CSCE C G C 1887
RESILIENT INFRASTRUCTURE
June 1–4, 2016
RESILIENT INFRASTRUCTURE LONDON 2016

A STANDARD LEXICON TO MEASURE THE LEVEL OF SERVICE OF PUBLIC TRANSPORTATION SERVICES THROUGH ONLINE TRANSIT-ORIENTED DISCUSSIONS

Moein Hosseini
MASc Candidate, University of Toronto, Ontario, Canada

Tamer El-Diraby
Associate Professor, University of Toronto, Ontario, Canada

Amer Shalaby
Professor, University of Toronto, Ontario, Canada

ABSTRACT

Public involvement in operation and planning of transit services is becoming a major focus of public transportation agencies. With the growing population and diversification of ideas, distilling knowledge from public opinion to incorporate into decision support systems is a major challenge for transit agencies. The common practice is to collect information from customers via Customer Satisfaction Surveys (CSat). Although such surveys provide useful insight for transit agencies, their design, data collection, and interpretation is expensive and time consuming. Because of the qualitative nature of the questions in these surveys, inconsistency in respondents' perception of the survey questions is another challenge. Furthermore, for management and decision-making purposes, the results of surveys should be quantified into various key performance indicators. Lack of a standard quantification system causes some biases in the reported results. The advent of online social media has introduced a new bidirectional communication system. Developing a linguistic-based system to interpret customer discussions in social media about the transit system and to transform the discussions into useful information for the agency could overcome some deficiencies existing in CSat. The current study develops a standard lexical resource to categorize online discussions on Twitter into different Level of Service (LOS) indicators.

Keywords: Level of Service, Customer Satisfaction, Lexicon, Social Media, Performance Measurement System

1. INTRODUCTION

Similar to many other organizations, transit agencies assess their activities using a large number of metrics from various perspectives. Utilizing a well-defined system of measures is fundamental to conducting the evaluation. This paper elaborates on the concepts related to performance measurement systems (PMS). More details are provided on the level of service (LOS) and customer satisfaction (CSat) measures which are a subset of the performance measures. The definition of performance measures and the LOS are outlined in various sources. Different performance perspectives of the transit service are discussed. Social media as an alternative source to collect information about the service quality is introduced. The process of developing a lexicon using standard linguistic resources is discussed. The lexicon facilitates the connection between information retrieved from social media and the proposed framework for the LOS measurement system. The performance of the proposed system is evaluated by analyzing the semantic of the discussions on Twitter about three transit agencies operating in Canada.

2. BACKGROUND

2.1 PERFORMANCE MEASUREMENT SYSTEM

Companies and agencies evaluate their actions in providing goods and/or services to their customers. The evaluation is not only important to adjust their internal processes but also to provide customers with relevant reports. The assessment is performed typically using a performance measurement system (PMS) which consist of both qualitative and quantitative measures (Bititci et al. 1997, Forza and Salvador 2000, Lynch and Cross 1991). The main role specified for a PMS is to link strategic plans, decisions, management processes, and operations of an organization to the objectives of the organization including its primary objectives set by the owners, and secondary objectives defined as a result of transactions and contractual relationships with critical external stakeholders (Atkinson 1998, Atkinson et al. 1997, Gates, 1999, Ittner et al. 2003, Maisel 2001, Neely et al. 1995). Using a PMS, deviations from the objectives could be detected and managers are able to suggest corrective actions. As an information system, appropriate technology infrastructure is necessary to collect raw data, process the information, and deliver results of the metrics throughout the organization. Information flow within the PMS should be bidirectional, considering feedback and feed-forward flows, resulting in the emergence of new metrics (McGee 1992, Otley 1999).

Transit agencies use performance measures because of three important reasons: 1) Transit agencies are required to provide various reports periodically to the government; 2) The performance measures inform other stakeholders outside the transit agency (transit boards, funding bodies, and public) about the operational conditions of the service; and 3) They are useful for the transit agency to provide an understanding of how well the transit agency is performing in its service area. Consequently, the measurement system could help the agency institute policy changes to improve its service. According to a study by the Kittelson & Associates, Inc. et al. (2003), some of the key purposes of the PMS include:

- Monitoring the service,
- Evaluating the economic performance of the system,
- Improving the management functionality,
- Enhancing the communications within the agency,
- Developing service design standards,
- Identifying the achievements and challenges, and
- Determining the community benefits.

