Quantifying the Offline Interactions between Hosts and Guests of Airbnb

Full Paper

Abdulkareem Alsudais
Claremont Graduate University
Abdulkareem.alsudais@cg.edu

Abstract

In this paper, the offline interactions between hosts and guests of Airbnb are investigated. While the platform-supported communications between hosts and guests are easily tracked, new solutions are required to quantify the offline interactions. These interactions were investigated through the development of an IT artifact that determines if a review written by a guest includes a mention of a host. Manual labeling of 1,024 randomly selected reviews indicated that 85% of reviews include a reference to a host. Two primary patterns in which hosts are mentioned were discovered. A new method to detect if a host is referenced in a review is proposed. The method is based on automatically detecting these patterns using Word Embeddings and Named Entity Recognition. The method achieved an accuracy score of 91.5% and was applied on thousands of reviews from Airbnb. Results demonstrated that over 80% of reviews include references to hosts.

Keywords

Sharing economy, text mining, natural language processing, virtual communities.

Introduction

In recent years, there has been an increase in the number of platforms that provide services in the “Sharing Economy,” which is also referred to as “Collaborative Consumption.” Zervas et al. (2016) defines the Sharing Economy as platforms that “enabled individuals to collaboratively make use of under-utilized inventory via fee-based sharing.” These platforms such as Uber, Lyft, and Airbnb tend to utilize technological advances in areas that include Mobile Computing, Geographical Information Systems, and Online Social Networks and the wide breadth of Internet access. The rapid rise and immense popularity of such platforms have motivated researchers and scientists from a variety of disciplines to study topics related to the Sharing Economy. These topics include investigations into the economic structures of these platforms (Horton and Zeckhauser 2016; Wallsten 2015), the rationale of individuals who use them (Hamari et al. 2016; Zekanovic-Korona and Grzunov 2014), and the impacts of specific platforms on well-established and traditionally stable industries (Cramer and Krueger 2016; Zervas et al. 2016). However, there remain multiple facets of the Sharing Economy that have yet to be explored. The purpose of this paper is to delve into one of these topics: Quantifying the real-world and offline personal interactions between hosts, or property owners, and guests of Airbnb.

Airbnb is a platform that allows hosts to list rooms and properties in an online marketplace that consumers can employ to search and book listings. While there are other platforms that offer similar services such as VRBO and Roomorama, Airbnb remains the most utilized platform in the space. Studying specific issues that pertain to Airbnb, the features that define the service, and the reasons behind its popularity might reveal new knowledge about the hidden characteristics that influence and motivate consumers to use a given platform in the Sharing Economy. These characteristics might be overlooked when investigating why some consumers prefer to use such platforms over traditional services. Furthermore, platforms in the Sharing Economy might have unique attributes based on the sector or industry they are in. Thus, a “one-size-fits-all” approach to study platforms in the Sharing Economy may not always be the optimal route to follow, whereas a focus on a specific platform might reveal more valuable information.
In platforms that are commonly classified as being part of the Sharing Economy, there are typically three primary actors: the provider who offers a tangible or intangible resource, a consumer who seeks to use that resource, and an IT artifact that facilitates the interaction between the provider and the consumer. Both the provider and consumer can either be a person or a business (Schor 2016). In the case of Airbnb, the provider is the host who offers a space in their property, or their entire property. The consumer is the guest who seeks a room or property to rent. The online and “virtual” or “cyber” interactions between the provider and the consumer are documented and easily tracked through private messages and online ratings. However, in order to elucidate the characteristics of the real-world and offline interactions between a guest and a host, additional processing steps are required.

