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.
© 2017 Elsevier Ltd. All rights reserved.

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/© 2017 Elsevier Ltd. All rights reserved.

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.

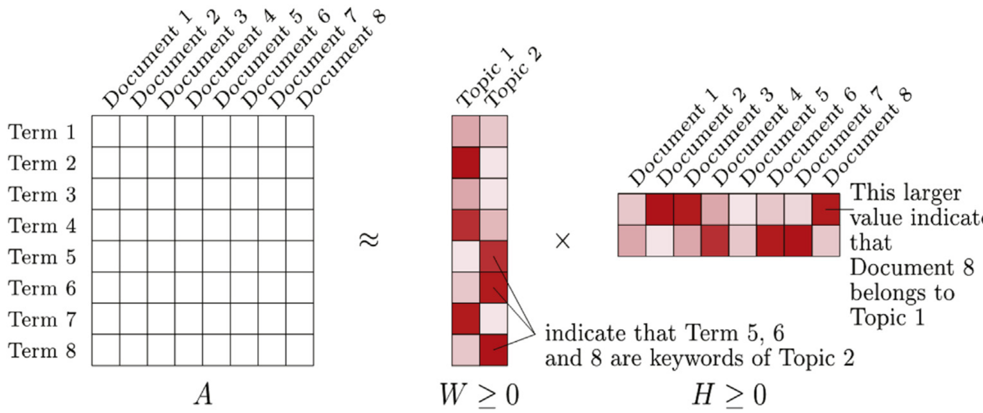

Fig. 1. Illustration of nonnegative matrix factorization (NMF) for topic modeling.

2.4. Latent class analysis

Individual responses to the questions in Tables A.1 & A.2 reflected personal preferences for AVs and TOD. We used 5235 complete responses to conduct the latent class analysis. We employed a latent class analysis to identify the optimal number of classes that exhibit a statistically significant difference in preferences for AVs and TOD, respectively. The probability formulation of individual *i* giving response *m* to question *t* (Vermunt & Magdison, 2005) is,

$$P(y_{it} = m|z_i) = \sum_{x=1}^{X} P(x|z_i)P(y_{it} = m|x) \tag{2}$$

where $P(y_{it} = m|z_i)$ is the probability of individual *i* giving response *m* to question *t*; *m* is the ordinal number representing the order of each option in question *t* (see Appendix Tables A.1 & A.2); $P(x|z_i)$ is the probability of individual *i* belonging to a certain class *x*; z_i represents socioeconomic characteristics of individual *i*; *x* is the latent class membership; and $P(y_{it} = m|x)$ is the class-specific conditional probability of individual *i* giving response *m* to question *t*, which is determined by Eq. (3).

$$P(y_{it} = m|x) = \frac{\exp\left(\eta^t_{m|x}\right)}{\sum_{m'=1}^{M} \exp\left(\eta^t_{m'|x}\right)} \tag{3}$$

where $\eta^t_{m|x}$ donates individual *i*'s utility associated with the response of *m* to question *t*, and

$$\eta^t_{m|x} = \beta^t_{m0} + \beta^t_{mx0} \tag{4}$$

where β^t_{m0} is the intercept; and β^t_{mx0} is the effect of the classes on the utility value, with restrictions $\sum_{x=1}^{X} \beta^t_{mx0} = 0$ and $\sum_{m=1}^{M} \beta^t_{mx0} = 0$.

The probability of individual *i* belonging to class *x* is parameterized as follows:

$$P(x|z_i) = \frac{\exp\left(\eta_{x|z_i}\right)}{\sum_{x'=1}^{M} \exp\left(\eta_{x'|z_i}\right)} \tag{5}$$

where $\eta_{x|z_i}$ donates the likelihood of belonging to one of the classes (Eq. (6)).

$$\eta_{x|z_i} = \beta_0 + \sum_{p=1}^{P} \beta_p \times z_{ip} \tag{6}$$

where β_0 is the intercept; β_p are the effects of socioeconomic variables on class membership; and z_{ip} are individual's socioeconomic variables (see Tables A.1 & A.2).

We used Latent Gold Choice 5.0 Software (Statistical Innovations Inc., Belmont, MA, USA) and Maximum Likelihood (ML) method to estimate the segmentation of classes. The determination of optimal class number is based on Bayesian Information Criterion (BIC) and classification errors (Lu, Southworth, Crittenden, & Dunham-Jones, 2015). A smaller BIC indicates a better model fit with a smaller number of parameters to be estimated. The statistic results is provided in Appendix Tables A.3 & A.4.

3. Results

3.1. Synthesized "topics" on AVs

Fig. 2 displays a category hierarchy to classify the synthesized topics. One of the two primary categories of the topics in the 1540 comments on AVs is "the threat to public transportation investment". In the survey, participants expressed their belief that connections with regional transit network essential for existing/future job centers to grow and be successful. If public funds are invested in AVs, there is concern about the threat to the funding and the development of existing public transportation systems. Here are two examples of what comments are characterized as AVs threaten public transportation investment. "I am concerned that resources may be diverted from public transportation or other more affordable options into autonomous vehicles." "I feel there are other advances we need to put first - transit, walking, biking. We need to make these a priority."

