

Exploring Collected Intelligence from Untappd to Support the Location Decision for New SMEs

Thiago H Silva
Federal University of Technology - Paraná
Applied Computing Graduate Program
Curitiba, Brazil
thiagoh@utfpr.edu.br

Alexandre R Graeml
Federal University of Technology - Paraná
Applied Computing Graduate Program
Curitiba, Brazil
graeml@utfpr.edu.br

ABSTRACT

Small and medium-sized enterprises (SMEs) usually have little budget for market research. That prevents them from developing sophisticated surveys or carrying out focus group sessions with current and potential customers to better understand their needs and wishes. Social media may be a rich source of information to support sound business decision making. This paper discusses the possibility of collecting data that people share in messages on social media and analyzing their content in order to build relevant knowledge for strategic business decisions, among which, location decisions. To exemplify that, we collected and analyzed messages generated by Brazilian users of Untappd, a mobile phone app for sharing beer drinking experiences. Among other things, it was very easy to notice that users from a few cities in the country have developed a more sophisticated taste for beer and are keen to try new types of beer and tell their friends about it. Is information like that useful for an entrepreneur who wishes to settle a new small craft beer business and needs to decide on the best place to do it? We believe so.

Keywords

Collected intelligence, social media mining, Untappd, customer knowledge, business location decision

1. INTRODUCTION

The decision on where to locate a business enterprise has always challenged entrepreneurs. There is a lot to take into account and it's a very strategic decision, considering its impact on the future of the enterprise and the difficulty in reverting it, in most cases. A large multinational, with operations around the globe, may strategically decide to locate a plant in a specific country or continent to increase its influence on that region of the planet. On the other hand a small supermarket chain may decide to concentrate its activities in the periphery of a metropolitan area, where the competition against the large and strong players in the retail industry is not as fierce. Sometimes, location decisions involve an analysis that is not so concerned with local details and particularities. That happens, for example, when a decision is made by an automaker with respect to the country, or even continent, where it will launch an automobile engine manufacturing plant. In other instances, deci-

sions are very punctual, though still strategic for their long term consequences. When an entrepreneur has to decide on the best location within a city for a new small venture, this is precisely the case. Some of the concerns involve the size of the realty that is necessary for the operation as originally planned, but also accounting for future expansion, local zoning (i.e., legal authorization to locate a specific kind of business in a particular area in the city), tax benefits for settling here or there, environmental issues, urban mobility (to make it easy for customers and employees to reach the company's premises) and so on.

In this paper, we are particularly interested in discussing business location from a market perspective, i.e., assessing where the business would find the most favorable conditions to prosper, based on information of previously existing demand. It should be remarked, though, that the reasoning used by the decision maker to make up his/her mind based on the information about the existence of previous demand in a specific geography escapes the scope of the work. After all, who has never heard the story about two shoe sellers sent to foreign distant lands to explore new potential markets? When they came back each had a completely different opinion: One thought that there was no opportunity for the company's product because no one wore shoes in the prospected region, while the other was enthusiastic about the huge unexplored market for the company's product. Our main purpose is, therefore, to analyze the possibility of using social media data to help an entrepreneur to decide on the most favorable location for settling his/her small enterprise. In order to accomplish that, we will use the fictitious case of an entrepreneur who is interested in starting a craft beer brewpub somewhere in Brazil. We could have chosen any other venture for our study, as people share information virtually on anything in social media, these days. But this specific kind of business tends to involve small enterprises, because craft product businesses, by definition, do not scale without turning into an industrial venture. And, although location decisions are important both for big and small companies, small and medium-sized enterprises (SMEs) need to find creative ways of obtaining good information to support their strategic decision making, because they do not have the same knowledge and resource means that are available to large organizations. Intelligence collected from social media may provide good support for this kind of decision, possibly representing a cheaper alternative to traditional sources of information or, at least, an interesting complement to other approaches.

2. COLLECTIVE INTELLIGENCE

Collective Intelligence is generating an increasing interest as an emerging discipline [1, 8]. It provides us with "a grand vision" but it is still not what we get from the social web, according to Gruber [3]. In its current stage of development, users' contributions in

In: Sessão de Pôsteres., 2016, Teresina. Anais do XXII Simpósio Brasileiro de Sistemas Multimídia e Web. Porto Alegre: Sociedade Brasileira de Computação, 2016. v. 2.