Transit agencies perform a tradeoff between their cost and their revenue to determine the extent, details, and quality of the service they provide. They usually use performance measures to compare the cost and revenue. A challenge to the transit industry is to use the most pertinent and useful performance measures for the comparative study (cost and revenue). If the study is performed in the planning or design stage, various default values and standard parameters are utilized to calculate the measures. However, if the study is performed at the operational planning stage, field measurements are the best source of information. Besides this challenge, a better analysis compares the cost against the benefits of the system. The measures used to perform this analysis allow the transit agency to consider customer and community concerns along with cost-efficiency and cost-effectiveness of the system (Kittelson & Associates, Inc. et al. 2003, Bititci et al. 1997, Kerssens-van Drongelen and Fisscher 2003, Neely 1998, Neely et al. 1995).

2.2 PERSPECTIVES OF A PERFORMANCE MEASUREMENT SYSTEM

The measures proposed in different PMS guidelines represent various perspectives. For example, Kaplan and Norton suggest a business PMS with four measurement perspectives: 1) financial 2) customer 3) internal business process 4) learning and growth (Kaplan and Norton, 1996). In the field of public transportation, there are generally four types of performance perspective: (Kittelson & Associates, Inc. et al. 2003)

- **Customers**

Transit users are classified into two major groups: 1) choice riders 2) transit-dependent (captive) riders. Captive riders could be considered the most important group of transit users since their travel highly depends on public transit. Therefore, a minimum standard should be applied to the quality of service (QOS) to keep transit viable for them. On the other hand, choice riders have some alternatives in their choice set including public transit. The QOS is an important aspect that influences their utility of choosing transit for making their desired trip. Usually, the expectation of the choice riders is higher than that of transit-dependent riders.

Customers of public transit are not restricted to choice riders and captive riders. In the current study, a wider spectrum of stakeholders is considered as the customers, namely the actual riders, neighbours (businesses and potential riders), and taxpayers (Figure 1). The dashed line specifies the area covered by the information derived from social media to calculate the performance measures. Although not all of these stakeholders are direct users of the transit service, their life is influenced by the actions and decisions of the agency providing the service. Therefore, they want to have a voice in the decisions made for the planning and operation of the service.

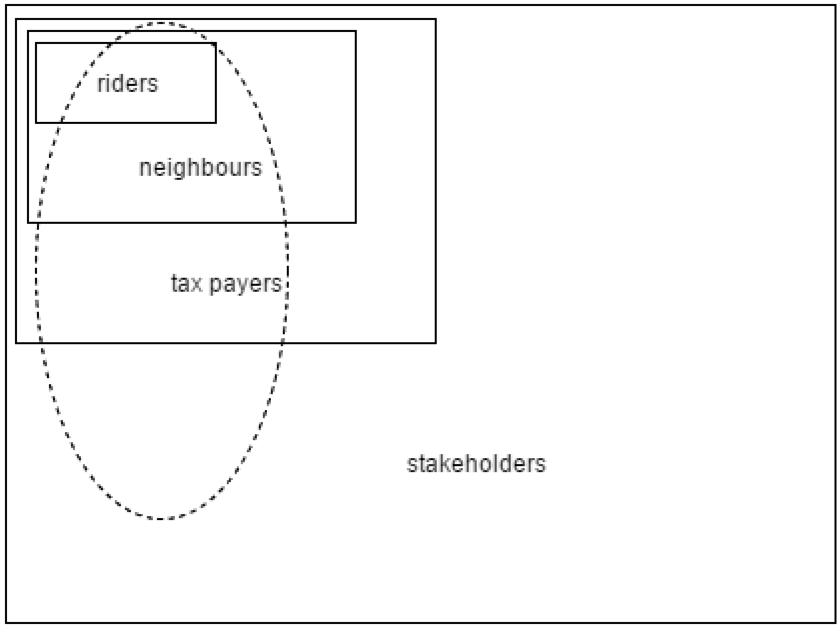

Figure 1: Customers considered in the analysis

The customer point of view reflects the level of satisfaction of the service. This perspective will be further discussed in the following section.

- **Community**

Transit services is a suitable option for people with no access to private automobiles. Therefore, it also provides job accessibility for the economically disadvantaged residents. It could also reduce air pollution, traffic congestion, and parking congestion. Besides all these benefits, the community has some concerns about public transit. The amount of direct and indirect tax that people should pay for the service, the visual impact of transit facilities on the community, noise levels and diesel fumes from the operating buses, the inefficiency of buses, and empty buses are some of the concerns of the community associated with public transit. A wide range of measures should be developed to reveal the community perception of these benefits and concerns.

- **Agency**

The agency tries to be efficient and effective at the organizational level (evaluated by traditional measures) while addressing customer and community issues (evaluated by non-traditional measures).

- **Vehicle/Driver**

This perspective is mostly pertinent to buses and bus drivers. The measures are representative of the interaction between buses and other road users (automobiles, pedestrians, and cyclists).

