Dataset Construction

In this page, We supplement more details about the dataset construction in our paper. Generally, the total dataset was constructed from Baidu Maps [1]. In the following subsections, we mainly discuss how we collect reviews and ground-truth description from noisy map data. The details about category information and context information collection are also reported.

A. Reviews Collection

Since the POI reviews serve as the main input in the POI description generation, we started our dataset construction from the review collection. At first, we filtered out two kinds of original reviews from map data: too short reviews and reviews with repeated meaningless phrases such as "Great, Great, Great". For the first case, we observed that a too-short review is more likely to be meaningless and noisy. We filtered out the reviews that are shorter than 40 tokens. For the second case, we calculated the number of words n_1 and the number of unique words n_2 in a review. The review with $n_2/n_1 \leq 0.6$ would be filtered out. In the rest reviews, we annotated the word with a frequency lower than 100 as unk (unknown) word, and build a vocabulary with 12,007 words. Then we filtered out the reviews that contained more than 5 unk words.

In this way, we got 43,225,977 reviews that cover 2,122,675 POIs. We concatenated the reviews for each POI as input. With attention mechanism especially self-attention, the order of the input reviews is not important to representation learning. It means the concatenation here is reasonable. To alleviate the computation complexity and the degradation of generative models with long sequence input, we kept the number of reviews in concatenation less than 6 and the number of tokens in concatenation less than 300.

B. Description Extraction

It is disappointing that most POIs suffer from the lack of ground-truth description in map data, while the POI description from Pedia are too stiff to serve as ground-truth and can only cover a few famous brands. Following the research conducted by Novgorodov et al. [1], which proved that some high-quality reviews can be seen as description directly, we applied a heuristic approach which combined the machine learning model and some simple filtering rules to extract description from reviews data.

In the machine learning part, we collected a small labeled dataset $\mathcal{Z}$ by manual assessment. Dataset $\mathcal{Z}$, composed of 16,836 samples, is annotated whether a review can serve as a description. We implemented the discriminator model based on LSTM to predict whether a given review can serve as description, following Novgorodov et al. [1]

For filtering rules, we analyzed the distribution of the words in reviews and description. Table I shows some words with a large difference between the frequency in reviews and description. We can find that "very" has a larger frequency because it tends to convey extreme emotions, while "again", "satisfy", "feel" and "friend" have a larger frequency in reviews because they tend to describe a personal experience rather than description. However, we expected the description to be objective, informative and vivid. Thus, we adopted the following filtering rules.

- Length control. While short reviews are not informative enough, the long reviews are often telling a specific story of personal experience. Thus, we filtered the reviews with length between 100 and 240 tokens experimentally.
- Reviews that with extreme words. Most of the reviews with extreme words are just venting emotion and insulting the POI. They can not help describe the POI. In particular, we removed reviews that contain "very sick", "disgusting" and other extreme phrases.
- Personal reviews. Different from reviews, the description is expected to be objective rather than subjective. To this end, we filtered out reviews contain "very like", "has(have) visited", "often" and other phrases, since most of these reviews are too personal and indicate a personal experience.

For those extracted description, we also built a vocabulary with 9,973 words in the same way with reviews. Finally, we extracted 691,224 description which cover 380,990 POIs with the combined method. We organized a manual evaluation on 1,000 random samples, and the precision of our extraction is 95.2%. The reviews that selected as description were removed from reviews data then.

C. Category and Context Information

As discussed in our paper, we also collected the category information for each POI. In our dataset, the number of categories m is 88. The top five categories in our dataset are "Chinese restaurant"(25.64%), "Fast-food restaurant"(10.46%), "Express Inn"(6.44%), "Cakeshop"(5.75%) and "Small store"(4.86%).

As for context information, we extracted the nearby map from a square area with side length $h = 3km$. Then we divided the nearby map into 10×10 grids. Since there are 88 POI categories in our dataset, the shape of the final context tensor is $10 \times 10 \times 88$.

TABLE I
FREQUENCY OF WORDS IN REVIEWS AND DESCRIPTION

Word	Reviews	description
again	41.04%	15.78%
very	30.83%	15.59%
satisfy	13.96%	0.22%
feel	11.00%	0.27%
often	10.24%	0.20%
friend	10.78%	0.99%
...	...	...

References

[1] S. Novgorodov, I. Guy, G. Elad, and K. Radinsky, "Generating product descriptions from user reviews," in *WWW*, 2019, pp. 1354–1364.

[1] https://maps.baidu.com/