ISBN: 978-85-7669-332-1

©SBC – Sociedade Brasileira de Computação

social media are collected and aggregated, generating results that could be more suitably described as the “collected intelligence” of independent users [7]. Maher [5] describes collected intelligence as that obtained from individuals who, each, provide their own independent contribution to resolve a specific challenge. It is a particular type of collective intelligence that does not require direct interaction among participants, because it does not aim at achieving any sort of synergistic solution based on the individuals’ collaboration, still according to Maher [5]. As stressed by Silva et al. [8], Schoder [7] and Gruber [3], result is obtained based on collective aggregation, which can be performed without further involvement or even awareness of those who provided the collected data.

For collected intelligence, the “crowd” hands the data over, and an external agent collects and processes it, feeding the collective results back to individuals that will make use of it [8], and who are not necessarily the same people who contributed to the collective aggregation, in the first place.

The fact that collected intelligence can be achieved without the necessary awareness of those involved, or even their willingness to put any additional effort into achieving goals or solving problems that may not concern them directly, turns it into a cost effective way to address such situations, many times. This is the main reason we wished to apply its concepts to support the decision making of an SME: as no one needs to be directly consulted and no one needs to get away from one’s own way, in order to support a project like this with the required data, it seems particularly interesting when there is no or little budget for data collection and analysis.

3. METHODOLOGICAL PROCEDURES

We collected data from Twitter messages (called tweets) triggered by users of Untappd, a mobile phone application that allows users to provide their friends with information on beer that they drunk and where they did it, among other things.

Almost 55 thousand tweets were collected from June 28 to July 28, 2013. Posts included a text message following a structured pattern, created by Untappd for automatic broadcasting on Twitter, from which all the data we used could easily be extracted.

The following messages are examples of collected tweets included in our dataset: “Drinking an Anniversary Ale by @lostabbey @ The Bad Apple”; “Love this place — Drinking a Cidre Bouché de Cru - Antoinette by Cidrerie Duch de Longueville at @murrayscheezbar”; “Drinking an American IPA by Win2 Brewing”¹.

As we can see, the type of beer comes after “Drinking a/an” and before “by <name of a company> @ <name of a place>”. This pattern makes data parsing easy using just about any programming language. Location information (latitude and longitude), the date of the post and the language in which it was written were also extracted from the collected tweets. However, the name of the country or city was not directly available. We needed to calculate the distance from the location of the post (latitude and longitude) to cities whose coordinates were publicly available. We used the “geonames” data set² for that purpose and calculated the smallest distance from the location of the post to the central point of a known city or town, in order to determine city and country for each tweet. In order to analyze the data, we used open software available online to generate the plots on maps and word clouds.

¹Here we included examples from English speaking users from different geographies to improve readability.

²<http://www.geonames.org>.

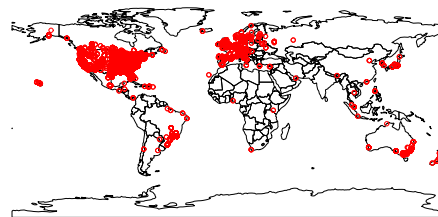


Figure 1: Geographical distribution of collected tweets.

4. UNTAPPING UNTAPPD

Untappd was launched in 2010 [6] and defines itself as “a new way to socially share a brew you’re currently enjoying, as well as where you’re enjoying it, with your friends!”³. According to Chorley et al. [2] there were more than one million users just three years after launching, who were responsible for over sixty million check-ins by then.

In this section we present a characterization of user activity in Untappd. The world map presented in Figure 1 plots the one hundred thousand plus tweets that we were able to collect over a month from Untappd users, showing their geographical location following the procedure described in the previous section.

Untappd is much more popular in North America (85.4% of the tweets) than anywhere else in the world. However, some countries in other continents also had relatively high levels of participation. There were, approximately, 2,200 tweets from England, 1,000 from Australia and 700 from Brazil, during the analyzed period.

As our fictitious decision maker is a Brazilian entrepreneur, and assuming that when thinking about starting a new small business people probably only consider their own country for location, in an attempt to reduce additional sorts of uncertainties, among which cultural differences and legal and fiscal issues, we will focus on analyzing the data we have collected for Brazil.

Although it represents only around 1% of the whole database of collected tweets, the records for the country will still hold interesting information to support our entrepreneur’s business location decision making.

Figure 1 already gives us a hint on the fact that the responses are not evenly distributed in the country. Most of Untappd users in Brazil live in the south-east of the country, an area that is more densely populated and also richer than other areas. São Paulo and Rio de Janeiro, the two largest metropolises in South America, were responsible for approximately 30% of the records for Brazil. Curitiba ($\approx 9\%$) and Belo Horizonte ($\approx 6\%$), though much smaller in size, also call attention for the interest of their inhabitants for craft beer, assuming that Untappd tweets are a good proxy for that.

