L_2)}{2\pi} \quad (10)$$

In equation (11), $TFD(L_1, L_2)$ denotes the distance of turning angle (Arkin & Chew et al., 1991; Fan & Zipf et al., 2014). As the max value of distance of turning angle between two lines is 2π , it is used to normalize the distance.

In the experiments, the value of the period threshold τ (used to determine if the two versions are belong to the same entity state) is 365. Users' reputations and trust degree of VGI are computed according to the generation time of the object versions. In our experiments, 2494 contributors are evaluated, the range of users' reputation value is from 0 to 1. The reputation are assigned to 10 levels according to the reputation value, i.e., 0-0.1, 0.1-0.2, ..., 0.9-1. The number of the users in each level is {2, 4, 22, 45, 257, 1629, 413, 90, 19, 13}, it obeys normal distribution in general.

As the disappeared entities usually are not represented in the current image, the quality of the deleted objects is hard to valid, the trust degree of the deleted objects (including 62548 line objects, corresponding to 206141 object versions) are ignored in the model validation analysis. Therefore, there are 104040 line objects (containing 350956 versions) are used in our model validation experiment. Like user reputation, the trust degree of the OSM line object final versions (i.e., the versions at 15 February 2013) are assigned to 10 levels according to the trust degree value, i.e., 0-0.1, 0.1-0.2, ..., 0.9-1. As the number of the objects with low-level trust degree values (i.e., the objects with 0-0.4 trust degree value) is small, the four levels are merged to one level, i.e., 0-0.4. The trust degree distribution of Berlin's line objects with 7 levels is shown in Figure 4. The assessment coverage is approximately 95% in this experiment.

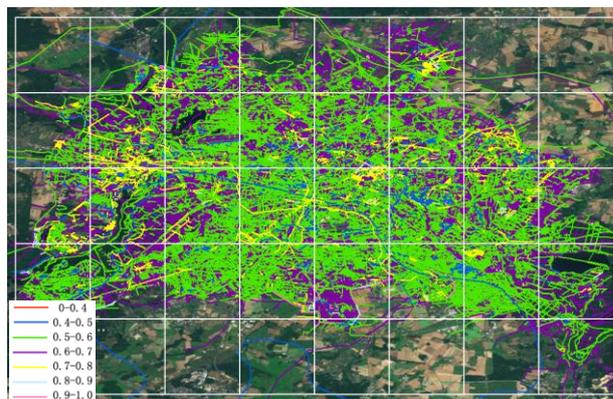


Figure 4. The trust degree distribution of Berlin's line objects with 7 levels

To evaluate our trust degree computation model, a set of samples are extracted by professional people from google image for each trust degree levels respectively. As the number of the objects with two end part trust degree values (i.e., 0-0.4, 0.4-0.5, 0.8-0.9 and 0.9-1) is very limited, these objects are selected as samples. The samples for the other level objects are selected considering the grid distribution (Figure 4) and subdivided trust degree levels. 1443 samples are extracted totally. 212 samples are removed because they are difficult to recognize from the image or other reasons, we obtain 1231 valid samples finally. These valid samples are contributed by 268 contributors, which account for approximately 25.2% of all the contributors of the final versions of line objects.

The samples are divided into four categories according to their quality, i.e., good, common, poor and error, which are represented by A, B, C and D respectively. The sample objects in category A are highly consistent with the actual geographic object. The sample objects in category B exist some positional offset or somewhat inaccuracy to the actual geographic features. The sample objects in category C are consistent with actual geographic targets basically with big offset or inaccuracy classification. The sample objects in category D do not exist. The distribution of the sample line objects is shown in Table 2.

Table 2. distribution of the sample line objects

Intervals of trust degree	Valid sample	A good	B common	C poor	D error
0-0.4	1	1	0	0	0
0.4-0.5	333	35	115	114	69
0.5-0.6	223	125	72	21	5
0.6-0.7	281	240	37	4	0
0.7-0.8	189	174	14	1	0
0.8-0.9	158	138	20	0	0
0.9-1.0	46	44	2	0	0

As Table 2 shows, there are a total of 1231 valid samples. Among them, 74 sample objects (approximately 6%) belong to the category D (error); 142 (about 11.5%) of the samples pertain to the category C (poor); 258 (approximately 21%) samples belong to the category B (common); and 757 (about 61.5%) samples pertain to the category A (good).

From Table 2, we can conclude that the number and proportion of poor objects is gradually decreasing as the increasing of trust degree level. Generally speaking, the quality of the objects in VGI and their trust degree are positively correlated. And all error objects locate in the range of lower trust degree intervals. In another words, our trust degree computation model is effective and can be used to evaluate the quality of VGI.