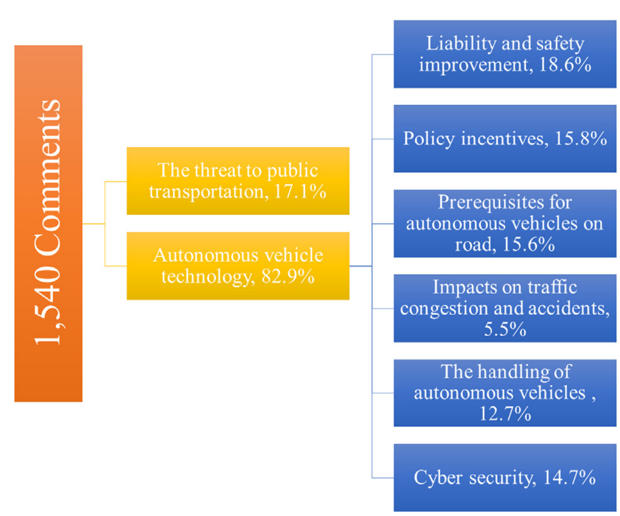

Fig. 2. The classification of synthesized "topics" on AVs: percentage refers to the proportion of comments belonging to each topic.

The second primary category labeled "AV technology" is mainly about public concerns and suggestions for operating AVs. As shown in Fig. 2, we identified six subcategories of topics. In each subcategory, the specific topic is composed of five key words. The first subcategory is about the liability and safety improvement. As shown in Table 1, specific topics include "liability, issue, concern, address, safety", "embrace, future, technology, fight, improve", "early, start, incentive, provide, adopt", and "technology, safety, prepare, vehicle, ready". Although five key words cannot embrace all the contents of the comments, they provide a best characterization of the central ideas embodied in the comments. For example, the five key words "early, start, incentive, provide, adopt" implies the discussion of the early start of providing incentives to increase the adoption. The number of comments on each topic in this subcategory implies more concern about liability than enthusiasm for adoption (see Table 1).

The second subcategory is about the call for policy incentives, which residents think is important to a successful application of AVs. First of all, people discussed the leadership of Georgia in AVs (topics: "Georgia, lead, future, leading, innovation" and "GA, area, lead, tech, state"). The call focuses on tax legislation (topic: "support, tax, funding, legislation, vote") and regulations (topic: "law, regulation, enact, support, government") to incentivize the adoption and business development for AV industries in Georgia (topic: "state, advance, working, government, business"). Citizens also discussed the possibility of providing affordable testing places to test AVs (topic: "testing, place, affordable cost, implement").

The third subcategory is about the prerequisites for AVs on roads. The most discussed topic is safety and governmental support of AVs (topics: "safety, vehicle, driver, provide, concern" and "autonomous, vehicle, car, safe, support"). Citizens also felt that passengers should be comfortable while riding in AVs (topic: "vehicle, feel, driving, liability, comfortable"). Another prerequisite is the usefulness of AVs for commuting (topic: "work, road, period, live, home"). Citizens also discussed the opening of a special lane similar to the high-occupancy vehicle lane (topic: "lane, special, highway, designate, HOV") and insurance for AVs.

The fourth subcategory is the impacts on traffic congestion and accidents. The topics included a discussion of the current traffic conditions in metro Atlanta (topic: "metro, Atlanta, small, town, accident") and the mitigation of traffic congestion and accidents by introducing AVs (topics: "congestion, reduce, traffic, accident, improve" and "traffic, autonomous, problem, alleviate, vehicle").

The fifth subcategory is the handling of AVs. People viewed AVs as an alternative commuting mode (topic: "drive, person, idea, work, people"), especially for older and disabled citizens (topic: "elderly, handicap, disabled, condition, disability"). The most discussed topic was concern about the speed and control of AVs (topic: "people, driving, control, time, speed"). Citizens suggested that there is a need for more research to inform the decision making (topic: "make, decision, people, informed, research"). The survey shows that current public opinion believes that AVs could be a good way to solve the congestion problem (topic: "good, thing, congestion, work, problem"). However, citizens do not think car lovers will choose AVs due to the lack of a "high" feeling of speed and control (topic: "idea, great, bad, car, love").

The last subcategory is the cyber security of AVs. Citizens are concerned about computer hacking during driving (topic: "car, driverless, computer, hack, driving"). For example, citizens do not wish to be taken to other unknown places by AVs. Also citizens are concerned about the errors that may be present in the AV hardware and software (topic: "human, error, situation, computer, safe"). The last topic citizens discussed is the interaction between AVs and human driven vehicles on the roads.

Overall, these six subcategories summarize a set of technical barriers and policy requirements to drive a high adoption of AVs. The summary shows a need for great efforts from automotive manufacturers, information technology developers, policy makers, transit agencies and city planners. Automotive manufacturers and information technology developers should solve the challenges of liability, comfort, controllability and cyber security. Policy makers should develop policies and incentives to encourage the early adoption of AVs and provide testbeds for running AVs. City planners and transit agencies should (re)design the land use and transportation system in order to satisfy commuting while expanding mobility, and mitigate traffic congestion and excessive increases in vehicle miles traveled. The collaboration among these stakeholders is critical during the introduction of AVs.

