Sustainable Cities and Society 38 (2018) 123–136

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Supporting sustainable system adoption: Socio-semantic analysis of transit rider debates on social media

Moein Hosseini, Tamer El-Diraby*, Amer Shalaby

Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario M5S 1A4, Canada

ARTICLE INFO

Keywords:
Customer satisfaction
Energy saving
Ridership behavior
Social media
Social network analysis
Semantic analysis

ABSTRACT

Online social media platforms provide a bi-directional communication channel between transit agencies and their customers. It can also be an effective venue for profiling users and their needs as a step towards customizing service and communication policies. This study analyzed online Twitter discussions for three major transit agencies in Canada. In our work, we integrate the analysis of the participants' social networks with the contents of their discussion. We also conduct the semantic analysis in a manner that parallels the structure and contents of customer satisfaction surveys—allowing for insightful comparisons between the results of both methods Analysis of the structure of the social networks of the Twitter accounts under study, including investigation of the formation of sub-communities and their interrelationship to the overall network. It was found that networks of the three cases portray a scale free and small world behaviors. This means that they are maturing networks; and that they represent viable communities—not just a randomly connected graph. This is important for future studies in relation to information diffusion and opinion dynamics: how people share information and how does this help shape their views. On the semantic front, a lexicon was developed based on existing thesauri for customer satisfaction analysis. Keywords form each tweet were extracted and the topic(s) of each tweet was defined based on the lexicon. It was found that, based on the sample investigated, the behavior of 100-follower networks (networks with nodes having at least 100 followers) closely mimics the behavior of the overall network. Studying these networks (of influential users) can make analysis faster and may not impact accuracy. We also clustered each network into sub-networks: small, medium, and large. Topics discussed in medium-size networks tended to be unique. This seems to be the level where active discussions of specific topics take place. Focusing on detecting these and analyzing their contents can provide better chance for capturing the evolution of community opinions.

1. Introduction

At the core of any successful sustainable policy for energy in cities is a widely-used transit system. Transit systems save energy and reduce emissions. More importantly, they are becoming one of the best venues for using clean energy due to positive economics and the increasing ease of implementation—thanks to advancement in green infrastructure technologies such as charging stations. Beyond the intrinsic savings, transit ridership influence transportation development and travel patterns positively—supporting walking, biking, shorter trips and trip reduction. In fact, it is estimated that total GHG (greenhouse gas) savings from reasonable increase in transit usage in the USA could reach at least 37 million metric tons of CO_2 (McGraw, Shull, & Miknaitis, 2010).

Profiling customers and understanding their needs are essential to effective planning and operations, hence achieving energy conservation targets. To illustrate, in the USA, since 1973, overall transit ridership has grown 55%. Of which, bus ridership has grown by only 6 percent while heavy rail and light rail ridership have each more than doubled (APTA, 2000). The disparity of growth in the two modes reflects the dominance of the socio-economic factors in transit systems (urbanization increased investments in regional rails and encouraged longer trips). Structural changes in society with concurrent preference changes have significant impacts in volume and composition of transit use. In addition to costs, non-price factors such as values and social norms have major role in supporting behavior change. Anable, Brand, Tran, and Eyre (2012) showed that lifestyle changes have significant impact to the extent that energy reduction targets can be achieved even if there are significant delays in electrification. If both are achieved, a more reliable future for energy savings can be realized. However, despite the long-term recognition that behavior change is essential and valuable, technical (more efficient vehicles) and economic (carbon pricing) considerations dominated policy making.

Surveys have been used for decades by transit agencies to measure the levels of customer satisfaction. They contain questions about socio-

* Corresponding author.
E-mail addresses: moein.hosseini@mail.utoronto.ca (M. Hosseini), tamer@ecf.utoronto.ca (T. El-Diraby), amer@ecf.utoronto.ca (A. Shalaby).

https://doi.org/10.1016/j.scs.2017.12.025
Received 7 August 2017; Received in revised form 18 December 2017; Accepted 19 December 2017
Available online 26 December 2017

economic characteristics of the users, trip features, the importance of different service attributes, and levels of satisfaction with these attributes. Surveys typically ask customers about their overall (average) satisfaction over a period of 30 days. The 30-day window may cause some biases in assessment: if the customer went through an incident of delay or bad service in recent days, this may impact his/her assessment negatively. Otherwise, if such an incident occurred before the 30-day window, the agency will lose the opportunity to learn about the impacts of such incident on satisfaction levels.

Analyzing social media feeds (on Twitter or Facebook, for example) can be an effective supplementary tool to the surveys. The majority of transit agencies already have social media outlets and customers interact with these continually. In fact, many agencies have used semantic and content analysis of social media feeds to study customer views. However, most of these studies did not consider the social network aspect: how the structure of the social network and interrelationships between user groups impact opinion making. If customer satisfaction analysis is to be used to help support understanding and promoting behavior change, instead of using aggregate satisfaction indices, we need to profile the needs of different groups to be able to develop matching reach-out, operational policies, and relevant service options (Anable et al., 2012). The consideration of the social (group) dynamics is becoming important to understanding behavior change given the increasing view that change is not just a reflection of automatically cued, repetitive behavior of individuals. Rather, a body-mind-world continuum, where actions by others and government policies matter (Schwanen, Banister, & Anable, 2012).