6. CONCLUSIONS AND DISCUSSION

The forming mechanism of the VGI object version is analyzed at first in this paper. According to the analysis: in VGI context, one state of an entity may have several versions, updates can be made due to two main reasons, i.e., a real entity's state change and to improve the quality of VGI object. There are implicit evaluations among the contributors who yield the different versions for the same entity state. The evaluation can be calculated using the similarity degree among the multi-versions for the same entity state. Therefore, a version-similarity based reputation computation model is presented. As the reputation of the contributor is movable, and the quality of VGI objects mainly determined by the professional skill and integrity (called reputation together in this paper) of its contributor, a version-similarity based trust degree computation model for crowdsourcing geographic data is presented in this paper. In this model, for the object state just has one version, i.e., $i=1$, the trust degree of the object version equals the contributor's reputation; for the object state with multi-versions, i.e., $i>1$, the trust degree of the object version v_i is determined by the trust degree of its previous version v_k , the contributor reputation of v_i , and the modification proportion between v_i and v_k . Experimental results demonstrate that the quality of the crowdsourcing geographic data is high positive correlation with its trustworthiness. Comparison to the existing models, our model has the following characteristics:

- 1) It is based on the assumption that the quality of VGI objects mainly determined by the professional skill and integrity (called reputation in this paper) of its contributor, and the reputation of the contributor is movable.
- 2) In VGI context, there are many objects with multi-versions for the same state, there are implicit evaluations among the contributors who yield the different versions for the same entity state.
- 3) The computation of the implicit assessment among the users in our model is based on the observation, i.e., "less modification made by the latter user to the former version means that higher support degree was evaluated to the former contributor, and vice versa". Version similarity is used to calculate the implicit assessment.
- 4) As the evaluation is based on version-similarity, not based on the direct subjective evaluation among users, the result is objective.
- 5) With the contributors' reputation movable property, the presented model has a higher assessment coverage than the existing methods.

It is needed to point that, as the first version objects contributed by a newly registered user can not be evaluated in this model, the assessment coverage still can not arrived 100%. In fact, in the E-service systems, the registration information usually is used as the initial reputation computation criteria; in addition, the indirect confirmation (i.e., a user edited an objects, he (or she) can be look as the vicinity of the objects in a certain near area to the edited objects) can be taken into account for the

assessment (Kefler & Groot, 2013), especially for the objects editing from imagery; furthermore, the quality of the object per se (e.g., the Lineage, the completeness of the thematic information, the consistency with the other objects in the region, etc.) and the uncertain assessment of the trust computation method will be considered in the further studies.

ACKNOWLEDGEMENTS

This study was funded by the National Natural Science Foundation of China (Project No. 41371366).

REFERENCES

- Arkin, E. M., L. P. Chew, D. P. Huttenlocher, et al., 1991. An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE transactions on pattern analysis and machine intelligence*. 13(3): 209-216.
- Bishr, M.; Mantelas, L., 2008. A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal*, 72, 229-237.
- Chang, E., Dillon, T., and Hussain, F. K., 2006. *Trust and Reputation for Service-Oriented Environments: Technologies for Building Business Intelligence and Consumer Confidence*, Wiley, West Sussex, UK
- Cipeluch, B., Jacob, R., Winstanley, A., Mooney P., 2010. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. *Proceedings of the ninth international symposium on transportation and traffic theory*
- Fan, H., A. Zipf, Q. Fu, et al. , 2014. Quality Assessment for Building Footprints Data On OpenStreetMap. *International Journal of Geographical Information Science*, 28(4): 700-719.
- Forghani, M., M. R. Delavar. 2014, A Quality Study of the OpenStreetMap Dataset for Tehran. *ISPRS International Journal of Geo-Information*, 3(2): 750-763.
- Girres, J., Touya G., 2010. Quality Assessment of the French OpenStreetMap 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
- Goodchild, M. F., Li, L., 2012. Assuring the quality of volunteering geographic information. *Spatial Statistics*,(1):110-120
- Haklay, M.; Weber, P., 2008. Openstreetmap:User-generated street map. *Pervasive Computing*, 7, 12-18.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *American Journal of Cardiology*, 93(11): 3-11
- Kefler, C.; de Groot, R.T.A., 2013. Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap, In *Geographic Information Science at the Heart of Europe*, Springer; 21-37.

Matyas, S., Kiefer P., Schlieder C., et al. 2011. Wisdom about the Crowd: Assuring Geospatial Data Quality Collected in Location-Based Games, ICEC, LNCS 6972, pp.331–336

Masuyama, A. 2006. Methods for Detecting Apparent Differences Between Spatial Tessellations at Different Time Points. *International Journal of Geographical Information Science*, 20(6): 633-648.

Shi, W. Z., Cheung C. K.. 2006. Performance Evaluation of Line Simplification Algorithms for Vector Generalization. *The Cartographic Journal*, 43(1): 27-44.

Veltkamp, R. C., Hagedoorn M., 2001. State of the Art in Shape Matching. *Principles of Visual Information Retrieval*, Springer: 87-119

Zhao, Y.J., Zhou, X.G., Li, G.Q., et al., 2016, A Spatio-Temporal VGI Model Considering Trust-Related Information, *ISPRS International Journal of Geo-Information*. *ISPRS International Journal of Geo-Information*.. 5(2), 10; doi:10.3390/ijgi5020010

Zhou, X., Zeng, L., Jiang, Y., et al., 2015. Dynamically Integrating OSM Data into a Borderland Database. *ISPRS International Journal of Geo-Information*, 4, 1707-1728.

Zielstra D., Zipf A., 2010. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. 13th AGILE International Conference on Geographic Information Science