Building Location Embeddings from Physical Trajectories and Textual Representations

Laura Biester
Carmen Banea
Rada Mihalcea

{lbiester,carmennb,mihalcea}@umich.edu
Outline

Introduction

Data Collection

Methods

Experiments

Conclusions
Overview

• Dense vector representations (embeddings) are commonly used in NLP to represent words, and have also been applied to locations

• We use location trajectories and text data to create embeddings

• To evaluate, we explore:
  • Surface level tasks to better understand what location embeddings encode
  • Downstream tasks to see if they can be used for predicting personal attributes

Liu, Xin, Yong Liu, and Xiaoli Li. "Exploring the Context of Locations for Personalized Location Recommendations." IJCAI. 2016.
Feng, Shanshan, et al. "Poi2vec: Geographical latent representation for predicting future visitors." Thirty-First AAAI Conference on Artificial Intelligence. 2017.
Solomon, Adir, et al. "Predict demographic information using word2vec on spatial trajectories." Proceedings of the 26th conference on user Modeling, adaptation and personalization. 2018.
Chang, Buru, et al. "Content-Aware Hierarchical Point-of-Interest Embedding Model for Successive POI Recommendation." IJCAI. 2018.
Sadilek, Adam, Henry Kautz, and Vincent Silenzio. "Modeling spread of disease from social interactions." Sixth International AAAI Conference on Weblogs and Social Media. 2012.
Research Questions

1. Do **location embeddings** encode meaningful **semantic information**?
2. What **sources of data** are most informative about locations?
3. Can location embeddings aid us in **downstream tasks**?
**Example Use Case: Health Monitoring**

- One downstream task is **depression detection**
- A possible application is individual-level monitoring for people who are **in treatment for depression**

| Mild symptoms | Severe symptoms |
|---------------|-----------------|
| .8 | .7 |
| .2 | .3 |
| -.3 | -.1 |
| .9 | .7 |
| -4 | -5 |
| .3 | .1 |
| -.4 | -.3 |
| .5 | .4 |
| -3 | -3 |
| .2 | .5 |
| -.4 | -.5 |
| -.3 | -.5 |

- Week 1
- Week 2
- Week 3
- ... (Week N-1)
- Week N
Data Collection
Location Data

- Data consists of **WiFi updates** from connections to **buildings at the University of Michigan**
- Locations are tagged with Metadata
  - GPS position
- Functionality

Pros and Cons of Wifi Data

**Pro**
- Locations are automatically discretized, unlike with GPS
- This data is likely available on most campuses

**Con**
- We have no data when students are off campus
- Students can be connected to two access points at once

IRB: HUM00126298
Personal Data

For downstream tasks, we use data from a survey and the university administration, including:

- **Class year**: freshman, sophomore, junior, and senior
- **School of Enrollment**: School student is enrolled in at the University of Michigan LSA, Engineering, and Business
- **PHQ-8 Depression Score**: Processed according to clinically validated algorithm
Location-Related Text Data

I'm a transfer who accidentally signed up for a Baits II room and now can't back out

How do people express themselves in different locations?

How do people informally describe locations?

How are locations formally defined?

https://reddit.com/r/uofm
https://campusinfo.umich.edu/building-search/building/163/west-hall
Methodology
Trajectory-Based Vectors (Loc2V)

- **Traditional word2vec input:** sentences
- **Our word2vec input:** sequence of locations, ordered by time visited
- For each hour, record one location
  - *This is the location with the longest time spent in that hour*
- Provides a precise meaning to adjacent locations in a sequence

What can we learn from how people physically interact with the world?
Text-Based Representations

• Collect datasets that link locations to text
• Create a **weighted average** of pretrained GloVe embeddings
• Weight embeddings using **term frequency-inverse document frequency** (TFIDF)

**Methodology**

What can we learn from **what people say** in and about physical spaces?
Combining Trajectories and Text

Combine vector and graph to get a completely new vector

Concatenate two vector representations

Faruqui, Manaal, et al. “Retrofitting word vectors to semantic lexicons.” NAACL. 2015.
Experiments
Six Embedding Models

- **Loc2V**
  - Timestamp-aware sequences as input to word2vec
- **Twitter**
  - Text-based embeddings using Twitter data
- **Reddit**
  - Text-based embeddings using Reddit data
- **Campus Website**
  - Text-based embeddings using campus website data
- **Loc2V-Reddit**
  - Concatenated W2V-Hour and Reddit embeddings
- **Reddit-Retrofit**
  - Weighted retrofitted Reddit embeddings

Experiments
Functionality Overlap

Higher is better

Experiments

- Overall Avg
- Loc2V
- Twitter
- Reddit
- Campus-Website
- Loc2V-Reddit
- Reddit-Retrofit

Class
Lab
Library
Student Life
Physical Distance Results

Lower is better

Duderstadt

Pierpont

Chrysler

Dow

Beyster

GG Brown

Experiments
Downstream Task Results

Experiments
Conclusions
Do location embeddings encode meaningful semantic information?

Yes, all of our embedding methods encoded physical distance and functionality.
What types of data are most informative about locations?

Using text data tended to inform us more about functionality.

Using location trajectories tended to inform us more about physical distance.

On downstream tasks, results were task dependent.
Can location embeddings aid us in downstream tasks?

When predicting school and depression, we saw slightly stronger performance with location embeddings.

For other tasks requiring more surface level information, one-hot vectors led to better performance.