Table 1
Summary of key words for each "topic" in the category of "AVs".

Summary of key words					
Liability and safety improvement, 286 comments					
o Liability o Issue o Concern o Address o Safety	o Embrace o Future o Technology o Fight o Improve		o Early o Start o Incentive o Provide o Adopt	o Technology o Safety o Prepare o Vehicle o Ready	
77 comments	21 comments		20 comments	156 comments	
The call for policy incentives, 244 comments					
o Law o Regulation o Enact o Support o Government	o Georgia o Lead o Future o Leading o Innovation	o Support o Tax o Funding o Legislation o Vote	o GA o Area o Lead o Tech o State	o State o Advance o Working o Government o Business	o Testing o Place o Affordable o Cost o Implement
95 comments	84 comments	15 comments	13 comments	30 comments	7 comments
Prerequisites for autonomous vehicles on roads, 240 comments					
o Lane o Special o Highway o Designate o HOV	o Safety o Vehicle o Driver o Provide o Concern	o Work o Road o Period o Live o Home	o Insurance o Year o Cost o Infrastructure o Company	o Vehicle o Feel o Driving o Liability o Comfortable	o Autonomous o Vehicle o Car o Safe o Support
24 comments	50 comments	20 comments	22 comments	21 comments	91 comments
Impacts on traffic congestion and accidents, 84 comments					
o Congestion o Reduce o Traffic o Accident o Improve		o Metro o Atlanta o Small o Town o Accident		o Traffic o Autonomous o Problem o Alleviate o Vehicle	
33 comments		15 comments		36 comments	
The Handling of autonomous vehicles, 196 comments					
o Make o Decision o People o Informed o Research	o People o Driving o Control o Time o Speed	o Elderly o Handicap o Disabled o Condition o Disability	o Drive o People o Idea o Person o Work	o Good o Thing o Congestion o Work o Problem	o Idea o Great o Bad o Car o Love
22 comments	52 comments	18 comments	45 comments	20 comments	39 comments
Cyber security, 226 comments					
o Car o Driverless o Computer o Hack o Driving		o Human o Error o Situation o Computer o Safe		o Driver o Car o Road o Regular o Driving	
136 comments		38 comments		43 comments	

3.2. Sentiment analysis on "topics" for social preferences

Topic modeling can help understand what people said about AVs. By conducting sentiment analysis, we can determine public attitudes towards AVs. When people discussed whether Georgia should lead future innovations (topic: "Georgia, lead, future, leading, innovation), 43% of the statements show a positive attitude as compared to 13% negative. It implies public support of Georgia in leading future innovations, including AVs. Correspondingly, we found that 57% of the statements are positive when people talked about whether the government should provide incentives to encourage the early adoption (topic: "early, start, incentive, provide, adopt"). In terms of traffic conditions in metro Atlanta and the surrounding small towns (topic: "Atlanta, metro, accident, small, town"), we found that 34% of the statements are negative, which is the highest in terms of the percentage of negative statements among all the topics. Meanwhile, 52% of the statements are positive about traffic improvement (topic: "congestion, reduce, traffic, accident, improve"). These results indicate a strong desire and passion from local residents to improve traffic conditions in metro Atlanta. However, everything is not black and white and not all topics have a sentiment. Overall, a gain of public attitudes and social preferences could provide a scientific basis to improve decision making.

3.3. Integration of AVs with TOD

According to the topic modeling and sentiment analysis, we conclude that there is a potential for introducing AVs in metro Atlanta. Meanwhile, we should address the concerns identified from topic modeling, which are especially critical to the early adoption of AVs. These concerns include liability, controllability, comfort and cybersecurity of AVs. More importantly, people feel concerned about the threat to public transportation systems from AVs. We propose an integration of AVs with TOD (Fig. 3) to help resolve this concern and coordinate the development of AVs and public transportation systems. Currently, the Metropolitan Atlanta Rapid Transit Authority or MARTA is operating a transit system in Atlanta, which is ready to provide a testbed to link AVs to transit. AVs owned by MARTA or other entities would bring people from where they live to transit stations and take people from transit stations to their final destinations. In this case, the investment in AVs is considered as a part of public transportation improvement. The sharing can maximize the utilization of automation and reduce the number of vehicles on roads as well as the likelihood of traffic congestion, especially at peak hour (D. J. Fagnant & Kockelman, 2014).

The integration of AVs with TOD can also boost the investment returns on transit systems. According to the design criteria of TOD, the typical radius that a transit station can serve is 0.5 mile, or a 10-min walking trip. By taking AVs, the radius can be significantly expanded. Currently, the MARTA transit system serves 25.5 square-mile areas and 111,000 residents within 0.5-mile TOD radius. If we adopt AVs to extend the TOD radius to 2 miles, the areas and residents served by the MARTA transit system will increase to 169.4 mile^2 and 606,000 residents, respectively (Fig. 4). An increase in ridership of MARTA by the same factor of 6 will fundamentally convert MARTA from the most carbon-intensive mode to the least one in Atlanta (Azevedo, Bras, Doshi, & Guldberg, 2009). It means that fewer stations are required in future, hence reducing the need for expansion of the transit system. Also the land available for higher-density (re)development around the transit stations is increased as well, which means more property tax revenue to maintain and improve the integrated AV and transit system. This scenario also increases the opportunities for households to live without owning a private car, which both increases affordability and matches the lifestyle preferences of the Millennial Generation.

