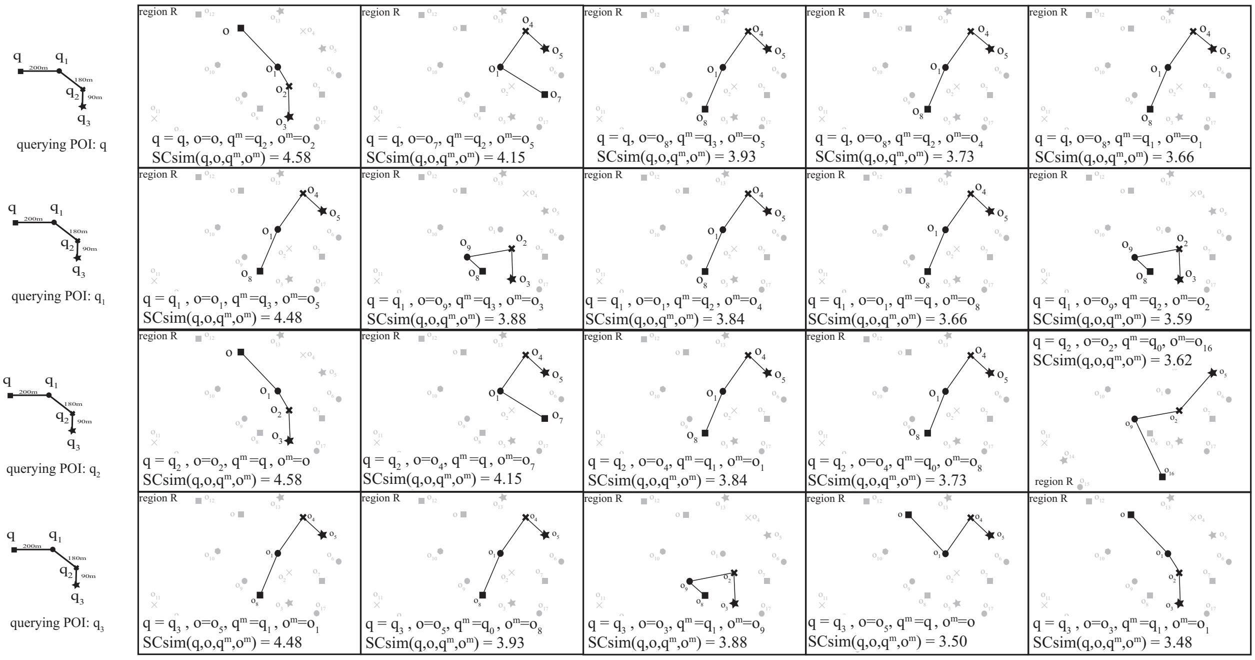

Figure 18: Top 5 instances with the highest spatio-textual similarity ($\beta = 0.01$).

APPENDIX

A. QUERY TRANSFORMATION

For computing $SCsim(q, o, q^m, o^m)$, q is placed at the same location as o, and then transform $Q_R(q, N, E)$ to place q^m at the same location as o^m. The *transformation* rotates and uniformly scales $Q_R(q, N, E)$. For each POI $q_i \in N \setminus \{q, q^m\}$, the position is $(q_i.x, q_i.y)$. After the transformation, it has a new position:

$$q_i.newx = \frac{\|o, o^m\|}{\|q, q^m\|}(q_i.x \cos\theta - q_i.y \sin\theta). \quad (13)$$

$$q_i.newy = \frac{\|o, o^m\|}{\|q, q^m\|}(q_i.x \sin\theta + q_i.y \cos\theta). \quad (14)$$

where

$$\theta = \begin{cases} \arccos\alpha & \text{if } -1 < \alpha < 1 \\ 0 & \text{if } \alpha \geq 1 \\ \pi & \text{if } \alpha \leq -1 \end{cases}$$

α is the angle between edge $< q, q^m >$ and edge $< o, o^m >$

$$\alpha = \frac{q^m.x * o^m.x + q^m.y * o^m.y}{\sqrt{(q^m.x^2 + q^m.y^2) * (o^m.x^2 + o^m.y^2)}}. \quad (15)$$

B. SPATIO-TEXTUAL CONTEXT SIMILARITY

For the query shown in **Fig.** 1, the truly matching POI is o. That is, the query is generated by selecting o (as the querying POI q) and then select o_1, o_2, o_3 (as the clue POIs q_1, q_2, q_3) which are slightly moved to new positions. The query result is presented in **Fig.** 18. For the original query, the 5 matching instances with the highest spatio-textual similarities are shown in the first row. For each query generated in the ensemble method, the 5 matching instances with the highest spatio-textual similarities are shown in the 2-4 row respectively.