In this paper, the fundamental objective is to investigate and quantify the offline interactions between hosts and guests of Airbnb. The goal is to determine the quantity of which guests explicitly mention in their reviews a reference to their real-world interactions with their hosts. A sample of randomly selected 1,024 reviews written by guests is employed to determine if a review includes a reference to a host and to recognize patterns used by guests to mention their hosts. Following this, a method is developed to automatically detect if a review contains a reference to a host. This method is based on several widely-accepted Natural Language Processing and Text Mining techniques. The method also relies on recent advances in the development of Word Embeddings models. These models create highly accurate word relatedness measures (Baroni et al., 2014). There are four primary contributions for this paper. First, a preprocessing pipeline that detailed necessarily steps required to remove problematic reviews from a dataset that consists of reviews from Airbnb. This pipeline can be replicated by others working on similar datasets. Second, an automated method that can be employed to detect if a review contains a reference to a host. This method can be utilized in subsequent research studies. Third, an investigation into the quantity in which reviews contain mentions of hosts. The findings indicate that over 80% of reviews contain references to hosts. Finally, an unexpected contribution is that reviews from Airbnb aggregated by cities can be used to potentially predict travel trends and discover attributes commonly shared by residents of cities.

**Related Work**

Researchers and scientists in the Information Systems field and related scientific communities have studied several aspects of the sharing economy and Airbnb. Ikkala & Lampinen (2015) conducted semi-structured interviews with 12 hosts and concluded that the two main motivations for hosting are financial gain and social rewards. Ma et al. (2017) investigated what hosts reveal about themselves, and how they use that as a way to signal trustworthiness. Furthermore, they studied how this perceived trustworthiness affect booking decisions guests make. In a similar study, Tussyadiah (2016) examined how hosts attempt to portray themselves in Airbnb and what nouns and adjectives they use in their profiles. A text mining solution was performed to cluster words according to their cooccurrences metrics. Based on that, the author proposed a classification of five types of hosts. The types were “the creative, the global citizen, the local expert, the personable, and the established” (Tussyadiah 2016). Ert et al. (2016) investigated the roles hosts’ profile images and perceived trustworthiness play in determining booking decisions. They found that consumers select a place based on both the place’s attributes and the host’s attributes. Their research indicates that perceived trustworthiness of hosts is partially determined by their posted profile images. Furthermore, these images affect perceived trustworthiness and prices’ prices. Fagerstrom et al. (2017) conducted a similar study but with a focus on hosts’ facial expressions in photos. They concluded that while prices are the primary motivator for renting decisions, hosts’ facial expressions and their posted ratings are the second determiners. Additionally, they found that positive or neutral facial expressions increase the likelihood of booking. Finally, Gutt & Herrmann (2015) investigated the changes hosts make to prices after they receive their first rating. They discovered that hosts increase their prices after they receive a first rating.

---

**Figure 1. The Basic Framework of Airbnb. Offline Interactions Between Hosts and Guests are Often Overlooked.**

In platforms that are commonly classified as being part of the Sharing Economy, there are typically three primary actors: the provider who offers a tangible or intangible resource, a consumer who seeks to use that resource, and an IT artifact that facilitates the interaction between the provider and the consumer. Both the provider and consumer can either be a person or a business (Schor 2016). In the case of Airbnb, the provider is the host who offers a space in their property, or their entire property. The consumer is the guest who seeks a room or property to rent. The online and “virtual” or “cyber” interactions between the provider and the consumer are documented and easily tracked through private messages and online ratings. However, in order to elucidate the characteristics of the real-world and offline interactions between a guest and a host, additional processing steps are required.

In this paper, the fundamental objective is to investigate and quantify the offline interactions between hosts and guests of Airbnb. The goal is to determine the quantity of which guests explicitly mention in their reviews a reference to their real-world interactions with their hosts. A sample of randomly selected 1,024 reviews written by guests is employed to determine if a review includes a reference to a host and to recognize patterns used by guests to mention their hosts. Following this, a method is developed to automatically detect if a review contains a reference to a host. This method is based on several widely-accepted Natural Language Processing and Text Mining techniques. The method also relies on recent advances in the development of Word Embeddings models. These models create highly accurate word relatedness measures (Baroni et al., 2014). There are four primary contributions for this paper. First, a preprocessing pipeline that detailed necessarily steps required to remove problematic reviews from a dataset that consists of reviews from Airbnb. This pipeline can be replicated by others working on similar datasets. Second, an automated method that can be employed to detect if a review contains a reference to a host. This method can be utilized in subsequent research studies. Third, an investigation into the quantity in which reviews contain mentions of hosts. The findings indicate that over 80% of reviews contain references to hosts. Finally, an unexpected contribution is that reviews from Airbnb aggregated by cities can be used to potentially predict travel trends and discover attributes commonly shared by residents of cities.

