

# SocRoutes: Safe Routes Based on Tweet Sentiments

Jaewoo Kim  
KAIST  
Daejeon, Korea  
ez2rc@kaist.ac.kr

Meeyoung Cha  
KAIST  
Daejeon, Korea  
meeyoungcha@kaist.edu

Thomas Sandholm  
HP Labs  
Palo Alto, CA, USA  
thomas.e.sandholm@hp.com

## ABSTRACT

Location-based services, and in particular personal navigation systems, have become increasingly popular with the widespread use of GPS technology in smart devices. Existing navigation systems are designed to suggest routes based on the shortest distance or the fastest time to a target. In this paper, we propose a new type of route navigation based on regional context—primarily sentiments. Our system, called SocRoutes, aims to find a safer, friendlier, and more enjoyable route based on sentiments inferred from real-time, geotagged messages from Twitter. SocRoutes tailors routes by avoiding places with extremely negative sentiments, thereby potentially finding a safer and more enjoyable route with marginal increase in total distance compared to the shortest path. The system supports three types of traveling modes: walking, bicycling, and driving. We validated the idea based on crime history data from the City of Chicago Portal in December 2012, and sentiments extracted from geotagged tweets during the same time. We discovered that there was a significant correlation between regional Twitter posting sentiments and crime rate, in particular for high-crime and highly negative sentiment areas. We also demonstrated that SocRoutes, by solely utilizing social media sentiments, can find routes that bypass crime hotspots.

## Categories and Subject Descriptors

J.4 [Computer Applications]: Social and Behavioral Sciences;  
H.3.5 [Online Information Services]: Web-based services

## General Terms

Human Factors, Design

## Keywords

Sentiment analysis, Web applications, Navigation system, Safety

## 1. INTRODUCTION

A number of studies have shown the relationship between geography and well-being, where the environment and places have a

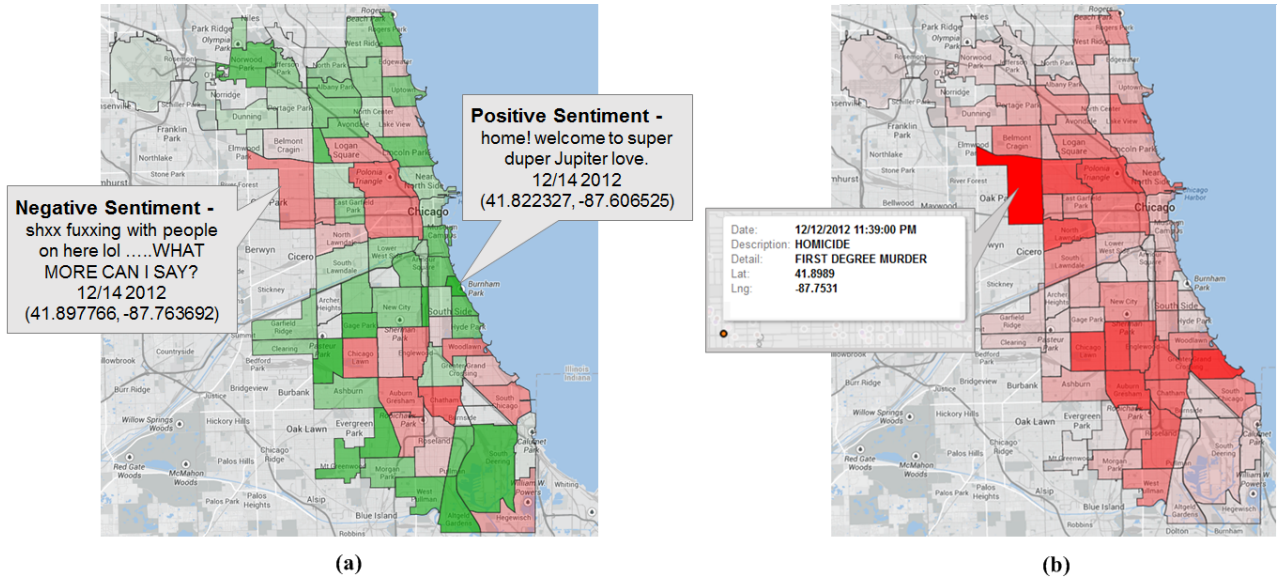
strong influence on one's happiness. In particular, the perception of safety in places is known as a major factor that determines social well-being [4]. People's sentiments change by the places they are in, where public parks incur more positive sentiment compared to transportation hubs [2]. Furthermore, people's perception and sentiments about places could potentially be used as an accurate measure of the actual crime rates [5]. These studies together indicate there are new opportunities to improve well-being in human mobility, by avoiding crime-prone areas and seeking more enjoyable places.

