Smarter Cities, Safer Travels: Intergrating Contextual Suggestion

Adriel Dean-Hall
University of Waterloo
adeanhal@uwaterloo.ca

Jack Thomas
University of Waterloo
j26thoma@uwaterloo.ca

ABSTRACT

Contextual suggestion is meant to provide meaningful venue recommendations tailored to a personal profile, but sometimes a user's information need goes beyond their taste in restaurants. This paper seeks to demonstrate that integrating information concerning crime reports and public safety alerts into suggestions allows for more informed decision-making by users. We will explore existing sources of this public safety data, demonstrate its potential use and examine the implications – positive and negative – implied by the approach.

1. INTRODUCTION

A common issue for travelers visiting a new city is not knowing the lay of the land. Finding restaurants, shops and entertainment venues compatible with their interest is an obvious concern, but this can also be extended to the public safety geography. A street map is not enough to tell visitors which roads are well-lit at night or which neighborhoods are currently experiencing higher-than-average crime rates. This information can be just as important to a user's decision as any part of their personal preferences.

Where once tourists might read guidebooks to identify high-risk neighborhoods, the modern traveler uses websites such as Trip Advisor[1] or Virtual Tourist[2]. The internet brims with reviews and anecdotes about cities, as well as users looking to make informed decisions while abroad.

The development of smart cities provides an opportunity to meet this information need. Municipal governments collect vast reams of data concerning public safety, from crime statistics to accident reports to the status of infrastructure. As the paradigm shifts toward making this data open and accessible to users, contextual suggestion developers can draw on this information to improve the suggestions they make.

"Contextual suggestion" covers an ongoing field of research at the border of search and recommendation meant to provide personalized recommendations to users for "points of interest", such as venues and other attractions. The subject has garnered interest[2] for providing travelers with a convenient way to explore an unfamiliar city using their personal preferences as a guide.

These early forays have been marked by limited scope, concerned solely with the user's tastes. Now that smart cities are expanding the data horizon to include new subjects, the time has come to consider safety's potential as a recommendation feature.

2. RELATED WORK

Recommender systems for contextual suggestion have already been the topic of considerable research, such as Garcia et al.'s paper in 2011 [4]. The authors make great strides in producing a system which adapts its recommendations according to the tastes, demographics and history of the user. Baltrunas et al.'s paper from the same year[1], emphasizes the importance of the user's context when making recommendations, such as the season, available transport, who they're traveling with and so on.

However, these papers do not consider the safety of an attraction, nor is any data relevant to that conclusion provided for the user. This is not to say that the value of public safety data has not been explored. As early as 2001, Estivill-Castro et al.'s work[3] demonstrated how crime statistics, one part of public safety, can be mined and mapped.

Smart city projects have made great strides in building safety databases. In 2011, the city of St. Louis participated in IBM's Smarter Cities Challenge[3] in an effort to improve synergy between separate law enforcement agencies within the city. At IBM's recommendation, the city created a plan for a centralized crime database accessible by all agencies so that a unified view of all offenders and offenses could be made. The focus so far has been on enhancing communication within government, but these same data sources could be made available to the public.

[1] http://www.tripadvisor.ca/Travel-g28970-s206/Washington-Dc:District-Of-Columbia:Health.And.Safety.html

[2] http://www.virtualtourist.com/travel/North_America/United_States_of_America/Washington_DC/Warnings_or_Dangers-Washington_DC-TG-C-1.html

[3] http://smartercitieschallenge.org/city_st_louis.html

ECIR'14 Information Access in Smart Cities Workshop (i-ASC 2014).
April 13, 2014, Amsterdam, the Netherlands.

Some smart cities infrastructure is already being used by other applications and fields of study to enhance their results. Even the Van Gogh museum in Amsterdam has found uses for smart city data to modify recommendations made to tourists. Through the European Union's CitySDK project[4], sensors have been installed in the museum to detect the presence of lines and crowds[5], letting prospective visitors know when the museum is busy and how long the expected wait will be. This museum system is a textbook application for how smart cities data is already being used to enhance the user experience.

Between existing research into contextual suggestion, the potential of public safety data and the success of similar smart cities initiatives, integration seems a natural avenue to explore. In the next section we will begin this exploration by taking profiles and data drawn from the TREC 2011 track on contextual suggestion and integrating it with real, publicly available crime statistics.

