Cities 63 (2017) 118–127

Contents lists available at ScienceDirect

Cities

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

Data-enabled public preferences inform integration of autonomous vehicles with transit-oriented development in Atlanta

Zhongming Lu [a,*], Rundong Du [b,c], Ellen Dunham-Jones [d], Haesun Park [c], John Crittenden [a]

[a] Brook Byers Institute for Sustainable Systems, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States
[b] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, United States
[c] School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States
[d] School of City & Regional Planning, College of Architecture, Georgia Institute of Technology, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Received 1 September 2016
Received in revised form 23 December 2016
Accepted 15 January 2017
Available online xxxx

Keywords:
Autonomous vehicle
Social preference
Topic modeling
Sentiment analysis
Latent-class choice model
Decision support

ABSTRACT

Autonomous vehicles (AVs) have emerged as a transformative technology with the potential to both fundamentally improve lives in cities but also to exacerbate suburban sprawl, vehicle miles traveled and the associated greenhouse gas emissions. Are communities willing to adopt best practices that can lead to early adoption of more sustainable outcomes? This paper presents innovative means to analyze social preferences, demand for AVs, and the potential to resolve community concerns with integrated solutions. We discuss our comprehensive analysis of unstructured and structured data from a survey on AVs that was conducted by the Atlanta Regional Commission in 2015. We used topic modeling to synthesize the "topics" from 1540 comments. The topics captured Atlanta residents' concerns and suggestions about implementing AVs. Further, sentiment analysis revealed people's attitudes on the topics. Accordingly, we proposed an integration of AVs and transit-oriented development (TOD: the development of compact and mixed-use communities around high quality mass transit services within a 10-min walking distance). The second type of data is people's responses to multiple-choice questions about AVs and TOD, which we call structured data. Using latent-class analysis, we identified heterogeneity in preferences for AVs and TOD. More Atlanta residents are willing to live in transit-oriented communities than traditional automobile-dependent ones if AVs save time and improve productivity. This finding portends the future success of combining AVs with TOD and reaping the sustainable benefits of this transformative technology.

1. Introduction

AVs are one of the transformative technologies that has emerged over the past few years (Burns, 2013). If AVs become mature, they could fundamentally change the way people live, work and travel in cities (Economist, 2015; D. J. Fagnant & Kockelman, 2014). The application of AVs could also change the way we build cities as well as how we manufacture cars (Daniel J Fagnant & Kockelman, 2015; Zhang, Guhathakurta, Fang, & Zhang, 2015). They could help mitigate local transportation problems, improve regional quality of life, decrease the demand for parking, and strengthen regional competitiveness, or they could exacerbate suburban sprawl, vehicle miles traveled and the associated greenhouse gas emissions (Litman, 2015; Wadud, MacKenzie, & Leiby, 2015). Are communities willing to adopt best practices that can lead to early adoption of more sustainable outcomes? Encouragement of the early and fast adoption of the best practices requires the development of proper policy incentives and increases in amenities. For example, Georgia has the largest share of electric vehicles in the US as compared to new vehicle sales. The greater adoption is a result of a generous $5000 tax credit that can be applied for up to five years of state tax filings or until the credit is depleted. However, AVs are less mature than electric vehicles and still have liability and functional issues to resolve (Campbell, Egerstedt, How, & Murray, 2010). There are only a few studies on the transport and environmental implications on AVs under different operation scenarios (e.g., AV taxis (Greenblatt & Saxena, 2015), and combination of AVs and on-demand mobility services (Greenblatt & Shaheen, 2015)). The development of AV technology is in its early stage and the lack of documented sustainability performance and lack of actual market acceptance data of different AV scenarios makes it difficult to inform policy making.

Successful policy development to incentivize the adoption of best practices requires the understanding of social preference and demand for AVs. Cities are complex adaptive systems and properties emerge from the millions of decisions (e.g., where to live, where to work and where to shop) and interactions (e.g., commute to work, housing location choice and business investment) between industries, citizens and infrastructure (Pandit, Lu, & Crittenden, 2015). As we construct urban infrastructures (e.g., buildings, roads, energy and water/wastewater provision), large scale properties emerge including land use pattern,

* Corresponding author.
E-mail address: zhongming.lu@gatech.edu (Z. Lu).

http://dx.doi.org/10.1016/j.cities.2017.01.004
0264-2751/

air quality, quality of life and carbon footprints. Some of the emergent properties are unintended adverse consequences, including bad air and water quality, urban flooding, traffic congestion, expensive per capita maintenance costs, social segregation, and sedentary lifestyles that contribute to poor public health. In sprawling areas laid out to require heavy reliance on automobiles, many of these adverse consequences result from decisions that benefit individuals but burden the collective. Whether AVs can make cities more sustainable will be determined by the types of sustainable amenities AVs provide to individual households and the policy developments that cities implement to incentivize their adoption (Bansal, Kockelman, & Singh, 2016). In order to create more desirable amenities that AVs can serve, we should identify synergy effects though the combination of other innovative urban design strategies and technologies. TOD is the one we identified in this study and we will illustrate how we identified this opportunity through the exploration of citizens' preferences.

