

個人に依存した時空間セマンティクスの分散表現の検討

庄子 和之¹ 米澤 拓郎¹ 廣井 慧¹ 河口 信夫^{1,2}

概要

近年, GPS (Global Positioning System) 機能を備えたスマートフォンやウェアラブル端末の普及により, 日々莫大な量の GPS 測位情報が生成されている. これらの GPS 測位情報は, ユーザの日々の生活パターンなどが如実に表れるものであり, これを分析することはそのユーザーの特徴 (性格, 趣味趣向) を抽出することに繋がる. 今後 IoT 社会が進み, 得られる GPS 測位情報がますます増えることが予想され, それらを有効に活用するための分析手法を考案することは大きな意味を持つと考えられる.

GPS 測位情報の分析の試みとして, word2vec の考えを取り入れてたものが多く存在する. 本来 word2vec は自然言語処理の分野で利用されるものである. しかし, 測位ロケーションを「単語」に, 一連の測位ロケーションの遷移を「文」と見立てると, といった工夫を施すことで人流解析の分野でも応用が期待できることが分かっている. その結果得られる分散表現は, 異なるロケーション間 [1] または異なるユーザ間 [2] の類似度を定量的に表すことになる. この分散表現を使えば様々なマーケティング戦略に活用できると考えられる.

word2vec とは, Mikolov ら [3] によって提案されたテキスト処理を行うためのニューラルネットワーク (NN) である. この手法は, 類似した文脈で用いられる単語は類似した意味を持つという分布仮説に基づいたものである. そして, word2vec には学習手法として, Skip-gram と Continuous Bag of Words (CBoW) の二つのモデルがある. Skip-gram は, ある単語を入力にその周辺の単語を予測するタスクを NN で学習する. 逆に CBoW は, 周辺単語を入力にその中心の単語を予測するタスクを NN で学習する. こうして得られた中間層の重みが各単語の分散表

現を表すようになる. この単語のベクトル空間において, ある単語の周辺によく表れる単語同士は近くに配置され, 逆に文章中に同時に出現しない単語同士は遠くに配置される. これを利用して単語間の類似度の計算や単語の持つ意味の加算, 減算が可能になる.

本研究では, word2vec の考えを利用して GPS 測位情報を学習させることで, 各ユーザーの「滞在場所」の分散表現を作成する手法を提案する. 既存研究には, ロケーションやユーザーの分散表現の作成というものは存在したが, 「滞在場所」という細かさのレベルは見られない. そのため, 従来よりもより詳細なユーザーの特徴が抽出できることが期待できる.

謝辞

本研究の一部は, JST CREST, 総務省 SCOPE, NICT 委託研究により支援して頂いております.

参考文献

- [1] Crivellari, A. and Beinat, E.: From Motion Activity to Geo-Embeddings: Generating and Exploring Vector Representations of Locations, Traces and Visitors through Large-Scale Mobility Data, *ISPRS International Journal of Geo-Information*, p. 134 (2019).
- [2] Esuli, A., May Petry, L., Renso, C. and Bogorny, V.: Traj2User: exploiting embeddings for computing similarity of users mobile behavior (2018).
- [3] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems 26*, pp. 3111–3119 (2013).

¹ 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

² 名古屋大学未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University

個人に依存した時空間セマンティクスのための分散表現の検討

庄子和之*, 米澤拓郎*, 廣井慧*, 河口信夫*

*名古屋大学 工学研究科



1. 背景

- GPS端末の普及による位置情報の活用需要が急増
- 座標のままでは扱いづらい
→ ユーザの“行動”特徴をモデル化
 - SVD, LSAを用いて**分散表現化**
- しかし、位置情報からユーザの行動特徴を詳細に抽出する手法はまだない



➢ ユーザへの個別化したサービスの提供に限界が...

2. 関連研究

- Generating and Exploring Vector Representations of Locations, Traces and Visitors through Large-Scale Mobility Data (Alessandro et al., 2019)
 - 地理的な“近さ”と、実際の“使われ方”に相関はない
 - Word2vec**を利用
 - ユーザの区別なく移動経路を学習
 - ✓ 入力: location
 - ✓ 出力: 入力locationの前後に居たlocation
- Traj2User: exploiting embeddings for computing similarity of users mobile behavior (Andrea et al., 2018)
 - “移動経路”から各ユーザを分散表現化
 - Word2vec**の考えを参考に
 - ✓ 入力: ユーザID
 - ✓ 出力: 移動手段, 天気, 時間帯を示す88次元の分散表

➢ **location**の一般化された分散表現を獲得

➢ **ユーザ**の分散表現を獲得

3. 目的

- ユーザを以下を考慮してモデル化
 - 移動経路
 - 時間帯, 曜日
 - 滞在場所

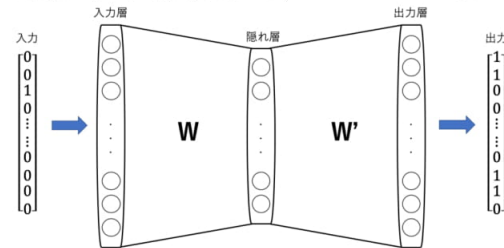
今回はユーザの“滞在場所”に着目し、これを分散表現化

- 分散表現のメリット
 - 定量的な比較が可能に
 - “長さ”や“近さ(類似度)”を算出
 - 機械学習を用いる際の入力に
- ✓ ユーザ間の滞在場所の“類似度”から
→ その場所の“使い方”の違いが判明
- ✓ 外出先のベクトルの“長さ”から
→ フットワークの軽さが判明



4. 提案手法

- “滞在場所”毎の分散表現化には、**word2vec**の考えを利用

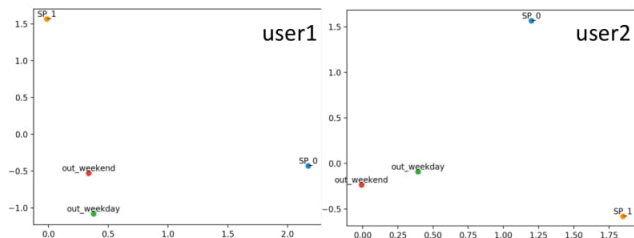


- 入力
 - 各滞在場所を表すone-hotベクトル
- 出力
 - いくつかの方法を検討中だが、Andrea et al.の手法が有効ではないかと考えている
 - 訪問日時, 滞在時間, 天気など

→ Wに“滞在場所”の分散表現が現れる

5. 実験

- ユーザの滞在場所の分散表現を学習
 - 入力: 各滞在場所を表すone-hotベクトル
 - 出力: 訪問時間(0~23時) + 退出時間(0~23時)の48次元ベクトル
(対応する時間にフラグが立てる)



- ✓ 滞在場所について
 - SP_0, SP_1: 頻繁に行く2箇所
 - out_weekday: 平日の外出
 - out_weekend: 休日・祝日の外出

➢ ユーザ毎に特徴が出ているのが分かる
• ここから各ユーザの休日の過ごし方などが定量的に算出できることが期待できる

6. 今後の方針

- 滞在場所の属性を推定
 - 自宅, 職場, 飲食店, スーパーなど
- “滞在場所”でなく, “時間”による分散表現の獲得
 - Wを曜日, 時間帯に変更
- “経路”も学習に使用
 - 地域別に以下の特徴が出てくるのではないかと期待できる
 - “通勤時間”, “通勤距離”