3. DEMONSTRATION

The 2013 Contextual Suggestion TREC track gathered, from several systems, personalized point-of-interest suggestions for multiple users and cities. In this track each system made ranked suggestions to users and the users rating the suggestions based on their interest in visiting the attraction [2]. One of these cities, the one which we will focus on, is Washington D.C., in this section we will combine crime incident reports with the suggestions that the systems made for the area.

District of Columbia Data Catalog [5] contains crime reports for a 10 km radius around the Washington, D.C. area. This information is provided as a live feed as well as reports that contain all the crimes for a given year. For this paper we use the 2013 data, a full implementation will probably find it valuable to use the most up to date information.

Our strategy is to provide an indicator to users about how many crimes have occurred in the area around the attraction they are considering visiting. A simple first attempt at doing this would be to count how many crime incidents occurred within a certain distance of the attraction:

$$S(y) = |x \in X | D(x, y) < \Theta|. \quad (1)$$

Here X is our set of crimes, y is our attraction, D is the distance, in kilometers, between two locations calculated using the Haversine formula, and Θ is our threshold for how far the crime has to be for it to be considered. So, $S(y)$ is the safety score of the attraction where lower numbers mean the attraction has had fewer crime incidents nearby and we assume are safer.

However, we also want to consider how long ago the crime occurred because crimes that occurred more recently are more likely to be an indication of safety, we can incorporate this into our equation:

[4] http://www.citysdk.eu
[5] http://data.dc.gov

TREC Rank	Title	Score
1	Zaytinya	1104.86
2	Bistrot Du Coin	610.58
3	Old Ebbitt Grill	406.15
4	Blue Duck Tavern	353.16
5	Founding Farmers	484.29

Table 1: Safety score for suggestions.

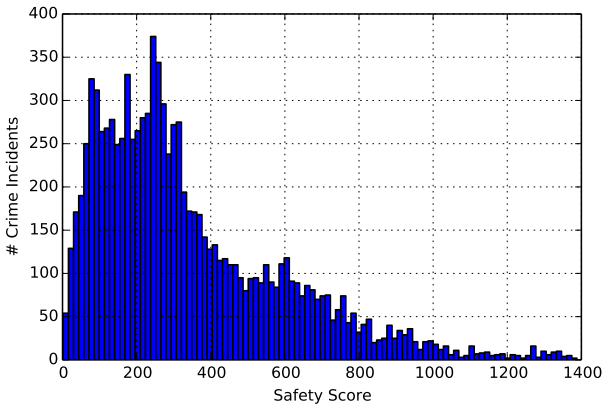

Figure 1: Safety score (equation 2) of locations where a crime occured.

$$S(y) = \sum_{x \in X | D(x,y) < \Theta} \frac{1}{\log(1 + \epsilon + T(x))}. \quad (2)$$

Here T is how many days have passed since the crime occurred divided by 365, so crimes that occured more recently are weighted higher. This gives us a safety score that warns users more heavily about crimes that occurred more recently. In this equation $\epsilon = 1.0$ and prevents division by zero. For this paper we pick a distance threshold $\Theta = 0.5$. The reason we choose this equation is that it gives more weight to crimes that have occurred more recently.

As an example, we can pick suggestions made by one of the systems (UDInfoCS1) for one of the profiles (554) as part of the contextual suggestion track. Here we we equation 2 to calculate the safety score of the top 5 suggestions, as ranked by the TREC system, for this user. These scores can be seen in table 1. In order to calculate the score we choose December 5th, 2013 as the date that the user is searching.

Now that we have a safety score we need to get context as to what this number means in terms of relative safety in the city. In order to do this we calculate the safety score at the location of every crime incident. Figure 1 shows the frequency of the safety scores given to locations where crimes have occurred in Washington, D.C.

In order to provide users with an easily digestible safety score we map the score given in equation 2 to 1-5 depending upon if the score is below the 20% percentile, between the 20% and 40% percentiles, between the 40% and 60% per-

Percentile	Score
0%	2.08
20%	180.55
40%	278.32
60%	428.66
80%	667.03
100%	1388.53

Table 2: Safety score percentiles.

