Computer Networks 141 (2018) 157–165

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Multi-model induced network for participatory-sensing-based classification tasks in intelligent and connected transportation systems

Heyuan Shi[a,b,c], Xibin Zhao[a,b,c], Hai Wan[a,b,c], Huihui Wang[d], Jian Dong[e,*], Kun Tang[e], Anfeng Liu[e]

[a] Key Laboratory for Information System Security, Ministry of Education (KLISS), Beijing 100084, China
[b] Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing 100084, China
[c] School of Software, Tsinghua University, Beijing 100084, China
[d] Department of Engineering, Jacksonville University, Jacksonville, FL 32211, USA
[e] School of Information Science and Engineering, Central South University, Changsha 410083, China

ARTICLE INFO

Article history:
Received 21 December 2017
Revised 6 April 2018
Accepted 30 May 2018
Available online 2 June 2018

Keywords:
Intelligent and connected transportation systems
Participatory sensing networks
Internet of things (IoT) applications
Classification,

ABSTRACT

Classification is an emerging Internet of Things (IoT) application in intelligent and connected transportation systems (ICTS). To reduce the costs associated with classification tasks, the concept of participatory sensing has been integrated into ICTS. In this approach, for performing classification tasks, a crowd of participants reports multi-modal data collected via smart devices and manual marking. However, the multi-modal data thus collected are always incomplete. Therefore, integrating these multi-modal data for classification in ICTS is a challenge. In this paper, we propose a multi-model induced network (MMiN) framework for participatory-sensing-based classification tasks in ICTS. We first explore the relationships of the multi-modal data through graph modeling and hypergraph learning. Then, the derived relationships and the multi-modal data are used to train the MMiN. Finally, end-to-end features are derived by the MMiN as the classification results. To evaluate the effectiveness of the MMiN framework, experiments conducted on three visual datasets and one trajectory dataset are reported. The experimental results and comparisons demonstrate the effectiveness of the MMiN framework in exploiting incomplete multi-modal data.

1. Introduction

Classification is an emerging Internet of Things (IoT) application that is widely employed in intelligent and connected transportation systems (ICTS), e.g., for traffic reporting, environmental monitoring and city-scale surveillance. The traditional way of performing classification tasks in transportation systems is by capturing raw data from a fleet of vehicles dedicated to such tasks. The classification results are then manually derived by specially assigned workers based on these raw data. At present, however, for most real-world applications, it is usually impossible to maintain such a dedicated fleet of vehicles and specially assigned workers due to resource limitations and the broad variations in task requirements. To avoid the fuel expenditure associated with dedicated vehicles and other costs for specially assigned workers for classification tasks, the concept of participatory sensing, which is popular in urban planning and traffic monitoring [1], has been integrated into ICTS. In this approach, a task owner recruits a crowd of participants to collect data and perform classification tasks [2] rather than using dedicated vehicles and specially assigned workers.

Thanks to the success of intravehicular devices [3], embedded systems [4] and urban sensing technologies, which are among the most relevant platforms for collecting data in ICTS, connected vehicles possess the ability to collect, access and process various sensor data based on sensing platforms [5]. Thus, participants can collect multi-modal data for classification tasks, e.g., feature data such as raw image data collected by cameras or feature data extracted by smart devices as well as label data created via traditional manual marking. In general, the performance of traditional classification based on single-modal data is limited by the quality of the single-modal data, whereas the use of derived multi-modal data has the potential to reduce the misclassification caused by the use of single-modal data. However, the multi-modal data collected by participants are typically incomplete because of the

* Corresponding author.
E-mail addresses: shy15@mails.tsinghua.edu.cn (H. Shi), zxb@tsinghua.edu.cn (X. Zhao), wanhai@tsinghua.edu.cn (H. Wan), hwang1@ju.edu (H. Wang), dongjian@csu.edu.cn (J. Dong), tk-0206@163.com (K. Tang), afengliu@mail.csu.edu.cn (A. Liu).