3. METHODOLOGY

3.1 Level of Service (LOS)

The Level of Service (LOS), as a constituent of the PMS, is a common term used in the field of transportation. It is sometimes called Quality of Service (QOS). The American Association of State Highway and Transportation Officials (AASHTO) defines LOS as a qualitative measure which describes the operational conditions of the service provided according to some service measures such as travel time, speed, freedom to maneuver, traffic interruptions, comfort and convenience (American Association of State Highway and Transportation Officials, 2009). The Highway

Capacity Manual (HCM) differentiates between the LOS and QOS. The LOS is a quantitatively stratified performance measure which is representative of the service quality provided under the operating conditions from the traveller's perspective. On the other hand, QOS is defined as a description of the quality of a transportation facility or a service operation from the traveller's perspective (National Research Council (U.S.) Transportation Research Board 2010). Therefore, the HCM considers the LOS as a quantitative measure and the QOS as a qualitative measure. Both definitions reach a consensus that the LOS is a subsystem of a PMS which could be evaluated independently from other measurement subsystems. The main difference is that AASHTO measures the LOS from different perspectives while the HCM only focuses on the traveller's perspective. Travellers are only a subgroup of the service customers. The definition provided by the HCM is closer to customer satisfaction (CSat) than the LOS which is evaluated based on different perspectives. Figure 2 explicitly shows the relationship between the PMS, LOS and CSat. The measures used to evaluate the CSat are a subset of the measures defined to assess the LOS. This is mainly because the CSat deals with only the customers' perspective while the LOS incorporates the four perspectives defined in the previous section.

Figure 2: relationship between PMS, LOS, and CSat

For each organization, based on the context and purpose of the organization, a multi-dimensional set of measures are introduced for the PMS (Bourne et al. 2003). Gates suggests a system composed of financial, strategic, and operating measures to assess the performance of a company (Gates, 1999). Neely et al. categorize measures of a PMS in four main categories including quality, time, flexibility, and cost (Neely et al. 2005). For the LOS, the measures should be consistent with the overall PMS. The LOS should encompass various attributes which influence the quality of the offered transportation service. In order to define a standard framework for the LOS measurement system, the 88th and 165th reports of the Transit Cooperative Research Program (TCRP) were utilized (Kittelson & Associates, Inc. et al. 2003, Kittelson & Associates, Inc. et al. 2013). A well-defined structure helps identify the pertinent information for calculating each measure. Figure 3 represents the framework of the LOS measurement system developed based on the TCRP reports. As could be seen, the LOS measurement system is divided into two groups at the first level: 1) Availability, 2) Comfort and Convenience. Each of these groups is further subdivided into four groups.

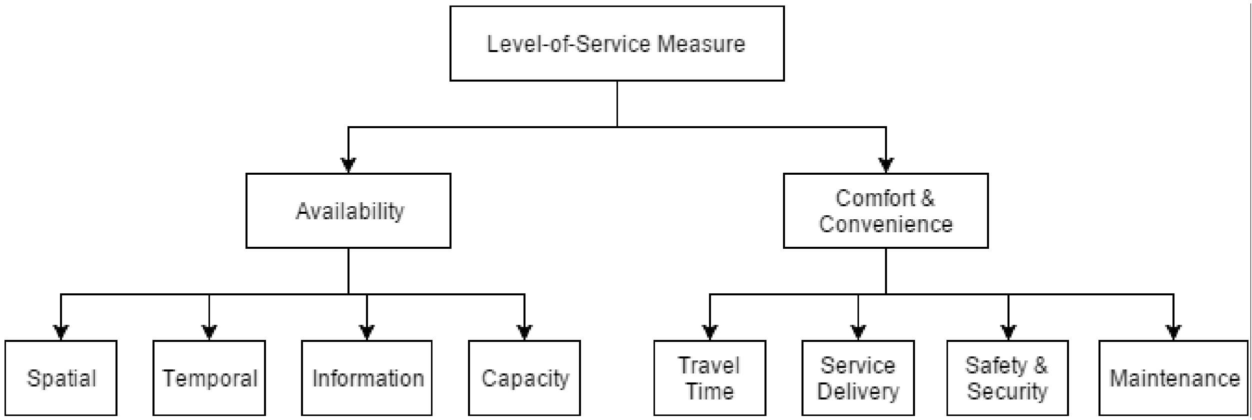

Figure 3: The LOS composition in public transit sector

Each user subjectively evaluates the components of the above framework and develops an overall value for the LOS. Because of the subjective properties of the values assigned to the LOS, there are inherent differences between the users. These differences influence the inclusion of the transit mode in the choice set of a customer. The availability criterion applies to both choice riders and captive riders. The main difference between choice riders and captive riders is in the comfort and convenience measures. Since the choice riders have alternative modes of transportation in their choice set, they compare the utility of the modes against each other. When considering public transit, choice riders look for some features including, but not limited to, saving money; traveling faster and in a reliable manner; avoiding driving in a congested roadway network; efficiently using travel time; and mitigating the negative impact of their trip on the environment. On the contrary, captive riders can only choose transit otherwise they should cancel their trip if the QOS is sufficiently negative. As a result, they usually have a lower expectation of the service quality compared to the choice riders. The transit agency should provide a service which has a quality above a minimum level for these users. A service with a sufficiently above the minimum level of quality is capable of retaining the riders when they are no longer transit dependent (Kittelson & Associates, Inc. et al. 2013).