**Related Work**

Researchers and scientists in the Information Systems field and related scientific communities have studied several aspects of the sharing economy and Airbnb. Ikkala & Lampinen (2015) conducted semi-structured interviews with 12 hosts and concluded that the two main motivations for hosting are financial gain and social rewards. Ma et al. (2017) investigated what hosts reveal about themselves, and how they use that as a way to signal trustworthiness. Furthermore, they studied how this perceived trustworthiness affect booking decisions guests make. In a similar study, Tussyadiah (2016) examined how hosts attempt to portray themselves in Airbnb and what nouns and adjectives they use in their profiles. A text mining solution was performed to cluster words according to their cooccurrences metrics. Based on that, the author proposed a classification of five types of hosts. The types were “the creative, the global citizen, the local expert, the personable, and the established” (Tussyadiah 2016). Ert et al. (2016) investigated the roles hosts’ profile images and perceived trustworthiness play in determining booking decisions. They found that consumers select a place based on both the place’s attributes and the host’s attributes. Their research indicates that perceived trustworthiness of hosts is partially determined by their posted profile images. Furthermore, these images affect perceived trustworthiness and prices’ prices. Fagerstrom et al. (2017) conducted a similar study but with a focus on hosts’ facial expressions in photos. They concluded that while prices are the primary motivator for renting decisions, hosts’ facial expressions and their posted ratings are the second determiners. Additionally, they found that positive or neutral facial expressions increase the likelihood of booking. Finally, Gutt & Herrmann (2015) investigated the changes hosts make to prices after they receive their first rating. They discovered that hosts increase their prices after they receive a first rating.
Methodology

To develop an automated method that detects if a review written by a guest includes a reference to a host, it is imperative to first investigate the ways in which guests mention their hosts in reviews. In this section, a description of a qualitative analysis of a sample of 1,024 randomly selected reviews written by guests is first given. Based on this examination, two patterns in which guests mention their hosts in reviews are identified. Two complementary methods to automatically detect these two patterns are proposed.

Qualitative Analysis of Reviews on Airbnb

The purpose of this qualitative analysis was to survey a sample of English reviews posted on Airbnb to identify patterns and ways in which guests explicitly stated a fact about their real-world and offline interactions with their hosts. To accomplish this task, a randomly generated sample that consisted of 1,024 reviews was developed. These reviews were selected from a large dataset comprised of 159,537 reviews that pertained to listings in the San Francisco area. The dataset was preprocessed to ensure the validity of the reviews (details on the preprocessing step are contained in the subsection “Preprocessing” under the section “Experiment”). The randomized selection was conducted using the “random” function of the python package pandas (Mckinney 2010).

For each review in the sample, two tasks were completed. The initial task was to decide if a guest mentioned their host in the review. The second task was to specify a portion of the text in the review that provided proof of the reference. Table 1 shows a selected number of reviews. The “host mentioned?” column indicates whether the guest included something in the reviews that referred to their offline and real-world interactions with the host. The “indicator” column presents the word that provided proof that the review should be labeled as one in which an interaction with a host was mentioned.

<table>
<thead>
<tr>
<th>Review</th>
<th>Host mentioned?</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ming was easy to communicate with and I could access her place even though she wasn’t at home at that time.</td>
<td>Yes</td>
<td>“Ming”</td>
</tr>
<tr>
<td>Had a great weekend, for us this was the perfect location. Walking distance to Golden Gate Park, the Legion of Honor and Ernesto’s Italian Restaurant. A nice spacious flat.</td>
<td>No</td>
<td>NA</td>
</tr>
<tr>
<td>This was one of the best experience I have ever had on AirBNB. The host is very nice and discreet, the place is clean and comfortable</td>
<td>Yes</td>
<td>“host”</td>
</tr>
<tr>
<td>It is great for our family, just besides one of the China Town. Location is great and the owner is very helpful.</td>
<td>Yes</td>
<td>“owner”</td>
</tr>
<tr>
<td>We had a decent stay at Leon’s for one night. Area was interesting with great late night eats. Only met Leon briefly upon checking out.</td>
<td>Yes</td>
<td>“Leon”</td>
</tr>
</tbody>
</table>