The results illustrate that matching instances with the higher similarities, intuitively, looks more similar to the clue. For the first row, the matching instance $\{o, o_1, o_2, o_3\}$ has the highest spatio-textual context similarity 4.58 against clue, the matching instance $\{o_7, o_1, o_4, o_5\}$ has the second highest similarity 4.15, the matching instance $\{o_8, o_1, o_4, o_5\}$ has the third, fourth and fifth highest similarity 3.93, 3.73 and 3.66 for different q^ms. It is interesting that $\{o_8, o_1, o_4, o_5\}$ looks very similar to the clue even though $\{o, o_1, o_2, o_3\}$ is the true matching instance. For the second row, q_1 is the querying POI. The most similar matching instance is $\{o_8, o_1, o_4, o_5\}$ with spatio-textual context similarity 4.48. The querying POI in the third row is q_2 and in the fourth row is q_3, the most similar matching instances are $\{o, o_1, o_2, o_3\}$ and $\{o_8, o_1, o_4, o_5\}$ respectively.

Even though the true matching instance is $\{o, o_1, o_2, o_3\}$, the test results suggest that both $\{o, o_1, o_2, o_3\}$ and $\{o_8, o_1, o_4, o_5\}$ are very similar to the clue. This is because the query is the approximation of the true matching instance. The true matching instance may be beat by some other matching instance. Since the true matching instance is unknown in advance, it is necessary to identify all very similar matching instances and return them to user.

If the ensemble method is not used, the result in the first row is returned to user. It indicates that the matching instance $\{o_7, o_1, o_4, o_5\}$ is more similar against the clue than $\{o_8, o_1, o_4, o_5\}$ even though $\{o_8, o_1, o_4, o_5\}$ is more similar against clue intuitively. If the ensemble method is used, the score of $\{o_8, o_1, o_4, o_5\}$ is higher than that of $\{o_7, o_1, o_4, o_5\}$, that is, the ensemble method can provide the results more consistent intuition.

In **Fig.** 19, it shows the results when two more clue POIs are specified in the query where the true matching instance is $\{o, o_1, o_2, o_3, o_{12}, o_{13}\}$. The first row is the result of the original query and the other rows are the results of the five generated queries in ensemble method. We can observe that the true matching instance dominates the results. It verifies that the more clue POIs in the query leads to higher accuracy.

Comparing the results in **Fig.** 19 and **Fig.** 18, it is worth to point out, if some clue POIs in the query such as o_{13} and o_{14} do not