There are two primary data sources to explore people's preferences and desires that can help with AV technology development and policy making. The first source comes from questionnaires. In a questionnaire, people are asked to respond to a series of questions that can capture individual preferences so that models can be developed to predict decision making. Questionnaire data collected with a structured format for model development, such as multiple choice, is called structured data. The information derived from the structured data is limited by the types and numbers of questions. The second source comes from people's comments and discussions. Preferences are embodied in these comments and discussions. However, comments and discussions do not have any specific format for analysis and modeling. Thus, we call this type of data source unstructured data. There are a vast source of comments and discussions, especially on social media (e.g., Twitter, Facebook and Flickr) and online shopping stores (e.g., Amazon). By analyzing the concerns and suggestions behind these comments and discussions, we can learn how to (re)design the technology, product and service. However, comments and discussions cannot help with choice modeling and market evaluation. Accordingly, we need a combination of both structured and unstructured data to explore people's preferences and desires that help with AV technology development and policy making (Fong, Hettinger, & Ratwani, 2015).

In 2015, the Atlanta Regional Commission (ARC) published a survey focusing on future transportation technologies including AVs that might change the way people live, work and travel in metro Atlanta. In the survey, a set of questions was asked to measure Atlanta residents' preference for AVs. The answers to these questions were used for preference modeling. Meanwhile, residents were asked to make comments on AVs. These comments make up unstructured data to analyze people's concerns and suggestions on AV implementation. In this study, we first used topic modeling to synthesize the "topics" from the unstructured comments. We also conducted sentiment analysis to obtain people's attitudes associated with specific topics. To address the concerns and suggestions in the synthesized "topics", we proposed a strategy to implement AVs in metro Atlanta. Further, we adopted latent class analysis to develop different classes with distinct preferences and relocation decisions associated with AVs. Because of the interrelationship between housing location and transportation, we also conducted the latent class analysis on heterogeneity in preferences for housing locations. According to individual preferences for AV and housing location, we assessed the potential for Atlanta residents to adopt AVs.

2. Methods

2.1. Data

The ARC survey was open to respondents from January 9, 2015 through March 31, 2015, which focused on future transportation options. Questions about AV tested whether residents were familiar with AV technology and sought to determine whether residents saw AVs as being able to address some of the region's transportation challenges (e.g., traffic congestion, lack of options for older and disabled people). Residents were also asked to leave comments after each question for AVs. Besides the questions and comments on AVs, the survey also asked residents about their preference for TOD. These questions sought to determine the importance of transit options in housing location choice and regional economic growth. The summary of the questions is provided in Tables A.1 & A.2. About 6300 respondents answered the survey, well-representing the residences in the region. Demographic data were collected including age, gender and race/ethnicity and the statistics compared to the census show that the survey's respondents are more whites and residents under 45 years old are underrepresented. The report of full questions and statistics can be found in the ARC Regional's Plan Survey Report (ARC, 2015).

2.2. Topic modeling

Residents provided 1540 comments in total on AVs. Topic modeling enables the classification of thousands of comments into several representative topics, so that such large number of comments could be interpreted and understood. This is an extension of frequency analysis, allowing for the interpretation of a larger number of documents. Nonnegative matrix factorization (NMF) has shown an excellent performance in document clustering and topic modeling (Kuang & Park, 2013; Xu, Liu, & Gong, 2003). The NMF is formulated as below:

$$\min_{W,H\geq 0} A-WH_F^2 \tag{1}$$

where the data is encoded as column vectors of the matrix $A\in\mathbb{R}_+{}^{m\times n}$, $W\in\mathbb{R}_+{}^{m\times k}$, and $H\in\mathbb{R}_+{}^{k\times n}$. Typically, k ≪ min (m, n). The matrix A is a term-document matrix, of which each column is a term-frequency vector to represent each document (Manning, Raghavan, & Schütze, 2008). The columns of W naturally become the representative vectors of the generated clusters (i.e., topics) and the values in each column of H are actually cluster indicators, as illustrated in Fig. 1. The optimization in Eq. (1) is actually approximating A's columns (which represents the documents) with nonnegative linear combinations of columns of matrix W.

In this study, we use an efficient hierarchical document clustering method (HierNMF2) based on a rank-2 NMF (i.e., k = 2). The HierNMF2 is be a very fast and high quality topic modeling algorithm (Kuang & Park, 2013). As a divisive clustering algorithm, HierNMF2 has two key components: (1) recursively splitting clusters using rank-2 NMF: once we obtain two clusters through rank-2 NMF, we recursively apply the same procedure to discovered clusters, splitting them into smaller ones and obtaining the desired number of clusters.; (2) a topic-aware criterion of choosing a cluster to split, which enhanced the topic quality: in each recursion step, the HierNMF2 algorithm will choose the best cluster to split, in the sense that the two new clusters are most well separated.

2.3. Sentiment analysis

There are many published works on sentiment analysis (Pang & Lee, 2008; Ravi & Ravi, 2015). However, those methods usually require a training data set that is either from a broad variety of sources or from a data source that is similar to the test data set. We chose a third-party Application Programming Interface (API) hosted by MeaningCloud LLC., because we think a commercial service provider should have access to more training data sets for sentiment analysis. For example, they can buy commercial data sets or hire people to label the data. The information about the API can be found on http://www.meaningcloud.com/products/sentiment-analysis.