Table 1. Selected Reviews to Demonstrate How Hosts are Mentioned

Pattern One: Referring to Hosts by Their Roles

In this discovered pattern, a guest referenced their host in a review by explicitly mentioning one of the hosts assumed roles, such as “host” or “owner.” To capture and accurately tag reviews in which hosts are mentioned by one of their roles, an important first step was to determine and discover the words that guests used in their reviews to refer to their hosts. In other words, a “lexicon” that consisted of a list of words which guests used to refer to their hosts was required. While it was possible to qualitatively examine reviews to manually populate this lexicon, such process is nevertheless incomplete and prone to Type II errors. Thus, a reliable alternative that leveraged existing and widely-accepted Natural Language Processing and Text Mining techniques to automatically populate entities in the lexicon was deemed as more appropriate.
Word Embeddings refers to “models that create dense vector representations for words or phrases in a text corpus, by utilizing their immediate syntactic context, defined by a window with proximate terms” (Alsudais et al. 2016). Recently, these models have gained popularity, partially due to advances in computational power that have made creating vectors for large corpora more feasible and less time intensive. Successful models for the generation of vector representations for words and phrases include Skip-gram and CBOW models (Mikolov et al., 2013), and GloVe (Pennington at al., 2014). One of the popular features of these models is one that captures semantic similarities between words in a corpus. For a given word, the model generates a list of the most similar words in the corpus to the given word, based on their shared usage context. For each word in the list, the model also generates a value that represents how similar a word is to the given word, with a value of one, indicating that the words are identical. According to Baroni et al. (2014), Word Embeddings models, which rely on an attempt to “predict” word vectors, perform better at capturing semantic similarity between words than traditional “counting” methods. In this study, this “word relatedness” feature is employed to populate a lexicon that consist of host-related words.

To populate the lexicon of host-related words, both CBOW and Skip-gram were run on a corpus that consisted of 439,414 reviews written by guests who visited the Los Angeles area. All characters in the reviews were changed to lower case to ensure that “Host” and “host” were captured as the same entity. The number of dimensions for both models was set to 300 and the window sizes was set to eight. After the models were ran, two lists of the most semantically similar words to the word “host” were retrieved. In Figure 2, a selected number of words are positioned in the figure based on their vectors as generated by CBOW with semantically similar words being placed in the same space.

The lists revealed the most common words guests used to refer to their hosts. For the CBOW model, the top five terms in the list and their semantic similarity scores were: “hostess (0.76),” “communicator (0.64),” “landlord (0.52),” “guy (0.51),” and “hosts (0.50).” Excluding “communicator,” these words are words that guests used to refer to their hosts, and were therefore synonyms for host in this context. In the list of the top 20 most similar words to “host”, words such as “communicator,” “conversationalist,” and “professional” also appeared. These words were not contextual synonyms for “host” and perhaps appeared in the list because they were commonly used in reviews in a similar context to “host”. Thus, an additional step was required to remove words in the lists that were not contextual synonyms for “host”, to ensure that the lexicon of host-related words was satisfactory. After performing this step, the lexicon was populated with the following words: “host”, “hostess”, “hosts”, “hostesses”, “landlord”, “landlords”, “owner”, “owners”, “guy”, “staff”, “man”, “woman”, “lady”, “hostes”, “hoster”, “gentleman”, and “person”.