There are numerous measures, defined by transit authorities and researchers, to measure the LOS and the level of satisfaction from the user point of view which fit within this framework. The purpose of the current study is to define equivalent measures which could be evaluted by the information retrieved from social media in order to assess the LOS.

3.2 SOCIAL MEDIA: A COMPLEMENTARY SOURCE TO MEASURE THE LOS

Social media, also known as Web 2.0, are a set of web-based applications which facilitate a bi-directional communication between the internet users. There are different types of social media including networking sites (such as Facebook and LinkedIn), micro-blogging sites (such as Twitter), media-sharing sites (such as YouTube and Flickr), and location-based sites (such as Foursquare). There are several reasons that transit agencies started using social media including 1) providing timely updates for customers; 2) disseminating information about fares, services, and long-term planning projects; 3) engaging the community in the decision-making process; 4) announcing positions in the organization to recruit new employees; and 5) entertaining riders. Besides all these opportunities for transit agencies, there are some challenges that they have to overcome. First of all, they need organizational support and the employer approval to introduce their social media accounts because it allows customers to bypass the agency bureaucracy. In addition, Engaging in social media activities is resource-hungry. Transit agencies require staff dedicated to continuously monitor the company's social media accounts and respond to criticisms from riders and employees. Furthermore, they need to archive and keep records of their online communications. Last but not least, security of their internal private information against cyber threats is another main challenge. Although there are many concerns and barriers associated with the use of social media, many transit agencies have decided to overcome them and benefit from this new communication system (Bregman, 2012).

The social medium considered in this study is Twitter. A key feature distinguishing Twitter from all other social media is the asymmetric properties of the relationship between the users (Gupta et al. 2013). This is especially important for popular users like famous people or official accounts of companies and agencies. It would have been a nuisance for popular users if it was necessary to respond to each "following" request. The establishment of a link between users without requiring permission makes Twitter a more realistic social network. An exception to this feature is access to private accounts which necessitates authorization of the account holder to access the network and information on their account. These accounts usually belong to individuals with basic status. Another feature specific to Twitter is the limitation on the length of messages. Each public message in Twitter is limited to 140 characters. As of August 2015, the limit on private messages has increased from 140 to 10,000 characters. The limitation (for public messages) makes users generate more unambiguous messages with a specific content and sometimes targeted users. Short messages also facilitate Web-surfing. Many messages in Twitter act as topics that might be of interest to a user. If users require further information, they could follow a link provided in the tweet. Besides the benefits of character restriction, there are also some disadvantages. The most important drawback is the formation of grammatically unstructured message with the extensive use of abbreviations and acronyms. This is one of the main challenges to researchers. In particular, some algorithms which employ Part-of-Speech as a language-specific concept, lose their reliability when studying the user-generated content on Twitter. Hence, the advantages of using Twitter outweighs its disadvantages in the current study.

3.3 TRANSPORTATION THESAURI

Information obtained from Twitter is in textual format with a limit of 140 characters for each sample (Tweet). Although the character limitation may introduce an unstructured use of language, standard keywords related to the context are still available. Since we are dealing with a set of vocabularies, the best way to exploit useful information is to use a linguistic-based tool. An efficient linguistic-based tool requires a standard linguistic database as well. A thesaurus accommodates all the necessary features to be an eligible database for this purpose. A thesaurus is a reference containing words and phrases and the relationship between each other. A thesaurus should not be misunderstood as a dictionary which lists the words alphabetically and provides a definition of the words. The relationship represented in a thesaurus can vary from hierarchical relationships such as *Broader Term* and *Narrower Term* to associative relationships such as *Related Term*, *Synonym*, and *Antonym.* There are many thesauri developed in different fields. The two thesauri utilized in the current study which are specific to the field of transportation, are described below.