TREC Rank	Title	Mapped score
1	Zaytinya	5
2	Bistrot Du Coin	4
3	Old Ebbitt Grill	3
4	Blue Duck Tavern	3
5	Founding Farmers	4

Table 3: Safety score for suggestions.

centiles, between the 60% and 80% percentiles or above the 80% percentile. of the scores given to areas where crimes have occurred Table 2 lists the percentiles for the scores in figure 1.

With this data we can now calculate our mapped safety scores for the sample suggestions as shown in table 3. Note that there are other possibilities that could be used to normalize safety scores. In particular, one possibility would be to factor in population density around the attraction location into the normalized scores.

Expanding on the use case of having safety scores to indicate the threat to personal safety we can note that the types of crimes have been annotated provide more detailed information about the type of crime in the area. The types of crimes are listed in table 4. One possible way to break down the type of threat is to differentiate between threat to personal safety and threat to the safety of a visitor's car. This information can be especially useful to visitors that are deciding where to park their car, or whether to leave their car at their hotel or not.

Using this information we can provide two safety scores: personal safety and vehicle safety. The only modification to equation 2 is to only consider certain types of incidents when calculating the score. In the types of incidents listed in table 4, "Motor Vehicle Theft" and "Theft From Auto" are the two that need to be considered when calculating the

Type	# of Occurences	% of Occurences
Theft/Other	12453	35.42%
Theft From Auto	9917	28.21%
Robbery	4080	11.60%
Burglary	3346	9.51%
Motor Vehicle Theft	2634	7.49%
Assault With Dangerous Weapon	2289	6.51%
Sex Abuse	295	0.83%
Homicide	103	0.29%
Arson	35	0.09%

Table 4: Offense occurences.

Rank	Title	Vehicle	Personal
1	Zaytinya	4	5
2	Bistrot Du Coin	4	5
3	Old Ebbitt Grill	1	4
4	Blue Duck Tavern	3	3
5	Founding Farmers	1	5

Table 5: Personal and vehicle safety score for suggestions.

vehicle safety score, these two combined make up 35.7% of the incidents. All the other types of incidents are considered when calculating the personal safety score.

Following the same process of developing a general safety score we can develop the vehicle and personal safety scores, the score for our example can be seen in table 5. Now we can see that in these suggestions there is more of a concern with regards to personal safety than to vehicle safety.

The mean general safety score associated locations that crimes occurred shown in figure 1 is 426. In comparison to that the mean vehicle safety score is 145 and the mean personal safety score is 281. The purpose of reporting these two score separately is so that areas that travellers need to be aware of one type of safety warning but not the other can be identified. We note that the Kendall tau coefficient of vehicle vs. personal safety scores is $\tau = 0.5224$, which shows that these areas are somewhat correlated but there are still areas where reporting these metrics separately can be useful.

We calculate the safety scores for all (roughly 600) attractions in Washignton, D.C. that were given an interest rating by users as part of the Contextual Suggestion Track. Here, the Kendall tau coefficient of interest rating vs. safety score is $\tau = -0.0044$. We can see that regardless of whether the user likes or dislikes the attraction the range of given safety scores is similar. Liked attractions don't have a tendency to be in either safe or unsafe areas and so providing a safety score to users will help them make decisions.

Other information could also be taken into account when calculating safety scores. Firstly, the time of day the user is searching could be taken into account: a visitor might not need to be concerned about an area when crimes occur during the night if they are visiting during the day. Secondly more severe crimes could be weighted more heavily, for example homicides or crimes with a gun could be given more weight in the safety score.

4. DIRECTIONS FOR FUTURE WORK

Our demonstration above has shown the power of just one data source in one city to allow users to make more informed decisions. This is just the beginning, and as more cities develop their information infrastructure these data sources will grow in number. Those cities leading the way will reap the earliest benefits by being the first to offer these services to citizens and visitors, which will in turn intensify advocacy for the creation and release of more databases elsewhere.

Making safety data collected by governments available to the public has been a cornerstone of many smart cities ini-

tiatives. Some sources, like the St. Louis one, remain closed, but data concerning public safety is not limited to the government, or to crime reports. The Open311 project, run by the nonprofit OpenPlans, effectively crowdsources the gathering of non-emergency public safety information, from accidents to utility breakdowns. Open311 collects data in a number of cities worldwide, even being integrated into other smart cities projects such as CitySDK to help cities communicate their data to developers. While standard adoption is always a delicate issue in the early days of any field, a savvy developer might be wise to keep their eye on Open311's progress.