Remarkably, the models captured a misspelled word: “hostes,” and a word that does not exist in the English language: “hoster.” While adding words such “person” and “man” to the lexicon might increase the likelihood of Type I errors (False Positive), an observation was made that there were more instances where these words were used to reference hosts than ones where they were used in other contexts. However, more rigorous testing is required to investigate the impact of the inclusion of these words.
Pattern Two: Referring to Hosts by Their Real Names

In this discovered pattern, guests referenced their hosts in a review by mentioning the host’s given name. To automatically detect if a host’s name is mentioned in a review, a first attempt was made based on matching a given name found in a review to the host’s name, as found in the listing page for the place. This attempt failed due to several problems, such as instances where guests mentioned hosts using their nicknames and where guests used an alternative spelling of a host’s name. As a result, this basic text-matching operation failed. An alternative method that leverages Named Entity Recognition (NER) proved superior. NER is the task of processing a body of text and classifying segments in the text that are named entities. Named entities refer to names of people and organizations that exist in the real world. An NER classifier was employed to label reviews and find if a name of a person was used in a review. If labeled as “True”, a review was classified as one where a guest mentioned a host. While there are instances of guests mentioning their significant other’s name, a name of a pet that lives in the place, or a name of another person they met while staying at the place, these instances were rare and should not significantly affect the results. However, additional testing is required.

A number of highly accurate algorithms to capture named entities currently exists (Finkel et al, 2005; McCallum & Li, 2003). Recent work in this area focuses on increasing the accuracy of taggers by leveraging advances in neural networks (Chiu & Nichols, 2016; Lample et al., 2016). However, these new models are still only slightly more effective than traditional ones. Thus, in this paper, Stanford Named Entity Recognizer (Finkel et al, 2005) was used. In this study, Stanford NER is performed on reviews that do not include a word that is in the lexicon of host-related words. Stanford NER can be implemented using several classifiers. The classifier entitled “english.all.3class.distsim.crf” was selected in this study after experimenting with others. For each word in a review, the tagger tries to classify if the word is a named entity. If it is, the tagger also provides a type for the entity. The type can be “person”, “organization”, or “location”. A review is labeled as one where a host is referenced when a named entity of the “person” type was detected in the review.

Experiment

Dataset

The dataset used in this paper was compiled by insideairbnb.com (2017). The site is not affiliated with Airbnb and their descriptions indicate that listings and reviews from several selected major cities in the world were scraped at various points of time. Datasets from this website have been used in other academic papers (Ma et al. 2017; Tussyadiah 2016). Three major cities in the United States were selected and used in this study. The cities were Los Angeles (LA), New York City (NYC), and San Francisco (SF). The number of reviews used from each location and the date these reviews were scraped are in Figure 2.
Preprocessing

Preprocessing is often a required step when text data is analyzed. The purpose of this step is to clean the data and to modify, or remove, problematic objects in the dataset. The type of preprocessing needed is problem-dependent. For this dataset, preprocessing was performed to remove automatically generated reviews that Airbnb create, empty reviews that contain no text, and non-English reviews.

The first set of reviews that needed to be removed were automatically generated reviews that Airbnb create to signal that a host has decided to cancel a reservation. These kinds of reviews are perhaps generated to discourage hosts from canceling reservations. These reviews were easily flagged since they included the same sentence: “This is an automated posting.” The second type of reviews that were removed were empty reviews. These reviews contained no text. The third type were reviews not written in English. To detect the natural language of reviews, a language detection tool entitled langid.py was used (Lui and Baldwin 2012). langid.py was trained on 97 natural languages. According to its authors, it achieved an accuracy ranging from 0.88 to 0.98 when it was evaluated on commonly used corpora. In this paper, langid.py was used to classify the natural language used in the reviews. Subsequently, any non-English reviews were removed.

Descriptive Statistics

One of the initial findings of this study was the relatively large number of problematic reviews that were required to be removed before attempting any further analysis. For example, at least 13% of reviews from NYC were unfit and were not used in the study because they were automatically generated, empty, or not written in English. Therefore, other researchers should consider this fact before utilizing similar datasets.
An unexpected finding was the large number of non-English reviews from the NYC area dataset compared to the data from LA and SF. A possible explanation is that, according to an annual report by the U.S. Department of Commerce’s National Travel and Tourism Office (2015) for the year 2014, NYC was ranked first in the number of overseas visitors to cities in the United States. According to the same report, NYC had a market share of 28.3% while LA and SF had a market share of 12.8% and 9.1% respectively. This is an indicator that patterns detected from Airbnb can potentially be used to predict country-wide tourism statistics and trends.