Existing personal navigation systems typically allow one to choose between the fastest and the shortest path. In improving well-being in human mobility, one can think of other criteria that could be meaningful, such as the friendliest, most scenic, safest, most enjoyable, or matching one's interest the best. For instance, regional information like "This road is so crowded" or "I saw someone pick pocket in this street again!"—the types of data that could be easily mined in online social media, could indicate that a user will likely have negative experience passing through that road, in particular when walking or bicycling.

While information about the crime and traffic history of public places are increasingly becoming available through the open government movements around the world, not all neighborhoods benefit from such data. Furthermore, there is a significant lag in the publication for the data that are available. This is where real-time, geotagged social media streams could play a role. Such big social data can be a rich source of information about places when gathered collectively, because messages are generated constantly and across wide geographic areas.

In this paper, we present a new system, named SocRoutes, that recommends routes based on regional context inferred from Twitter. By considering the regional context—namely sentiments and perceptions about places, SocRoutes could potentially avoid recent crime hotspots and find a safer, friendlier, and more enjoyable route. Regional context is important in a number of cases. For instance, one can use the application to plan a safe route to school for children. One may also like to avoid bad neighborhoods as a tourist.

SocRoutes utilizes the complete set of real-time geotagged data from Twitter and suggests routes by avoiding places with extremely negative sentiments, hoping to prevent any negative future experiences. Noting that users can have vastly different levels of interaction with places depending on how they travel, SocRoutes supports three types of traveling modes: walking, bicycling, and driving, and assumes that safety is of highest concern in the walking mode and of least concern in the driving mode, based on the exposure of as well as the ability to escape from unsafe situations. Given a source and a destination, SocRoutes finds safer and more enjoyable routes



**Figure 1: (a) Tweet sentiment map of Chicago at community level with example tweets. Green color means regions with more positive (than negative) sentiments and red means the opposite. (b) Crime density map with example homicide case in the Austin community.**

with marginal increase in total distance compared to the shortest path, by searching paths with small total amount of negative sentiments on the regions along the route.

We validated the idea of utilizing regional contexts with a use case from a representative city, Chicago, and examined a number of route pairs suggested by SocRoutes. When compared with the crime data published from the City of Chicago Data Portal, SocRoutes could effectively bypass recent crime hotspots solely by utilizing social media updates. This result matches with the recent finding that physical safety and crime rates show correlation with people’s perception about places [5].

This paper is organized as follows. We first present a pilot study that shows correlation between sentiments and crime in Section 2. We then describe the details of the SocRoutes system in Section 3. In Section 4, we conclude and discuss future work.

## 2. MOTIVATION

The crux of our idea is whether real-time, geotagged social media streams can be used to enhance everyday experience in the way we interact with places, for instance by avoiding crime-prone areas and seeking more enjoyable routes. A good example is provided in a recent effort of building a fine-grained mood map of New York City from Twitter streams [2], where the authors found that public parks incur positive sentiments, whereas transportation hubs incur negative sentiments. Also, people’s perception of a dangerous or safe place had acute relationship with the actual crime rates [5].

In order to identify a more concrete relationship between public sentiments and public safety, we conducted a pilot study and created the map of major residential areas based on both Twitter sentiments and crime history. We used the Gnip<sup>1</sup> API to Twitter, which accesses the full tweet streams, and gathered geotagged tweets over a 10-day period from December 8th to 17th in 2012. In order to consider only urban areas, we extracted all geotagged posts within a 25-mile radius from the 20,000 largest cities in the world, which yielded a total of 60 million tweets. In this paper, for lack of space, we demonstrate our system only with the data from

Chicago, which contained nearly 120,000 tweets. We use the Web analysis tool presented in [3] for extracting sentiments.

As a measure of safety in the community granularity, we used open government data. We downloaded the crime history data published from the City of Chicago Data Portal<sup>2</sup>, which contains information about the entire 4.2 million crimes committed from 2001 to 2012. Metadata consists of date, crime type, description, and the exact latitude-longitude geolocation of the crime scene. We eliminated all crimes such as deceptive, gambling, liquor violation, trespass, and prostitution, which may not directly impact the experience of travellers. After such filtering, we found a total of 12,503 reported crimes during the 10-day period in 2012.

