International Journal of Cognitive Informatics and Natural Intelligence
Volume 15 • Issue 2 • April-June 2021

Learning Trajectory Patterns via Canonical Correlation Analysis

Ping Huang, School of Mathematical Sciences, South China University of Technology, Guangzhou, China
Jinliang Lu, School of Mathematical Sciences, South China University of Technology, Guangzhou, China

ABSTRACT

A substantial body of research has been devoted to the analysis of motion trajectories. Usually, a motion trajectory consists of a set of coordinates, which is called a raw trajectory. In this paper, the authors first use vectors for some artificially constructed global features, such as the mean discrete curvature and standard deviation of acceleration, to represent the raw trajectory data, and then apply a multiset canonical correlation analysis method to extract latent features from the artificially constructed features. The performance of the latent features is then measured by evaluating the accuracy and F1 score of a gradient boosting decision tree model for different datasets, which include paired sample datasets and unpaired sample datasets. The experimental results show that the classifier performance for MCCA features is much better than that obtained for the artificially constructed features, such as that for the motion distance or mean velocity.

KEYWORDS

Discrete Curvature, Multiset Canonical Correlation Analysis, Raw Trajectory Classification, Unpaired Sample

1. INTRODUCTION

At present, it is important to analyze motion trajectories. By analyzing motion trajectories, different types of trajectories can be identified from a large number of trajectory data, such as distinguishing the choice of transportation - car or bus, classifying the traffic level-congested or smooth.

Trajectory data are generally a sequence of spatial location information data recorded alongside time that can be collected by an algorithm (K Li, Wang, & S Li (2019)). As given by Definition 1 in (Silva, Petry, & Bogorny (2019)), the raw trajectory data are the latitude and longitude information or the location of pixels in the image recorded alongside the time.

Trajectory classification is a very important part of trajectory analysis. Unlike other fields such as image recognition, trajectory classification focuses on how to extract effective features from the spatial position information sequence instead of constructing novel and efficient classifiers. Then, these features are used to train common classifiers such as Logistic Regression (LG), Gradient Boosting Decision Tree (GBDT), etc.

In this paper, we first extract the global features of a spatial position information sequence: the mean, standard deviation and entropy of the discrete curvature, the mean and standard deviation of the velocity and acceleration, as well as the length of the trajectory and motion duration. Second, we use multiset canonical correlation analysis (MCCA) to re-extract features from the artificially constructed features. Finally, we use the MCCA features to train a Gradient Boosting Decision Tree model, and conduct experiments using 7 different datasets to analyze the performance of the MCCA features.

DOI: 10.4018/IJCINI.20210401.oa1

The remainder of the paper is organized as follows: section 2 introduces the related research work for trajectory classification, the Gradient Boosting Decision Tree and canonical correlation analysis; section 3 introduces 9 features constructed in this work; section 4 introduces MCCA and how we can carry out MCCA for datasets with unpaired samples; section 5 includes experiments and results; section 6 includes the conclusion and future directions.

2. RELATED WORK

Recently, many studies have been carried out for trajectory classification. (Silva, Petry, & Bogorny (2019)) summarized three types of trajectory features: global features, local features, as well as global and local features, and presented an experimental comparison for several datasets with methods proposed by others. (Zheng, Li, Chen, Xie, & Ma (2008)) and (Sharma, Vyas, Schieder, & Akasapu (2010)) used global features to classify trajectories. The difference being that (Zheng, Li, Chen, Xie, & Ma (2008)) focuses on the transportation mode classification, so global features extracted by (Zheng, Li, Chen, Xie, & Ma (2008)) and (Sharma, Vyas, Schieder, & Akasapu (2010)) are different. Different studies show different choices for the classifiers. (Mlıch & Chmelar (2008)) and (Bashir, Khokhar, & Schonfeld (2007)) used Hidden Markov Models(HMM) to complete the classification task, while (Wang, Chu, Jiang & Li (2019)) used a Naive Bayesian Model (NBM) and (Liu & Lee (2017)) used BiLSTM. (Xiao, Wang, Fu, & Wu (2017)) used a Multilayer-Perceptron (MLP) and achieved a state-of-the-art accuracy of 75.56% and an F1 score of 73.83% for raw trajectories dataset "Geolife[1]" and achieved a state-of-the-art accuracy of 93.58% and an F1 score of 93.58% for the raw trajectories dataset "Animals[2]", while (Ferrero, Alvares, Zalewski, & Bogorny (2018)) proposed a new method called MOVELETS to achieve the goal of robust trajectory classification. With a Support Vector Machine, the MOVELETS method achieved a state-of-the-art accuracy of 92.30% and an F1 score of 90.82% for the multiple-aspect trajectories dataset "Animals[2]". The Gradient Boosting Decision Tree algorithm, which is widely used in many fields, and, in this paper, was proposed by (Friedman, (2001)). (Son, Jung, Park, & Han (2015)) applied the GBDT algorithm to object tracking tasks, and (Rao et al. (2019)) applied it to feature selection problems. At the same time, there are many optimizations for GBDT, such as the LightGBM algorithm proposed by (Ke et al. (2017)). Other optimization algorithms such as (Wang et al. (2018)) and (Wang et al. (2019)) can also be used. Canonical correlation analysis (CCA) was proposed by Hotelling, H., which is fully described in (Hotelling (1935)) and (Hotelling (1992)). There exists an expansion of CCA for a multiset called multiset canonical correlation analysis (MCCA). (Lisanti, Karaman, & Masi(2017)) used MCCA for person reidentification tasks, while (Kanatsoulis, Fu, Sidiropoulos, & Hong (2018)) used MCCA for largescale data, and (Nielsen, (2002)) used MCCA for the GIS system, with discussion of the different restrictions on the MCCA optimization problem.

3. FEATURES OF A TRAJECTORY

Assume that a raw trajectory is represented by a set of points $T = \left\{\left(x_i, y_i, t_i\right)\right\}_{i=0}^{n}$, where the set of times $\left\{t_i\right\}_{i=0}^{n}$ occur in ascending order. Although we can use the method in (K Li, Chen, W Li, He, & Xue (2018)) to model trajectory data that is ordered in time, similar to (Zheng, Li, Chen, Xie, & Ma (2008)) and (Sharma, Vyas, Schieder, & Akasapu (2010)), we will construct several global features in this section. Then, a vector of these features, which is called original features, is used in the experimental section to represent trajectories.