3.4. Market potential assessment of the integrated AVs and TOD

Whether Atlanta residents will choose the integrated AVs and transit system for transportation and live in transit-oriented communities depends on individuals' preferences for AVs and TOD. According to the latent class analysis on people's attitude towards AV, we found three classes that have significant heterogeneity in preference for AVs. There is one class of individuals that believe that AVs will be viable in the next couple of decades. This class feels comfortable when transferring driving control to a fully AV. If AVs lead to less congestion and allow people to be productive while riding, this class is likely to consider relocating in metro Atlanta. Moreover, this class considers AVs as a viable option for those who cannot drive themselves and Georgia should support the implementation of AVs. We defined this class as “AV supporting class”. The second class is uncertain whether AVs will be viable in the next couple of decades and whether Georgia should support the implementation. But the second class feels comfortable with the idea of taking AVs and will consider relocating if AVs lead to less congestion and allow people to be productive while riding. Thus, we defined the second class as “AV uncertain class”. The third class is the opposite of the first class. The third class does not believe that AVs will be viable in the next couple of decades. This class also feels uncomfortable about taking AVs and is unlikely to relocate. We defined the third class as “AV unconvinced class”.

According to latent class analysis on people's attitude towards TOD, we found four classes for heterogeneity in housing preference. The first and second classes prefer TOD. However, the first class supports more development and employment opportunities in the less-developed south metro region of Atlanta than the second class does. The third and fourth classes prefer automobile-dependent communities but the third class supports more development and employment opportunities in the south metro region of Atlanta than the fourth class does.

As mentioned before, the AV “supporting” and “uncertain” classes are likely to consider the relocation if AVs lead to less congestion and allow them to be productive while riding. Whether these two classes will choose houses in transit-oriented or automobile-dependent communities depends on their housing preference. We know the percentage of survey respondents who belong to the AV “supporting”, “uncertain” classes and prefer TOD from the latent class analysis. However, the sample of survey respondents cannot represent the residents in metro Atlanta because of sample bias in demographics. To correct the bias, we adjusted the weight of each survey respondent to calculate the percentage (see Appendix: Weight adjustment). We found that there are 39.3% of residents in metro Atlanta who belong to the AV “supporting”, “uncertain” classes and prefer transit-oriented communities. It is higher than the percentage of residents (25.1%) who belong to the AV ‘supporting”, “uncertain” classes but prefer automobile-dependent communities (Fig. 5). The comparison indicates that the implementation of AVs could lead to more TOD than automobile-dependent sprawling in

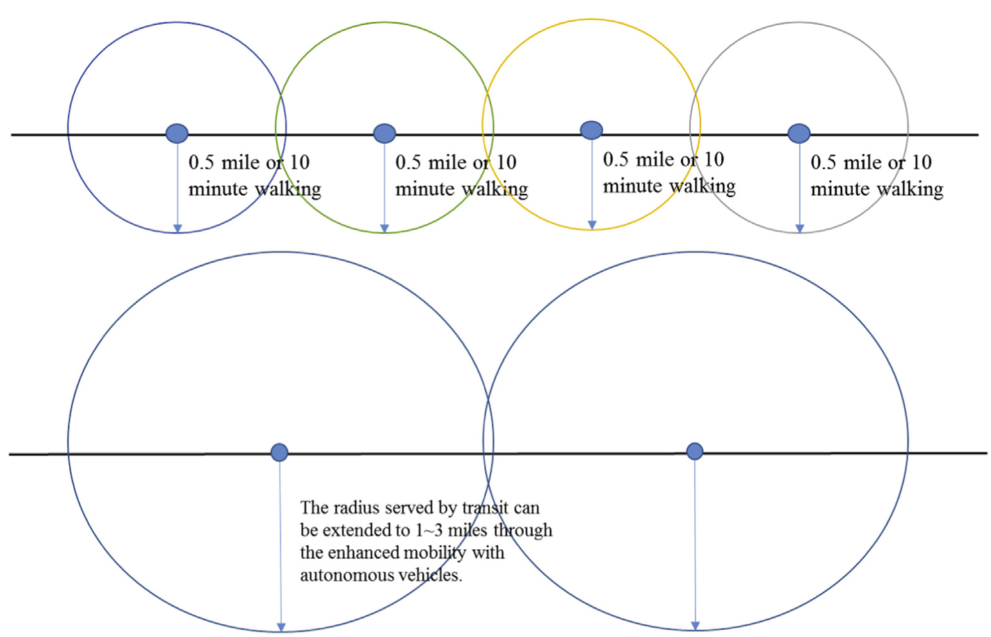

Fig. 3. The integration of AVs and TOD, informed by public preference.