**Method Evaluation**

To evaluate the method proposed to detect if guests mentioned hosts in reviews, the randomly generated and manually tagged sample of 1,024 reviews was used. For this dataset, each review was labeled as either “True” when a host was mentioned or “False” when a host was not mentioned. The method was evaluated by comparing the “True” or “False” labels as identified by the method to those created manually by the annotator for each individual review. To provide clear quantitative evaluation for the method, three well-established measures were used. The three measures were precision, recall, and accuracy. To calculate the performance of the method in these three measures, four variables needed to be calculated. The variables were “True Positive” (TP), which refers to the number of reviews that were correctly labeled “True” by the method, “True Negative” (TN), which refers to the number of reviews that were correctly labeled as “False” by the method, “False Positive” (FP), which refers to the number of reviews that were falsely labeled as “True” by the method, and, “False Negative” (FN), which refers to the number of reviews that were correctly labeled as “False” by the method. As the results in Table 4 demonstrates, the method performed greatly on all three measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
<th>Calculation</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>( \frac{TP}{TP + FP} )</td>
<td>( \frac{806}{806 + 26} )</td>
<td>96.8%</td>
</tr>
<tr>
<td>Recall</td>
<td>( \frac{TP}{TP + FN} )</td>
<td>( \frac{806}{806 + 61} )</td>
<td>92.9%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>( \frac{TP + TN}{TP + TN + FP + FN} )</td>
<td>( \frac{806 + 131}{806 + 131 + 26 + 61} )</td>
<td>91.5%</td>
</tr>
</tbody>
</table>

**Table 3. Summary of Measures**

The final task in this study was to apply the method proposed to detect if a review included a reference to a host on randomly selected large datasets from Los Angeles, New York City and San Francisco. The objective of this step was to attempt to reach generalizable findings based on applying the method on large datasets.
In this paper, the offline interactions between guests and hosts of Airbnb were investigated. A new method was developed to detect if reviews written by guests included mentions of hosts. The method was based on the automated detection of two patterns guests use to reference their hosts. Well-established Natural Language Processing and Text Mining techniques were leveraged to develop the method. The method achieved a precision score of 96.8%, a recall score of 92.9%, and an accuracy score of 91.5% when tested on a set comprised of manually labeled reviews. Nine datasets from three locations were used to apply the method on larger datasets. Findings suggest that the offline interactions and communications guests have with their hosts are important, as evidenced by the quantity in which guest mentioned their hosts in their reviews. This could be a new clue for understanding the barriers and motivators to participation in platforms in the Sharing Economy space. Perhaps some users prefer Uber or Lyft over traditional taxi services because the quality of personal interactions with drivers is higher. Other unexpected findings suggest that Airbnb can be used to forecast travel trends and discover features shared by residents of cities.

While the results of the quantitative measures used to evaluate the method proposed in this paper were high, the method nevertheless has several limitations. First, the method is still prone to failure in capturing instances where guests mentioned their hosts in reviews. For example, a guest might refer to a host by using a word that is not one of the commonly used words in the host-related lexicon. Additionally, it is possible for the Stanford NER to fail at tagging named entities of the person type. Thus, even though a review does contain a given person's name, the method will not be able to capture it. Second, the method relies on the assumption that if a given name is used in a review, that given name is the host's name. However, reviews where guests mentioned pets' names were found. Finally, the dataset only includes reviews from guests who write reviews. Thus, individuals who use Airbnb but elect not to write reviews are not represented in this dataset.

**Table 4. Summary of Number of Reviews where an Offline Interaction with a Host is Mentioned. The Percentages Indicate the Percentages of Reviews with Hosts Mentioned**

<table>
<thead>
<tr>
<th>Location</th>
<th>First Sample of 10,240 Reviews</th>
<th>Second Sample of 10,240 Reviews</th>
<th>Sample of 40,960 Reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles</td>
<td>8,176 (79.8%)</td>
<td>8,129 (79.3%)</td>
<td>32,557 (79.4%)</td>
</tr>
<tr>
<td>New York City</td>
<td>8,507 (83.0%)</td>
<td>8,432 (82.3%)</td>
<td>33,778 (82.4%)</td>
</tr>
<tr>
<td>San Francisco</td>
<td>8,360 (81.6%)</td>
<td>8,442 (82.4%)</td>
<td>33,640 (82.1%)</td>
</tr>
</tbody>
</table>

For each one of the three locations, three sets that consisted of randomly selected reviews were constructed. The number of reviews in each dataset were 10,240; 10,240; and 40,960 reviews. In total, nine sets were used. All the reviews were selected from the processed dataset, and reviews were selected regardless of their sentiments.