In the next section, we analyze the data for these four cities in a little more detail.

5. ZOOMING INTO THE UNTAPPD DATA

Word clouds involve highlighting the more frequently used words or terms that are present in a dataset one is analyzing, so that they are more prominently represented. Although, as a qualitative analytical tool, word clouds are only recommended as a preliminary analysis tool for scientific work, or for validation of previous findings, they are able to provide a decision maker who already has some knowledge about the phenomenon s/he is observing with some relevant additional information and good insights.

³See the About section in the app’s blog page at <http://blog.untappd.com/>.



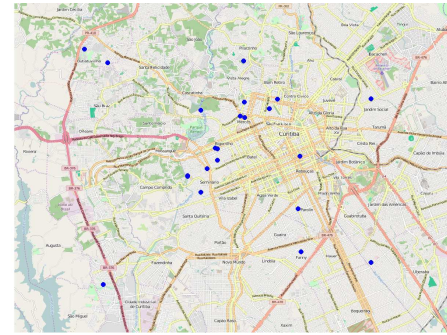
Figure 2: Word clouds for types of beer and brands.

When we analyze the word clouds of tweets for the four chosen Brazilian cities (see Figure 2), it is not difficult to notice that Untappd users from Curitiba and Belo Horizonte are more sophisticated than the ones from São Paulo and Rio, with respect to their beer choices. They mention styles of beer more often than brands and, although São Paulo and Rio had almost twice as many tweets and, although São Paulo and Rio had almost twice as many tweets as Curitiba, and three times more tweets than Belo Horizonte, the difference in size of word clouds is not proportionally larger.

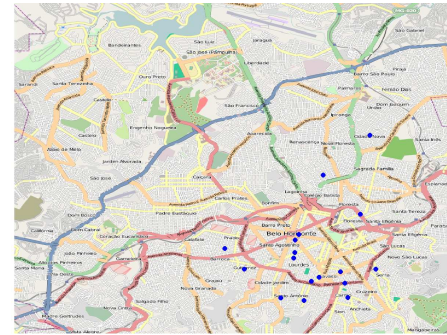
This kind of information is interesting for an entrepreneur who intends to start a brewpub, even though it does not allow us to know how it will affect his/her decision. Settling the business in Curitiba or Belo Horizonte would expose the new venture to customers that are better educated about the product a craft brewer has to offer. Perhaps, the market is already there for craft beer. But, at the same time, there are already competitors around, which could make the entrants' life harder. On the other hand, locating the business in São Paulo or Rio de Janeiro would mean that competition would probably not be so direct, as craft beer has not yet taken off in those cities. However, mass beer breweries, which are able to practice much lower prices, would compete for the money of the uneducated customers. And for those who can't make the difference between mass market beer and craft beer, the product may be perceived as a commodity, being sold by an abusive price.

For the continuation of our analysis, we will suppose that our fictitious entrepreneur thought it would be better to settle the business in a city where there is already a tradition among customers of drinking craft beer. After all, most craftsmen do not see other craftsmen as competitors. Many times, they consider them even as partners, which cluster geographically to provide customers with plenty of choice. And besides, although competition may be more intense within the category, it may be easier to sell the product at a premium price when other entrepreneurs are helping to make the market understand that there is value in the more elaborated product a brewpub or a craft brewer can offer.

The blue dots in Figure 3 show the coordinates of tweets. One interesting thing that we noticed was that they are well spread around the cities. They do not seem to cluster in gastronomic areas of



(a) Curitiba



(b) Belo Horizonte

Figure 3: Scatter plot of all places where users shared data for Curitiba and Belo Horizonte.

Curitiba and Belo Horizonte, for example. We decided to check the names of places where those beers were being drunk. We suspected that maybe people were drinking their beers at home, as this is a possibility with craft beer, particularly when there are many home brewers that distribute their production directly to friends and neighbors. But that did not turn out to be the case. Figure 4 shows the result, once again in the form of word clouds. The names that appear there for both cities are names of brewpubs and craft beer shops, in most of the cases.

Such a result shows that these two cities are indeed different to other large cities in Brazil where people will drink beer at restaurants or regular pubs, with much less emphasis to the craft beer movement.