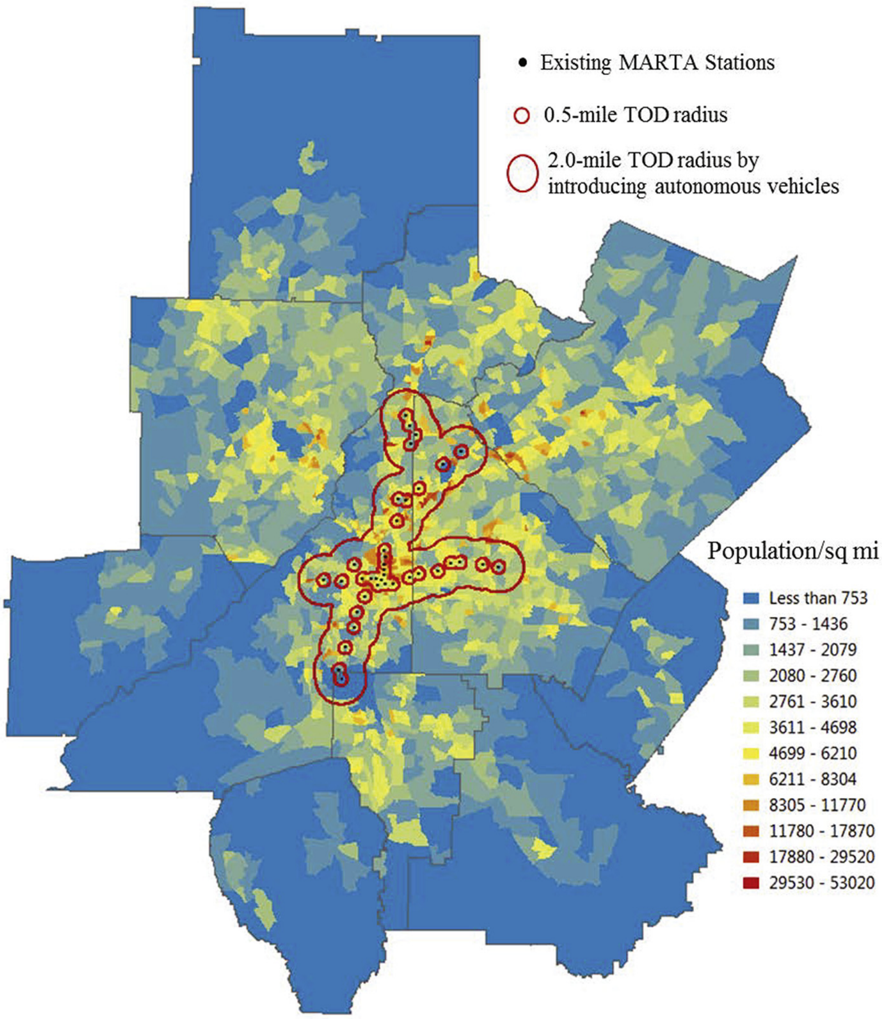

Fig. 4. Expansion of the areas and population by increasing the TOD radius of existing MARTA stations from 0.5-mile to 2.0-mile through the integrated AVs and transit system.

metro Atlanta. The market potential suggests that the integration of AVs with TOD is a feasible and proper strategy for the early implementation of AVs. However, it should be noted that we still need to confirm the time savings and productivity benefits of integrating AVs and TOD before public support is forthcoming.

4. Discussion

Although there is no direct evidence from previous studies to support the integration of AVs and TOD in Atlanta, a study about the impact of car sharing on public transit shows that car-holding households are more likely to shift towards public transit across United States (Martin & Shaheen, 2011). AVs are designed for sharing as well, which will be more flexible in time scheduling than current platform such as Uber. So the integration of AVs and TOD is likely to boost transit ridership. However, we need to emphasize that our analysis and results were built on Atlanta, where people strongly need a better public transportation system. How to best use AVs in other regions will depend on local conditions (e.g., existing transportation infrastructure, travel demand and socioeconomic status).

In this study, the comments in the survey for topic modeling were limited by the questions that were asked as well as the number of respondents. For example, the ARC did not ask who should own the AVs. Should individuals or government own the vehicles? In fact, the ownership is a critical question which affect the policies to incentivize the use of AVs. In the future when AVs become mature, the government will need to compare the cost effectiveness and sustainability impacts of individual ownership versus publicly operated AVs. This can be addressed using larger data sets from social media from which the preferences of millions of people can be assessed.

Understanding the preference heterogeneity in AVs and TOD using the latent-class analysis is also limited by the survey in collecting social-economic data. We did not consider whether people who live near MARTA are more or less favorable to the combination of TOD and autonomous vehicles. To the extent that Atlanta is highly segregated by race and income, the geographic biases are likely to persist without a more careful consideration of how the responses are weighted. Meanwhile, due to the sample bias in ARC survey, we can overestimate/underestimate the preference by modeling the sample as a homogeneous group. The latent-class analysis takes account of the heterogeneity in preference. However, the sample bias cannot be fully eliminated using latent-class analysis. A statistically valid random sample is required in the future.