As the results in Table 5 demonstrates, in all datasets, the percentages of reviews where hosts were mentioned were remarkably high. This was an additional indicator that guests indeed value their offline interactions with their hosts. Additionally, the results were similar to ones found when a sample of randomly selected 1,024 reviews were qualitatively tagged. This further validated the method proposed and demonstrated its effectiveness when used on larger datasets. Finally, the similarities between the values for the smaller two samples and the larger one suggest that it may not be necessary to test the method on the entire dataset. This fact is useful and will be referred in forthcoming studies when the method is tested on new locations and additional datasets.

One of the first noticeable findings is that the values for NYC and SF are similar while they are seemingly lower for LA. There is no clear reason why this is the case, and additional investigation is required. However, one possible explanation is that residents of NYC and SF might be more welcoming and friendly than those who live in LA. This is an indicator that these reviews could be used to develop an understanding of the characteristics shared by residents of cities.

**Conclusion, Discussion, and Limitations**

In this paper, the offline interactions between guests and hosts of Airbnb were investigated. A new method was developed to detect if reviews written by guests included mentions of hosts. The method was based on the automated detection of two patterns guests use to reference their hosts. Well-established Natural Language Processing and Text Mining techniques were leveraged to develop the method. The method achieved a precision score of 96.8%, a recall score of 92.9%, and an accuracy score of 91.5% when tested on a set comprised of manually labeled reviews. Nine datasets from three locations were used to apply the method on larger datasets. Findings suggest that the offline interactions and communications guests have with their hosts are important, as evidenced by the quantity in which guests explicitly reference their hosts in their reviews. This could be a new clue to the barriers and motivators to participation in platforms in the Sharing Economy space. Perhaps some users prefer Uber or Lyft over traditional taxi services because the quality of personal interactions with drivers is higher. Other unexpected findings suggest that Airbnb can be used to forecast travel trends and discover features shared by residents of cities.

While the results of the quantitative measures used to evaluate the method proposed in this paper were high, the method nevertheless has several limitations. First, the method is still prone to failure in capturing instances where guests mentioned their hosts in reviews. For example, a guest might refer to a host by using a word that is not one of the commonly used words in the host-related lexicon. Additionally, it is possible for the Stanford NER to fail at tagging named entities of the person type. Thus, even though a review does contain a given person's name, the method will not be able to capture it. Second, the method relies on the assumption that if a given name is used in a review, that given name is the host's name. However, reviews where guests mentioned pets' names were found. Finally, the dataset only includes reviews from guests who write reviews. Thus, individuals who use Airbnb but elect not to write reviews are not represented in this dataset.
study. It is possible that guests who write reviews are more social, and the offline interactions with hosts are more important to them compared to those who do not write reviews.

There are at least five implications for this study. First, the preprocessing performed on the dataset indicates that other researchers utilizing review data from Airbnb should perform similar steps to reduce the number of problematic reviews. Second, the same steps and tools can be replicated to process similar datasets. Additionally, the method to determine if a host was explicitly mentioned in a review can be employed in several subsequent studies. For example, it can be used to generate sentiment scores for hosts and to investigate if the prevalence of explicit mentions affect hosts’ overall ratings. Third, the high prevalence of reviews where guests explicitly mentioned their hosts could be an indicator that the ease of offline interactions with hosts is a strong attribute that influences the quality of the experience. Perhaps it is also a motivator, or a barrier, to why individuals prefer to use or not use services in the Sharing Economy. Fourth, recent developments suggest that major industries are moving toward the automation of several of their services. Many of these automated services will eliminate the need to have traditional face-to-face interactions between a consumer and a representative of the establishment. Findings from this study may be indicators that consumers value such interactions.

References