6. RELATED WORK

With respect to the object of the study, we are not the first ones to try to extract knowledge from data generated by users of Untappd. Chorley et al. [2] crawled data from that same social app, though with a much more obvious intent: understanding the drinking habits of people through the lenses of social media. One of the pros of using Untappd data, according to Chorley et al. [2] is that a formal schema is used for all check-ins, which makes them very straightforward, requiring no interpretation or translation when organizing the data for analysis. We fully agree with that. Just about



Figure 4: Names of places where beer was drunk.

anyone can collect data from that social media and the data is easy to aggregate and analyze.

With respect to the objective of the work, there have been a few attempts, in recent years, to use location-based applications' data in order to study the strategic problem of deciding the best location for a new business. Karamshuk et al. [4], for example, studied the problem of optimal allocation of retail stores in the city. They used Foursquare data to assess the popularity of three retail chain stores in New York, based on the number of check-ins by users of that social media platform. The authors evaluated a diverse set of features, modeling spatial and semantic information about the places and patterns of user movement in the area around the analyzed locations. They observed that the existence of places in the vicinity that naturally attract large volumes of people, such as a train station or an airport, as well as the existence of other retail stores of the same type close by, were the strongest indicators of popularity.

Our study is different to those concerning the same object or related objective because we focus on the possibility of improving strategic decision making in the case of SMEs, which have little budget for market research. In order to accomplish that, we presented a fictitious case study about where to settle a new craft beer business. We hope that, even having used simple tools to analyze a data set that could be explored in much more sophisticated ways, we were able to make our point: there is a lot of good information out there, waiting for whoever is willing to put effort into harvesting it and making sense of it.

7. CONCLUSION

In this ongoing study, we only analyzed the most direct information Untappd users provided us with their beer drinking check-ins. We believe that what we did is enough to prove the relevance

of social media data for location decision analyses, even (or particularly) in the case of small enterprises that have little financial resources for market research. There are many other possibilities that were not explored here for brevity, which are also feasible and inexpensive. As we have data on the time check-ins took place, it is possible to analyze, for example, when people drink their beers. Wouldn't that information also help decide on location? Presumably, people drink after work, during weekdays. On weekends, the patterns of entertainment may be different. But how different would the work day and weekend patterns be? Some areas in a large metropolis are safer than others in the evening (information could even be crossed with other sources, like <http://wikicrimes.org>) to further improve the quality of decision on location.

It is important to point out some possible limitations of data obtained the way we did. First, it may reflect the behavior of a fraction of consumers. Our collection was based on data shared by users of Untappd on Twitter. Therefore, there could be biases relating to the fact that the users of such application are not necessarily representative of the population of craft beer drinkers. They are likely to be young, owners of smartphones, and urban dwellers. Untappd application has an English interface, which would surely scare away the majority of the less educated citizens of countries that do not have English as their mother tongue. Consequently, urban areas with older and poorer populations will provide less data and be underrepresented in whatever analysis is made. Besides, users may not share data concerning all of their destinations, considering the info will be made public on Twitter. Thus, our dataset might offer a partial view of consumers habits, which needs to be taken with care.

In spite of all that, it does provide solid aggregate information that, when contrasted with the entrepreneur's own experience, can surely help improve his/her confidence on the decision being made.

Acknowledgment

This work was partially financed with resources from FAPEMIG, Fundação Araucária, Capes and CNPq.

8. REFERENCES

- [1] E. Bonabeau. Decisions 2.0: The power of collective intelligence. *MIT Sloan management review*, 50(2):45, 2009.
- [2] M. Chorley, L. Rossi, G. Tyson, and M. Williams. Pub crawling at scale: tapping untappd to explore social drinking. In *Proc. of ICWSM'16*, 2016.
- [3] T. Gruber. Collective knowledge systems: Where the social web meets the semantic web. *Web Semantics*, 6(1):4–13, 2008.
- [4] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo. Geo-spotting: mining online location-based services for optimal retail store placement. In *Proc. of KDD'13*, pages 793–801, 2013.
- [5] M. L. Maher. Design creativity research: From the individual to the crowd. In *Design Creativity 2010*, pages 41–47. Springer, 2011.
- [6] T. Mather. Untappd connects beer drinkers everywhere. untappd - drink socially. In *Untappd Blog*, 2010.
- [7] D. Schoder, P. A. Gloor, and P. T. Metaxas. Social media and collective intelligence—ongoing and future research streams. *KI-Künstliche Intelligenz*, 27(1):9–15, 2013.
- [8] T. Silva, P. Vaz De Melo, J. Almeida, and A. Loureiro. Large-scale study of city dynamics and urban social behavior using participatory sensing. *Wireless Communications, IEEE*, 21(1):42–51, Feb 2014.