The first thesaurus utilized is the *Transportation Research Thesaurus (TRT)* published by the National Research Council (U.S.) Transportation Research Board in 2001. It summarizes efforts of a team of lexicographers and database experts. It is intended to address some issues such as inconsistency in terminology when describing the same concept and extensive misuse of a proper noun to describe a subject. The thesaurus was developed based on a rich dataset from various sources including the *Transportation Research Information Services (TRIS)*, the *Highway Research Information Services (HRIS)*, *Northwestern University*, and the *University of California, Berkeley.* Although it covers all areas of transportation, it is focused more on ground transportation. The main scope of this thesaurus is to unify the language and terminology between providers and users of the information available in the TRIS database. The main concept behind the TRT is facet analysis (National Research Council, Transportation Research Board, 2001). A facet is a principal characteristic which establishes a classification system by discriminating between different concepts (keywords and key-phrases) in a dataset. Facets should be homogeneous, mutually exclusive, and collectively exhaustive. They should also conform to the scope and subject of the classification system (Spiteri, 1998). The TRT was developed in an iterative process. Each iteration constitutes two stages. In the first stage, all the vocabularies in the dataset were clustered in multiple semantic groups in order to define the facets and possible sub-facets considering the independency criterion for the facets. The second step involved the assignment of terms available in the database to the facets and sub-facets defined in the first stage. After this stage, some terms did not belong to any cluster or they belonged to multiple clusters. This two-stage process was iterated until the structure of the facets were revised in such a way that all terms were included in the thesaurus. Each facet in the TRT was assigned a letter of the alphabet. This provides the basis to define a unique code for each term in the Thesaurus which helps to identify the location of the term in the hierarchy (National Research Council Transportation Research Board, 2001).
The second thesaurus used in the current study is the *Australian Transport Index Thesaurus (ATRI)* published by the ARRB Group Ltd. The focus of this thesaurus is on road-based transportation. It uses the following codes to show the relationship between the terms in the database: (Cox et al. 2013)

- SN: Scope Notes

  This code is used to show the contextual usage of a keyword or to provide further clarification for the meaning of a keyword.
- UF: Used For

  This code represents other commonly used synonyms for a term.

- BT: Broader Term
  This code defines the hierarchical structure of the thesaurus.
- NT: Narrower Term
  This code defines the hierarchical structure of the thesaurus.
- RT: Related Term
  This code is representative of an associative relationship between different keywords.

Although the logic behind the two thesauri seems quite similar, and there are some common terms between them, there is a main structural difference between them. The TRT has an explicitly defined hierarchical design and each term has a unique code. Therefore, each term could be easily located within the thesaurus independently from other keywords. On the other hand, the hierarchy defined in ATRI Thesaurus does not accommodate all terms in the database and some of the terms introduced in the thesaurus are loosely connected with the associative relationship (RT). Although the TRT dominates the ATRI Thesaurus in terms of the structural definition, the abundance of keywords in the ATRI Thesaurus makes it a useful source to build the intended lexicon. Therefore, the lexicon developed to define the LOS categories takes advantage of both thesauri.

3.4 LEXICAL RESOURCE

The two transportation thesauri introduced are standard collections of words and phrases which cover almost all aspects of transportation. The focus of the current study is on a specific concept in transportation, the LOS. Therefore, only a portion of each thesaurus was used. First, similar to the TRT, the categories of the lexicon were determined. In developing a thesaurus, facets may be subject to changes over each iteration of word assignment. However, for the purpose of the current study, the categories are immutable. The categories defined in the lexicon are not facets because they violate some of the fundamental features of facets. For example, they are not mutually exclusive and some of the keywords fall under multiple categories. The next step involved searching throughout the thesauri for keyphrases related to the study such as "*Transport Performance",* "*Level of Service"* and *"Quality of Service"* in addition to keywords and key-phrases originated from the eight categories of the LOS. Then, all the keywords found related to the LOS categories were assigned to the relevant class(es). The final collection benefits from both the unique structure of the TRT and the abundance of keywords in the ATRI Thesaurus. The lexicon contains 227 keywords and key-phrases in total. It consists of 207 unique terms. There are 20 terms which are shared among two categories. These shared keywords are another proof for the violation of the mutual exclusiveness attribute. As a result, in order to enhance the categories to accommodate the properties of facets, a thorough linguistic study is required.