It should be noticed that we are not arguing that the integration of AVs with TOD is the best strategy to follow. AVs could potentially be seen as alternatives to public busses, which would negatively affect public transport investments. Instead, we provided innovative means to inform initial conceptual designs by analyzing unstructured and structure data. The proof of the conceptual design needs more validations. Future studies should address whether AVs can lead to less congestion and allow people to be productive while riding, especially when we integrate AVs, autonomous busses, and transit systems.

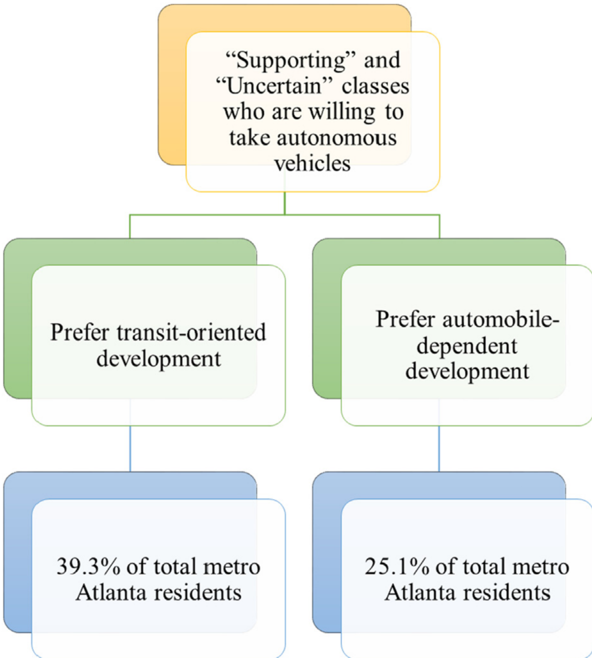

Fig. 5. Comparison of the demand for transit-oriented versus automobile-dependent development while using AVs for commuting in the metro Atlanta.

There have been a few studies on the travel, land use, and environmental implications of shared AVs. These studies run the simulations using hypothetical landscapes. The simulation itself does not include the dynamics of traffic flows and model the interaction between autonomous and human driving vehicles. In the future, simulation and analyses should be further explored that build upon real cities to evaluate and validate the best practices for running AVs.

Following model analysis to prove transport efficiency and land use serviced by the integrated AVs and TOD, another survey is required in the future to present this study"s analysis to the public. They can then decide how AVs and TODs should be integrated. While attitudes shape behavior, they do not determine it, especially considering the "Not In My Back Yard" bias (i.e., an opposition by residents to a proposal for a new development because it is close to them). Using these calibrated choices, it is possible to quantify the change in land use and environmental impacts to supplement our analysis of the expansion of the community serviced by integrating AVs and TOD. Moreover, financing and policy incentives should also be studied to further drive adoption, especially since the potential benefits are significant.

5. Conclusion

Urban infrastructures play a critical role in supporting human activities in cities. Technological options are rapidly becoming available, which allows us to build and retrofit cities so that our cities can be more sustainable and resilient. However, it is especially challenging to increase early adoption of innovative technologies. Herein, we demonstrated a comprehensive analysis including topic modeling and sentiment analysis that are based on citizen's comments, and latent-class modeling on heterogeneous preferences. The analysis enables the discovery of social preferences and demand for innovative technologies. As an example, we proposed and validated a strategy for implementing AVs in Atlanta that was informed by public preference. In the future, modeling and testing will be required to verify the performance of this integration.

Acknowledgements

Financial support from the National Science Foundation through grants (#0836046 and #1441208), is gratefully acknowledged. This work was also partially supported by Brook Byers Institute for Sustainable Systems, the Hightower Chair and Georgia Research Alliance. The authors would like to thank anonymous reviewers for the comments and suggestions. The views and ideas expressed herein are solely of the authors and do not represent the ideas of the funding agencies in any form.

Appendix A. Weight adjustment

Because of sample bias, the respondent sample cannot represent the demographic data. Thus, we weighted individual responses to adjust the ratio of different socioeconomic groups (i.e., age, ethnicity and gender) (Eq. (A.1)).

$$W_{(age_i, ethnicity_i, gender_i)} = \frac{P_{(age_i, ethnicity_i, gender_i)\_pums}}{P_{(age_i, ethnicity_i, gender_i)\_survey}} \tag{A.1}$$

where, $W_{(age_i, ethnicity_i, gender_i)}$ is the weight of the group belonging to the level i of age, ethnicity and gender; $P_{(age_i, ethnicity_i, gender_i)\_survey}$ is the percentage of the group belonging to the level i of age, ethnicity and gender in the sample; and $P_{(age_i, ethnicity_i, gender_i)\_pums}$ is the percentage of the group belonging to the level i of age, ethnicity and gender in the Public Use Microdata Sample (PUMS) data. The PUMS is a subsample of individual person and housing unit records (e.g., sex, education, and employment status) from American Community Survey (ACS). The PUMS is considered as an unbiased sample of individuals in the metro Atlanta.

Table A.1
Questions on autonomous vehicles.