1.1. Objectives and approach

Our objective is to progress the discipline of profiling transit users and encouraging behavior change through conducting an analysis that has three main features:

1. Linking the analysis of user social networks with the contents discussed. Do all groups (sub-networks) discuss the same topics? How do topics differ across groups?
2. Analyzing the difference in social and semantic features during normal days and days with service interruptions. This can help understand how people react to interruptions; what information or help do they need during such incidents? This can support better handling of emergency situations.
3. Conduct the above analyses across different networks to examine the possible role of local contexts on the analysis. To this end, we conducted the analysis for three Canadian transit agencies: Toronto Transit Commission (TTC)—a bus and subway system within the city; GO transit—a regional rail and bus service in the greater Toronto Area; and TransLink—a bus and light rail system in the city of Vancouver.

Our analysis approach can be summarized as follows:

1. Collect and clean the data corpus; detect days with increased level of discussions—especially those that occurred during an interruption incident.
2. Use clustering algorithms to detect sub-communities in the user network: groups that are more connected to each other than to the overall network.
3. Develop a lexicon of topics and keywords to look for within tweets.
4. Detect keywords in tweets and cluster them into the topics identified above.
5. Study topics discussed within sub-groups and in different operational situations (regular days and days with incidents).

The paper starts with a brief review of recent studies on measuring customer satisfaction in transit systems, including traditional approaches and through social media analysis. This review emphasize that work on the use of social media in analyzing customer satisfaction typically used limited sets of data and, in almost all cases, did not consider the nexus between social (user groups) and semantic (topics of discussion) networks. Next, the methodology of this study is explained—including steps used for data collection and data pre-processing, development and use of customer satisfaction lexicon, and social and semantic network analyses. This is followed by a discussion of the major findings of the analysis. First, we present the results of our social network analysis in the three agencies. Second, we present the results of the semantic analysis of tweets. Thirdly, we discuss the relationships between these two analyses. At the end, a set of conclusions and suggestions for future work are discussed.

2. Background

Achieving energy and carbon management targets in transport systems will require significant changes both on the technology and policy dimensions. One of the essential means to achieve this is to promote behavior changes that are supportive to the use of public transit (Waisman, Guivarch, & Lecocq, 2013). Comparisons of the contributions of technical options (such as a shift to biomass or hydrogen for the non-electricity energy requirements) and behavioral options (such as a reduction in total passenger-km of travel per capita) has been assessed to favor technical options in the short to medium terms in many studies (see for example, Harvey, 2013; McGraw et al., 2010). However, the perspective of modeling approach matters in this regard. Creutzig (2016) compared three analysis approaches: integrated assessment modeling, transport-sector modeling, and place-based modeling. The first focuses on fuel composition, the second on efficiency measures and the third on localized and idiosyncratic behavioral mitigation options. He reports that “specific local approaches could mitigate urban transport emissions by 20–50%, higher than that revealed in aggregate global models.” This means that the role and value of behavioral change is important to long-terms conservation and could be the most important factor in localized changes. Having a longer inertia, it is important to note that making progress in this regard is not an easy task; and if we are to reach conservation goals by mid-century, we must start now (Pietzcker et al., 2014). In fact, in addition to prices and income, cultural and life-style patterns or constraints are now essential to any work into urban energy modeling (Ajanovic, Dahl, & Schipper, 2012).

2.1. Customer satisfaction surveys

Profiling rider views is important to understanding drivers for behavior change. Typically, capturing rider preferences and levels of satisfaction is done through surveys. The first step in conducting the traditional customer satisfaction surveys is to define the indicators to measure. Indicators are chosen by conducting a preliminary survey or through focus groups. The selection of indicators depends on various factors including, but not limited to, the transit modes under consideration, nature of the service provided, and local culture. Therefore, there is no comprehensive standard list of service quality indicators. Table 1 shows a summary review of relevant studies and the indicators used, sample sizes, and analysis approach.

The indicators are then assessed through input from customers with the aim of investigating the interaction between the scores of service indicators and the overall quality perceived by the users (de Oña et al., 2013; Mouwen, 2015). Simple indices such as ServQual (initially introduced by Parasuraman, Zeithaml, & Berry, 1985) is widely used in the sector (see for example, Chou, Liu, Huang, Yih, & Han, 2011; Pakdil & Aydin, 2007). Five service quality dimensions and 22 indicators are used. Service quality spans the following dimensions: tangibles, reliability, responsiveness, assurance, and empathy. A questionnaire is used to collect customer assessments on a Likert scale with seven levels (from