Another possibility to consider is extending the personalizing of profiles beyond merely the user's taste in venues. By integrating their security needs and habits into their profile, a system can provide more personalized alerts. A user who owns a car will naturally be more concerned about reports of car theft in an area they intend to park, while a user who regularly spends time in high-crime areas of a city will be less likely to be discouraged by security warnings surrounding a venue. Security profiles can be built alongside the existing preference profiles, by looking at the safety scores for those attractions the user has already visited.

5. IMPLICATIONS

While we have focused so far on the positive potential of integrating public safety information, it must be acknowledged by anyone looking to implement these ideas that there are real hazards. The most immediate one is that recommending a restaurant according to someone's tastes and recommending a course of action for their personal security are two wildly different actions with different consequences to match.

A poor restaurant recommendation might ruin an evening, whereas a mistaken safety report could put someone in jeopardy. Even if the report is accurate, having that information might lull some users into a false sense of security when they should still be alert. A mere suggestion system cannot possibly take full responsibility for the safety of its users, nor can a developer guarantee that an area reported as safe could not be the site of a crime. Developers must be aware of this when designing systems, noting that they provide only supplementary information and do not take the place of caution and common sense.

Another unavoidable issue is the locality of crime. Neighborhoods and vicinities with a high crime frequency still have businesses who may not appreciate customers being "scared off". Residents of these neighborhoods may also not appreciate their homes being labeled a high-risk area, or the implications of tourists being steered away from them. Many databases also currently include personal information, such as the perpetrators or victims of crime, whose use should be avoided to protect the privacy of residents.

These are valid concerns, which is why the decision to act on available data must remain firmly with the user. Suggestion systems, like tourist guidebooks and websites, should remain focused on merely providing publicly available data to promote informed decision-making. Likewise, those who maintain smart cities databases should take care not to release unnecessary personal information to the public. The misrepresentation of a neighborhood and those who live in it can be damaging in many ways, and a thoughtful developer will consider this in the design of their system.

Having discussed these pitfalls, it might seem as though the risks of integrating public safety information outweigh the benefits. However, keeping public safety information outside of contextual suggestion does not mean that users won't factor in safety when making their decisions. If anything, by not presenting public data, users are more likely to make ill-informed decisions based on hearsay, prejudice or outdated information. While we can avoid responsibility by not making any pretense of speaking to safety, we cannot diminish the user's real information need.

6. CONCLUSIONS

In this work we have demonstrated the power and potential of public safety data when integrated with contextual suggestion. With just crime data from a single city, we can provide users with the information and analysis necessary to promote safer, smarter decision-making - the overarching goal of the smart cities paradigm. A greater number of more varied public safety databases are emerging all the time, but it will take far more work to go from talking about their integration to implementing it.

There can be no doubt that this work is necessary. As information technology infrastructure continues to grow, cities will continue to get smarter, and developers will need to keep up. It behooves us to take advantage of these opportunities in order to better serve users, despite the many risks and challenges involved. Even with these new sources of information being made available, tackling the safety of users is an enormous responsibility. To shirk it because we fear the consequences would be the greater failure.

7. REFERENCES

[1] L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci. Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing*, 16(5):507–526, 2012.

[2] A. Dean-Hall, C. L. Clarke, J. Kamps, P. Thomas, and E. Voorhees. Overview of the trec 2013 contextual suggestion track. In *To appear in 22st Text REtrieval Conference, Gaithersburg, Maryland*, 2013.

[3] V. Estivill-Castro and I. Lee. Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data. In *Proceedings of the 6th International Conference on Geocomputation*, 2001.

[4] I. Garcia, L. Sebastia, and E. Onaindia. On the design of individual and group recommender systems for tourism. *Expert Systems with Applications*, 38(6):7683 – 7692, 2011.

[5] M. Groen, W. Meys, and M. Veenstra. Creating smart information services for tourists by means of dynamic open data. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 1329–1330, New York, NY, USA, 2013. ACM.