Questions as indictors	Options and value assignment
Questions and responses to model individual's preference for autonomous vehicle	
Do you believe autonomous vehicles will become a realistic option within the next couple of decades?	Yes = 1 Unsure = 0 No = −1
If autonomous vehicles led to less traffic congestion and also allowed you to be productive while riding instead of driving, how likely would you be to move to another part of the region?	Very = 2 Somewhat = 1 Unsure = 0 Not at all = −1
How comfortable would you be in transferring driving control to a fully autonomous technologically advanced vehicle?	Definitely = 2 Some = 1

Table A.1 (*continued*)

Questions as indictors	Options and value assignment
	Unsure = 0 Not at all = −1
Do you see driverless cars as a viable option for people who cannot drive themselves such as older adults or people with disabilities?	Yes = 1 Unsure = 0 No = −1
Technology to implement autonomous vehicles will be available in the next 5–10 years; however liability issues may delay their introduction. Should the State of Georgia support the implementation of autonomous vehicles, including enactment of laws and regulations encouraging this technology?	Yes = 1 Unsure = 0 No = −1
Socioeconomic variables for class membership modeling	
Have you heard of autonomous vehicles ("driverless cars")?	Yes = 1 Unsure = 0 No = −1
Do you own a smartphone or plan on purchasing one in the near future?	Yes = 1 Unsure = 0 No = −1
Age	0–18: 0 19–24: 1 25–34: 2 35–44: 3 45–54: 4 55–64: 5 65–74: 6 75+: 7
Gender	Female: 0 Male: 1
Race/Ethnicity (modeled as a categorical variable)	White/Caucasian Black/African-American Asian Hispanic/Latino Two or more races Some other race

Table A.2
Questions on transit-oriented development.

Questions as indicators	Options and value assignment
Questions and responses to model individual's preference for transit-oriented development	
Should the Atlanta region strive to achieve a more equitable distribution of economic opportunities by placing greater emphasis on developing new employment centers in the southern, eastern and western parts of the region?	Yes = 1 Unsure = 0 No = −1
Do you believe connections with a regional rail and/or express bus network will be essential for existing job centers and potential new ones to grow and be successful in the future?	Yes = 1 Unsure = 0 No = −1
How important is it for the Atlanta region to promote a variety of housing options that are connected to existing and future job centers via expanded transit?	Very important = 3 Important = 2 Not that important = 1 Unimportant = 0
Consider where you have lived, worked and attended school. Have you ever made a choice regarding employment, education or housing based on access to transit?	Yes = 1 Unsure = 0 No = −1
How important is it to you to have a public transit option available where you live in the Atlanta region right now?	Very important = 3 Important = 2 Not that important = 1 Unimportant = 0
If you think you would move, would it be closer or further from work?	Closer to work = −1 Further from work = 1 Same distance to work = 0 Unsure = 0 Not likely to move = 0
Socioeconomic variables for class membership modeling	
Do you own a smartphone or plan on purchasing one in the near future?	Yes = 1 Unsure = 0 No = −1
Age	0–18: 0 19–24: 1 25–34: 2 35–44: 3 45–54: 4 55–64: 5

(continued on next page)

Table A.2 (*continued*)

Questions as indicators	Options and value assignment
	65–74: 6
	75+: 7
Gender	Female: 0
	Male: 1
Race/Ethnicity (modeled as a categorical variable)	White/Caucasian
	Black/African-American
	Asian
	Hispanic/Latino
	Two or more races
	Some other race

Table A.3
Summary of the latent-class analysis on autonomous vehicles.

	Class 1	Class 2	Class 3	*p*-Value
Models for responses to questions (β^t_{mx0} in Eq. (4))				
Do you believe autonomous vehicles will become a realistic option within the next couple of decades?	5.5914	−0.3906	−1.6207	0.0000
If autonomous vehicles led to less traffic congestion and also allowed you to be productive while riding instead of driving, how likely would you be to move to another part of the region?	0.3826	0.2980	−0.0614	0.0000
How comfortable would you be in transferring driving control to a fully autonomous technologically advanced vehicle?	2.7615	1.0491	−0.4015	0.0000
Do you see driverless cars as a viable option for people who cannot drive themselves such as older adults or people with disabilities?	5.1021	−0.3106	−1.4744	0.0000
Technology to implement autonomous vehicles will be available in the next 5–10 years; however liability issues may delay their introduction. Should the State of Georgia support the implementation of autonomous vehicles, including enactment of laws and regulations encouraging this technology?	6.4939	0.7568	−1.7301	0.0000
Model for clusters				
Intercept	−2.1715	1.5005	1.3364	0.0000
Covariates				
Have you heard of autonomous vehicles (“driverless cars”)?	2.0044	−0.4644	−0.7345	0.0003
Age	−0.1486	−0.0087	0.0066	0.0000
Race/Ethnicity				
Asian	0.4621	0.4187	−0.2586	0.0014
Black/African-American	−0.1867	0.1672	0.1508	
Hispanic/Latino	0.1499	0.1041	−0.4574	
Some other race	−0.8509	−0.2555	0.2536	
Two or more races	0.4915	−0.3197	0.2534	
White/Caucasian	−0.0659	−0.1148	0.0583	
Gender	0.5121	−0.2610	−0.3882	0.0000
Do you own a smartphone or plan on purchasing one in the near future?	0.9826	−0.1965	−0.2697	0.0000

Table A.4
Summary of the latent-class analysis on transit-oriented development.