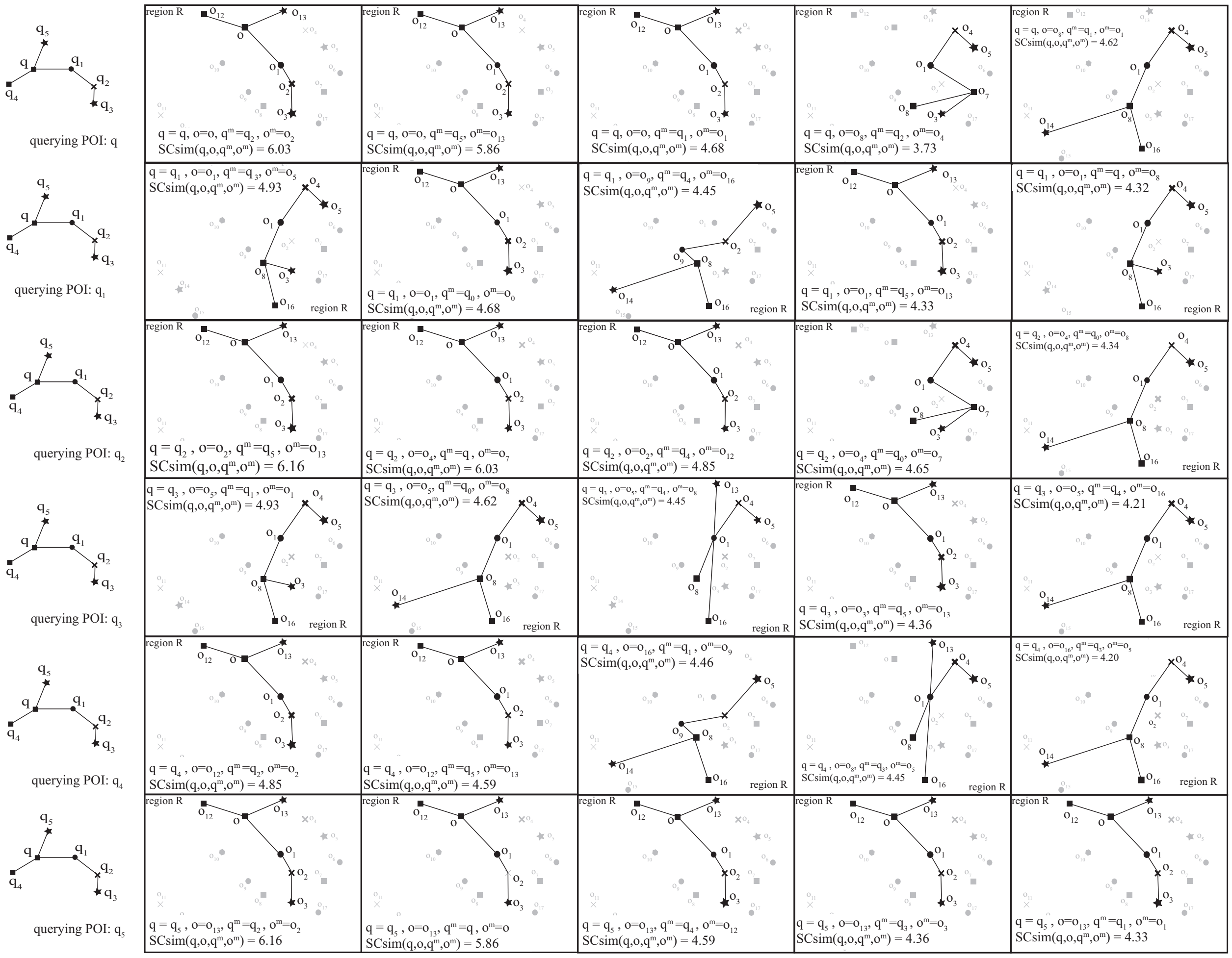

Figure 19: Top 5 instances with the highest spatio-textual similarity ($\beta = 0.01$).

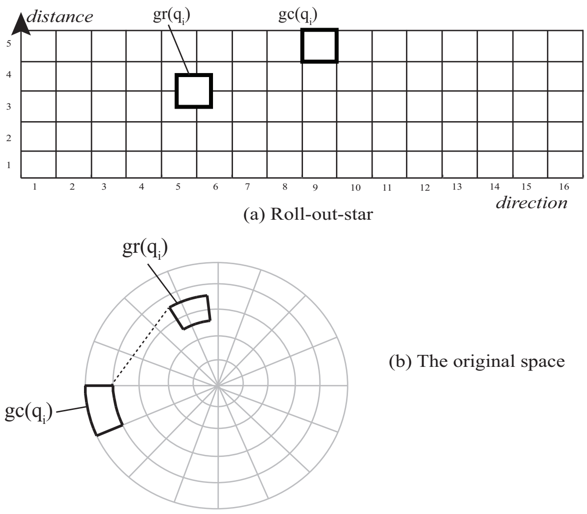

Figure 20: Computing the minimum distance between $gr(q_i)$ and $gc(q_i)$.

have truly matching POIs in the database due to data quality problems (e.g., incorrect category, category change, database incompleteness/update), they will be largely ignored in the similarity metric on purpose such that the remaining clue POIs can still work. This is verified as shown in **Fig.** 18. The true matching instance is still successfully identified even though it also returns other similar matching instance.

C. MINIMUM DISTANCE TO GRID CELL

Given $gr(q_i)$ and $gc(q_i)$, let $\{cr_1, cr_2, cr_3, cr_4\}$ and $\{cc_1, cc_2, cc_3, cc_4\}$ be the four corners of $gr(q_i)$ and $gc(q_i)$ respectively. The positions of four corners in original space can be obtained based on the corresponding four corners in roll-out-star. The example shown in Figure 20 illustrates the positions of $gr(q_i)$ and $gc(q_i)$ in the original space. It is easy to prove that the minimum distance between $gr(q_i)$ and $gc(q_i)$ is the minimum distance between the corners of $gr(q_i)$ and the corners of $gc(q_i)$, i.e., $\min_{i=1,j=1}^{i=4,j=4} dist_E(cr_i, cc_j)$ in the original space.