3.5 SEMANTIC ANALYSIS

Semantic analysis on text is performed by analyzing words and phrases utilized in developing a text and the relationship between these elements. The appearance of keywords and key-phrases are studied via different feature sets in semantic analysis such as Term Occurrence, Term Frequency (TF), and Term Frequency Inverse Document Frequency (TF-IDF). The relationship between words in a sentence is established by "grammar". Therefore, one feature which could be defined for each word in a sentence is their grammatical roles or in Natural Language terminology, their "Part-of-Speech" (POS). Although in some cases POS has been a useful feature set, it is mostly used for sentiment analysis. Because of the short format of tweets (Twitter restricts a tweet to 140 characters) and utilization of a specific lexical resource in the current study, the feature sets chosen were the occurrence of monograms (tokens or keywords) and occurrence of phrases. The occurrence of phrases was implemented in such a way that inclusion of other irrelevant words between keywords of a key-phrase, and the order of keywords of a key-phrase in a tweet do not influence the labelling process. It is worth mentioning that the method employed to label the tweets only specifies the subject (semantic). Two methods were utilized to perform the label assignment: 1) without stemming 2) with stemming. The original lexicon was used in the first analysis without any changes. In the second analysis, the lexicon was replicated with the stem of the words and the labelling task was also performed based on the stem of words. Comparing the labeling performance of the lexicon with word stems and those without word stems could reveal the influence of stemming.

4. RESULTS AND DISCUSSION

The tweets about the Toronto Transit Commission (TTC), Translink, and GO Transit, as the largest public transit agencies in Canada, were analyzed. There are several reasons for choosing these three agencies. For example, due to their high passenger volumes, a large number of tweets could be collected containing concerns of customers about the service. There are some key differences between these agencies such as differences in geographical location and the type of service (local vs. regional) which are useful in validating the proposed lexicon as a standard linguistic resource for the LOS. The data collection was performed over a period of approximately 11 months (Feb 5th 2015 to Dec 31st 2015). The tweets were collected by a streaming process using the name of the agencies as the searching keywords. After preprocessing the tweets and removing all the irrelevant tweets, 496389, 140694, and 159647 tweets respectively relevant to the TTC, Translink, and GO Transit were available. The "without stem" ["with stem"] approach was capable of labelling 15.93%, 9.82%, and 21.32% [20.91%, 17.10%, and 27.57%] of the tweets of each agency. The results show that there were more formal discussions about GO Transit compared to the two other agencies. Tweets pertaining to Translink had the least level of formality. The labelling process works in such a way that a tweet is capable possessing multiple labels. A tweet assigned multiple labels has keywords from different categories or a specific keyword that belongs to multiple categories. Using the stem of words had a positive effect in labelling the tweets. The stemming became more effective as the number of topics discussed in a tweet increased.

The effect of the number of terms in each lexicon on the number of labelled tweets is another source of bias which needs attention. Table 1 contains the number of terms in each category and the number of tweets associated with that category. Pearson Coefficient of Correlation is used to draw conclusion on the possible bias introduced by the number of terms in a category. The coefficients of correlation measured for the correlation between the number of terms on one hand and number of labelled tweets ("without stem" approach); number of labelled tweets ("with stem" approach); and the number of differences on the other hand are respectively 0.6512, 0.6200, 0.0474. Although the result shows a slightly high correlation between the number of terms in the lexical resource and the number of labelled tweets, the diversity of key terms in each lexicon is still the main influencer on the number of labelled tweets. The positive correlation could become closer to zero if the lexical resource is enriched by other resources such as keywords used in academic papers, folksonomy, and news articles. The jump in the number of labelled tweets when switching from the "without stem" approach to "with stem" approach is not correlated to the number of terms in the corresponding resources. Since including the stem of words improves the effectiveness of the system and the enhancement is more related to the nature of terms utilized, further analysis on the data could be conducted based on the stemming approach.

Figure 4 compares the two methods based on the percentage of tweets tagged to each category of the LOS. The largest difference is in the "service delivery" category. Service delivery in public transit comprises a variety of aspects. Therefore, the diversity of keywords in this context is higher than others which in turn influences the number of detected tweets in the "with stem" method. In both methods, a great portion of the discussions is about the "travel time" and "safety and security" of the service. Furthermore, there are more discussions about the convenience and comfort of the system rather than the availability of the service. "Information availability" dominates the other categories of the availability. This shows that a major topic of concern, within the availability spectrum, among the customers is the information that the agency provides. There are not many discussions about the maintenance of the facilities. This is mainly because among all components of the LOS, it is more related to the agency perception of the system rather than the customer perception. In other words, in most cases, maintenance issues require technical knowledge of the system and facilities.