	Class 1	Class 2	Class 3	Class 4	*p*-Value
Models for responses to questions (β^t_{mx0} in Eq. (4))					
Should the Atlanta region strive to achieve a more equitable distribution of economic opportunities by placing greater emphasis on developing new employment centers in the southern, eastern and western parts of the region?	1.9612	−0.3866	0.5768	−1.0161	0.000
Do you believe connections with a regional rail and/or express bus network will be essential for existing job centers and potential new ones to grow and be successful in the future?	4.7844	1.2626	−0.8114	−1.6202	0.000
How important is it for the Atlanta region to promote a variety of housing options that are connected to existing and future job centers via expanded transit?	1.7834	1.6239	0.0119	−1.0136	0.000
Consider where you have lived, worked and attended school. Have you ever made a choice regarding employment, education or housing based on access to transit?	0.6298	1.1869	−0.4281	−0.1912	0.000
How important is it to you to have a public transit option available where you live in the Atlanta region right now?	2.0740	1.9983	−0.5780	−0.3906	0.000
If you think you would move, would it be closer or further from work?	−0.4673	−0.2574	0.0444	0.2812	0.000
Model for classes					
Intercept	0.3765	1.4930	0.0660	−0.5054	0.000
Covariates					
Age	0.0001	−0.0003	0.0001	−0.0001	0.000
Race/Ethnicity					
Asian	0.5031	0.4550	0.6073	0.1485	0.000
Black/African American	1.4728	−0.1571	0.8180	−1.9511	
Hispanic/Latino	0.0749	0.4882	−0.8764	0.5391	
Some other race	−0.6197	−1.0583	0.0685	0.3434	
Two or more races	−0.2501	−0.0203	−0.2558	0.0471	
White/Caucasian	−1.1810	0.2925	−0.3616	0.8731	
Gender	−0.7598	0.2024	−0.6095	0.6405	0.000
Do you own a smartphone or plan on purchasing one in the near future?	−0.1201	0.3266	0.0972	−0.0674	0.007

References

ARC (2015). The region's plan phase II survey report. http://atlantaregionsplan.com/community-engagement/

Azevedo, K., Bras, B., Doshi, S., & Guldberg, T. (2009). *Modeling sustainability of complex systems: A multi-scale framework using SysML* (pp. 1437–1448)ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers.

Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. *Transportation Research Part C: Emerging Technologies*, *67*, 1–14.

Burns, L. D. (2013). A vision of our transport future. *Nature*, *497*, 181–182.

Campbell, M., Egerstedt, M., How, J. P., & Murray, R. M. (2010). Autonomous driving in urban environments: Approaches, lessons and challenges. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*, 4649–4672.

Economist (2015). *If autonomous vehicles rule the world: from horseless to driverless.* (Retrieved June 10, 2016, from http://worldif.economist.com/article/12123/horseless-driverless. In).

Fagnant, D. J., & Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C-Emerging Technologies*, *40*, 1–13.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181.

Fong, A., Hettinger, A. Z., & Ratwani, R. M. (2015). Exploring methods for identifying related patient safety events using structured and unstructured data. *Journal of Biomedical Informatics*, *58*, 89–95.

Greenblatt, J. B., & Saxena, S. (2015). Autonomous taxis could greatly reduce greenhouse-gas emissions of US light-duty vehicles. *Nature Climate Change*, *5*, 860–863.

Greenblatt, J. B., & Shaheen, S. (2015). Automated vehicles, on-demand mobility, and environmental impacts. *Current sustainable/renewable energy reports. 2*. (pp. 74–81).

Kuang, D., & Park, H. (2013). *Fast rank-2 nonnegative matrix factorization for hierarchical document clustering* (pp. 739–747)Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.

Litman, T. (2015). *Autonomous vehicle implementation predictions.* Victoria Transport Policy Institute.

Lu, Z., Southworth, F., Crittenden, J., & Dunham-Jones, E. (2015). Market potential for smart growth neighbourhoods in the USA: A latent class analysis on heterogeneous preference and choice. *Urban Studies*, *52*, 3001–3017.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval. Vol. 1*.Cambridge: Cambridge university press.

Martin, E., & Shaheen, S. (2011). The impact of Carsharing on public transit and non-motorized travel: An exploration of North American Carsharing survey data. *Energies*, *4*, 2094.

Pandit, A., Lu, Z., & Crittenden, J. (2015). Managing the complexity of urban systems. *Journal of Industrial Ecology*, *19*, 201–204.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, *89*, 14–46.

Vermunt, J. K., & Magdison, J. (2005). *Technical guide for latent gold 4.0: Basic and advanced.* Belmont Massachusetts: Statistical Innovations Inc.

Wadud, Z., MacKenzie, D., & Leiby, P. (2015). Help or hindrance? The travel, energy and carbon impact of highly automated vehicles. *Transportation Research Part A: Policy and Practice*.

Xu, W., Liu, X., & Gong, Y. (2003). Document clustering based on non-negative matrix factorization. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 267–273). ACM.

Zhang, W. W., Guhathakurta, S., Fang, J. Q., & Zhang, G. (2015). Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. *Sustainable Cities and Society*, *19*, 34–45.