https://doi.org/10.1016/j.comnet.2018.05.030
1389-1286/

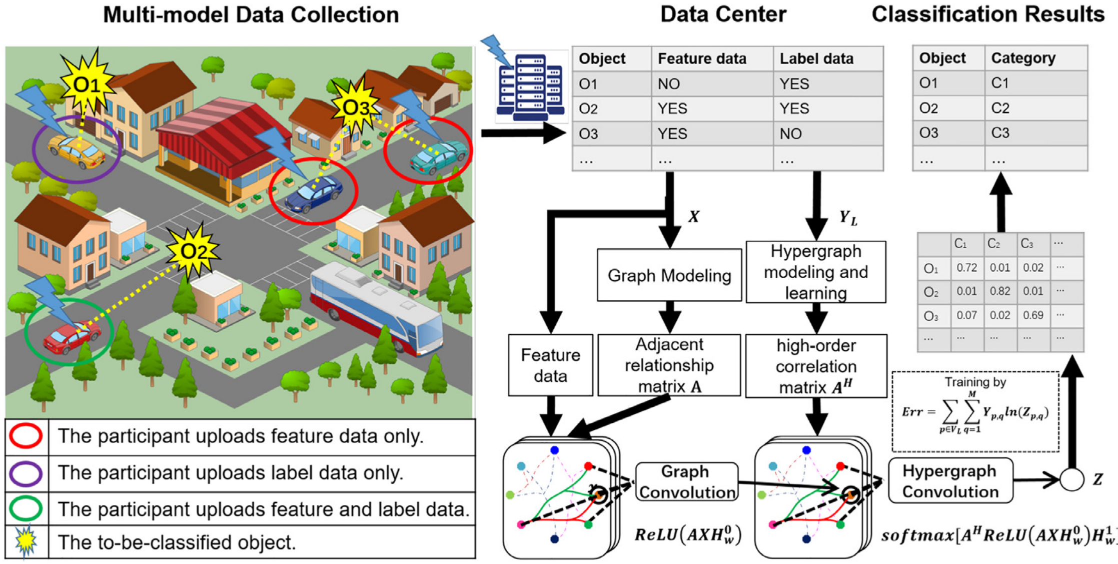

Fig. 1. Illustration of the proposed MMiN framework.

varying sensing capabilities of the participants and the quality of the reported data. For example, some participants may report only feature data (multi-modal data that are missing label data), while others may upload both feature data and manual labels. Therefore, the question of how to exploit such incomplete multi-modal data to achieve improved performance in classification tasks in ICTS is a challenging issue.

In this paper, we propose a multi-model induced network (MMiN) framework for performing classification tasks in ICTS with incomplete multi-modal data. The proposed framework is based on graph modeling, hypergraph learning, and a neural network. An illustration of the MMiN framework is shown in Fig. 1. We first model the incomplete multi-modal data, i.e., feature data collected by smart sensors on vehicles and label data generated by people, to explore the relationships among objects. In this way, the relationships among the feature data of the objects are explored via graph modeling, and the underlying high-order correlations among the label data are formulated via hypergraph learning. Through multi-modal data modeling and learning, the adjacency matrix and high-order correlation matrix are derived. Then, the multi-modal data and the derived relationship description matrices are used to train the MMiN. The correlations among the multi-modal data are used to improve the classification performance. Finally, an end-to-end feature description is derived by the MMiN as the classification result for a participatory-sensing-based task in ICTS. The contributions of this paper can be summarized as follows.

- We formulate the participatory-sensing-based classification problem in ICTS and explore the underlying relationships among multi-modal data via graph modeling and hypergraph learning.
- The multi-model induced network (MMiN) approach, which is a neural-network-based method, is proposed for performing classification with incomplete multi-modal data.
- Experimental results and comparisons based on three visual datasets and the T-Drive trajectory dataset demonstrate the effectiveness of MMiN framework.

The remainder of this paper is organized as follows. We briefly review related work in Section 2. The problem formulation is given in Section 3. We introduce the exploration of the relationships among multi-modal data based on graph modeling and hypergraph learning in Section 4. Then, the MMiN architecture is proposed in Section 5. Experiments, comparisons and related discussions are presented in Section 6. Finally, we conclude this paper in Section 7.

2. Related work

Classification is widely employed in transportation systems, e.g., for traffic congestion detection [6], transportation mode classification [7], traffic sign classification [8], action classification [9] and attribute classification [10]. Recently, with advances in embedded systems [11–13] for IoT applications and cyber-physical systems [14], the integration of IoT and intelligent transportation systems has been explored [15], and classification tasks specific to the scenario of ICTS have begun to be considered.

There are numerous classification methods for specific types of single-modal data, e.g., image and temporal series data, such as support vector machine (SVM) classification [16], graph-based methods [17], Bayes models [18], decision trees [6], and neural-network-based methods [19]. For classification based on single-modal data, Mallah et al. [6] proposed a decision-tree-based framework for real-time traffic congestion classification in vehicular ad hoc networks (VANETs) based on extracted temporal features called BEACON features, which contain basic vehicle state information with time stamps. Yang et al. [8] proposed a detection module based on single-visual-feature data for traffic sign proposal extraction and classification via a color probability model. Kubera et al. [16] investigated the automatic classification of vehicle types based on audio signals using decision trees, random forests, artificial neural networks, and support vector machines. Dong et al. [19] used a semi-supervised convolutional neural network based on single-view feature data, i.e., vehicle frontal-view images, for the vehicle type classification task. For road marking classification, Paula and Jung [18] exploited visual data captured by an onboard camera and a Bayesian classifier based on mixtures of Gaussians to classify lane markings.

With the diversity of sensors available for data collection due to the success of ICTS, data collectors possess the ability to derive multi-modal data. Many works have focused on classification based on data collected from multiple sensors. Garcia et al. [20] proposed a complete perception fusion architecture for object detection that integrates the composite representation and uncertainty management of multi-modal data from smart devices, i.e., radar, lidar, and camera data. Ma et al. [21] exploited wireless accelerom-