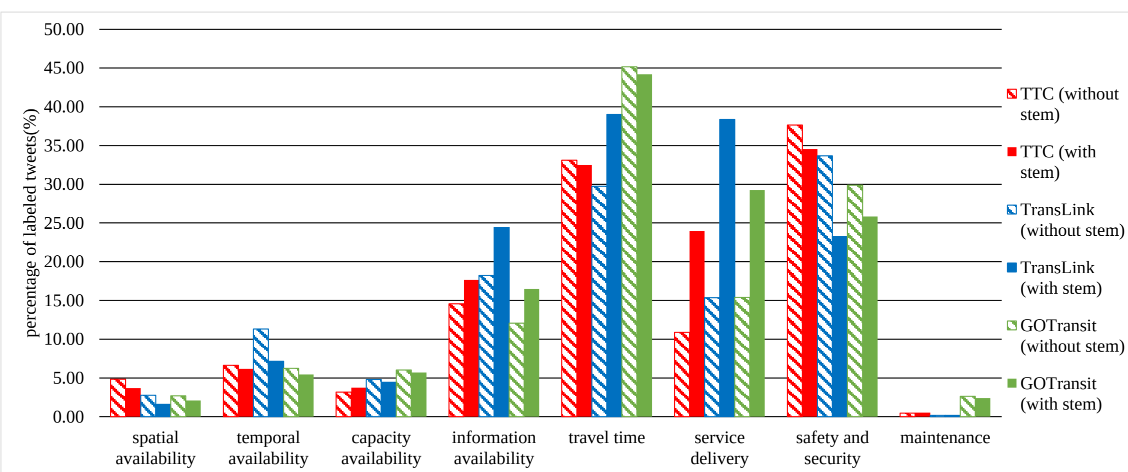

Figure 4: percenatege of tweets tagged for the two approaches of defining the feature set (with[out] stemming)

Table 1: Effect of the number of items in each category of the lexicon on the labelling process

category	number of items in lexicon	transit agency	number (%) of labelled tweets		
			without stem	with stem	difference
spatial availability	3	TTC	3814 (4.83%)	3814 (3.68%)	0 (-1.15%)
temporal availability	12		5245 (6.64%)	6394 (6.16%)	1149 (-0.47%)
capacity availability	3		2512 (3.18%)	3901 (3.76%)	1389 (0.58%)
information availability	14		11512 (14.56%)	18346 (17.68%)	6834 (3.12%)
travel time	25		26163 (33.10%)	33742 (32.52%)	7579 (-0.58%)
service delivery	60		8610 (10.89%)	24848 (23.95%)	16238 (13.05%)
safety and security	101		29756 (37.65%)	35858 (34.56%)	6102 (-3.09%)
maintenance	9		357 (0.45%)	554 (0.53%)	197 (0.08%)
spatial availability	3	Translink	382 (2.77%)	383 (1.59%)	1 (-1.17%)
temporal availability	12		1564 (11.33%)	1722 (7.16%)	158 (-4.17%)
capacity availability	3		661 (4.79%)	1070 (4.45%)	409 (-0.34%)
information availability	14		2517 (18.23%)	5876 (24.42%)	3359 (6.20%)
travel time	25		4108 (29.75%)	9388 (39.02%)	5280 (9.27%)
service delivery	60		2119 (15.35%)	9231 (38.37%)	7112 (23.02%)
safety and security	101		4649 (33.67%)	5601 (23.28%)	952 (-10.39%)
maintenance	9		21 (0.15%)	40 (0.17%)	19 (0.01%)
spatial availability	3	GO Transit	918 (2.70%)	918 (2.09%)	0 (-0.61%)
temporal availability	12		2121 (6.23%)	2400 (5.45%)	279 (-0.78%)
capacity availability	3		2048 (6.02%)	2509 (5.70%)	461 (-0.32%)
information availability	14		4113 (12.08%)	7246 (16.46%)	3133 (4.38%)
travel time	25		15376 (45.17%)	19453 (44.20%)	4077 (-0.97%)
service delivery	60		5243 (15.40%)	12881 (29.27%)	7638 (13.87%)
safety and security	101		10171 (29.88%)	11369 (25.83%)	1198 (-4.04%)
maintenance	9		889 (2.61%)	1049 (2.38%)	160 (-0.23%)

5. CONCLUSION AND FUTURE WORK

In this paper, the framework of a system to calculate the LOS is defined. Appropriate language-based features in the form of keywords and key-phrases were allocated to the blocks of the LOS classification system using two standard thesauri. The resulting system was utilized to classify the information obtained from Twitter. A semantic analysis based on the occurrence of the keywords and the key-phrases was performed. Two methods were considered in the classification. The first method only dealt with the original terms while the second method considered the original terms as well as the stem of the words in each term. The results revealed that the second method is more effective. These two methods could be improved by using more advanced natural language processing methods. Despite the existence of informal conversations in social media, the proposed system was able to classify approximately 20% of the retrieved tweets. The information labelled by the system could be used as a benchmark to derive informal keywords which are specific to a transit agency. These keywords could be added to the lexicon to improve the performance of the classification system. As a future step, sentiment analysis will be added to the current semantic analysis. This will convert the suggested LOS classification system to a measurement system which could be comparable to traditional measurement systems evaluated by CSat surveys.

REFERENCES

American Association of State Highway and Transportation Officials, 2009. *AASHTO Transportation Glossary.* 4th ed. American Association of State Highway and Transportation Officials (AASHTO), Washington, D.C., USA.

Atkinson, A. A., 1998. Strategic Performance Measurement and Incentive Compensation. *European Management Journal,* 16(5): 552-561.

Atkinson, A. A., Waterhouse, J. H. & Wells, R. B., 1997. A Stakeholder Approach to Strategic Performance Measurement. *Sloan Management Review,* 38(3): 25-37.

Bititci, U. S., Carrie, A. S. & Mcdevitt, L., 1997. Integrated Performance Measurement Systems: a Development Guide. *International Journal of Operations & Production Management,* 17(5): 522-534.

Bourne, M. C. S., Neely, A. D., Mills, J. F. & Platts, K. W., 2003. Implementing Performance Measurement Systems: a Literature Review. *International Journal of Business Performance:* 1-24.

Bregman, S., 2012. *Uses of Social Media in Public Transportation,* Transportation Research Board of the National Academies , Washington, D.C., USA.

Cox, L., Butcher, L., Byrne, S. & Meier, A., 2013. *Australian Transport Index Thesaurus,* ARRB Group Ltd. Melbourne, Australia.

Forza, C. & Salvador, F., 2000. Assessing Some Distinctive Dimensions of Performance Feedback Information in High Performing Plants. *International Journal of Operations & Production Management,* 20(3): 359-385.

Franco-Santos, M. et al., 2007. Towards a Definition of a Business Performance Measurement System. *International Journal of Operations & Production Management,* 27(8): 784-801.

Gates, S., 1999. Aligning Strategic Performance Measures and Results. *The Conference Board Trusted Insights for Business Worldwide*, New York, USA.

Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., Zadeh, R.,, 2013. WTF: The Who to Follow Service at Twitter Committee (IW3C2). *International World Wide Web Conference*, Rio de Janeiro, Brazil.

Ittner, C. D., Larcker, D. F. & Randall, T., 2003. Performance Implications of Strategic Performance Measurement in Financial Services Firms. *Accounting, Organizations and Society,* 28(7-8): 715-741.

Kaplan, R. S. & Norton, D. P., 1996. Linking the Balanced Scorecard to Strategy. *California Management Review*, 39(1): 53-79.

Kerssens-van Drongelen, I. C. & Fisscher, O. A., 2003. Ethical Dilemmas in Performance Measurement. *Journal of Business Ethics,* 45(1): 51-63.

Kittelson & Associates, Inc.; Parsons Brinckerhoff; KFH Group, Inc.; Texas A&M Transportation Institute; Arup, 2013. *Transit Capacity and Quality of Service Manual,* Transportation Research Board of the National Academies, Washington, D.C., USA.

Kittelson & Associates, Inc.; Urbitran, Inc.; LKC Consulting Services, Inc.; Morpace International, Inc.; Queensland University of Technology; Nakanishi, Yuko, 2003. *A Guidebook for Developing a Transit Performance-Measurement System,* Transportation Research Board of the National Academies, Washington, D.C., USA.

Lebas, M. J., 1995. Performance Measurement and Performance Management. *International Journal of Production Economics,* 42(1-3): 23-35.

Lynch, R. L. & Cross, K. F., 1991. *Measure Up!: The Essential Guide to Measuring Business Performance*, Mandarin, London, UK.

Maisel, L. S., 2001. *Performance Measurement Practices Survey Results*, American Institute of Certified Public Accountants, USA.

McGee, J., 1992. *What is Strategic Performance Measurement?*, Ernst & Young Center for Business Innovation, Boston, USA.

National Research Council Transportation Research Board, 2001. *Transportation Research Thesaurus and User's Guide,* National Academy Press, Washington, D.C., USA.

National Research Council Transportation Research Board, 2010. *Highway Capacity Manual.* 5th ed. Transportation Research Board of the National Academies , Washington, D.C., USA.

Neely, A. D., 1998. *Measuring Business Performance: Why, What and How,* The Economist and Profile Books Ltd., London, UK.

Neely, A. D., Gregory, M. J. & Platts, K., 1995. Performance Measurement System Design: A Literature Review and Research Agenda. *International Journal of Operations & Production Managment,* 15(4): 80-116.

Neely, A. D., Gregory, M. J. & Platts, K., 2005. Performance Measurement System Design: a Literature Review and Research Agenda. *International Journal of Operations & Production Management,* 25(12): 1228-1263.

Otley, D., 1999. Performance Management: A Framework For Management Control Systems Research. *Management Accounting Research,* 10(4): 363-382.

Spiteri, L., 1998. A simplified model for facet analysis: Ranganathan 101. *Canadian Journal of Information and Library Science,* 23: 1-30.