Figure 1 shows the resulting maps of sentiments (left) and crimes (right) for 77 major communities in Chicago. Some correlation is discernible visually. The Austin community, located at West side of Chicago, was the most crime-prone area during the data collection period and is marked with the darkest red color on the crime map. Several other communities in the Southeastern area of Chicago such as Auburn Gresham, Chatham, South Shore, and Chicago Lawn also show high crime rates. For these areas, one can visually identify that negative sentiments are more dominant than the positive sentiments. The Pearson correlation coefficient was measured to be  $-0.55$  between crime rates and sentiments, where negative correlation means a high crime rate is related to less positive (i.e., more negative) sentiment. Given that tweets are a mix of daily chatter and personal updates as well as news, this level of correlation is promising. Furthermore, the top-10 precision and recall values for predicting the highest crime-rates using just tweet sentiments were both 0.5, compared to 0.14 for a theoretical random predictor.

This result is in tune with the recent finding, which identified that local environment and neighborhood have great impact on perception of safety related issues [1]. Based on findings of this pilot study, in the following section, we adopt Twitter sentiment as a feature for enhancing positive experience in human mobility and propose a new route navigation method, called SocRoutes.

<sup>1</sup><http://gnip.com>

<sup>2</sup><http://data.cityofchicago.org>

### 3. THE SOCROUTES SYSTEM

We propose a new system, called SocRoutes, that recommends routes based on regional context inferred from Twitter. By considering the regional context – namely sentiments and perceptions about places, SocRoutes could potentially avoid recent crime hotspots and find a safer, friendlier, and more enjoyable route. Our project is described in detail at <http://mia.kaist.ac.kr/project/socroutes/>.

#### 3.1 Key Components

The SocRoutes system consists of four major components, which we describe in detail next. First, in the *data filtering* module, public geotagged tweets are filtered so that potentially irrelevant tweets to local places are eliminated. For example, mentions or replies to a specific person (i.e., a tweet starting with a @-sign), retweets, and tweets containing any URLs are not considered for understanding regional context. In order to account for bias across users, we only consider the first legitimate tweet from each user, so that those who tweet proportionally more times than others do not account accordingly large in the trend.

Second, the *sentiment analysis* module infers sentiment of tweet posts. We take state-of-the-art implementation in [3] that combines the existing sentiment tools to achieve high coverage and accuracy in detecting sentiments in social media data. The sentiment analysis tool could predict sentiments from 70% for our tweet data. The sentiment value of each tweet is determined on a scale from -1 to 1, with increasing (positive) sentiment. A value of 0 means that neither a positive nor a negative sentiment could be inferred and these tweets are hence dropped from the analysis. In our dataset, we got nearly 47,000 tweets that have non-zero sentiment values from a total 120K. Thus, 39.2% of tweets express positive or negative sentiment – 26.4% are positive while 12.8% are negative. The mean value is 0.249, indicating tweets are slightly positive on the average.

Third, the *regional clustering* module aggregates tweets by region to determine a representative sentiment of the region. We utilized the Javascript library RouteBoxer<sup>3</sup>, which computes bounding boxes within a specified distance of a route. Compared to simple bounding box method, RouteBox automatically merges cells into a set of non-overlapping rectangular areas. Hence, our system only needs to calculate sentiments for tweets in a very limited, and targeted area.

Forth, the *route finding & visualizing* module finds a route based on the Google Maps Directions API<sup>4</sup>. This API provides a way to find a route from a source to a target destination, potentially through several waypoints, for various travelling modes such as walking, bicycling, and driving. We find new routes by modifying the direct default routes provided by the API to include waypoints that cause detours around areas with a high negative sentiment.

#### 3.2 Route Recommendation

SocRoutes offers a new metric, called *safety score*, by which users can pick a route for three types of travelling modes: walking, biking, and driving. The safety score is calculated by computing the percentage of the number of areas with sentiment value higher than a baseline *sentiment threshold* compared to the total number of areas on the route. We have set the default sentiment threshold as -0.7 for this study, yet this number can later be adjusted by a user. The areas a.k.a. boxes are determined by the regional clustering module using RouteBoxer algorithm.

<sup>3</sup><http://tinyurl.com/lyx5gtc>

<sup>4</sup><https://developers.google.com/maps/documentation/directions/>

Depending on the mode of travel, the safety score is computed at various levels of granularity (i.e., box sizes) to match the different levels of risk exposure inherent in the mode. In general, the faster one travels (e.g., driving instead of walking), the more shielded that user is from the environment. Therefore, SocRoutes provides smaller safety range for faster mode of travel. In the current configuration, 200, 100, and 80-meter ranges are used for walking, biking, and driving respectively. This means that a walking user would like to avoid extremely negative sentiment within a 200-meter radius of his path.

---

**Algorithm 1:** If the direct route contains a box (obtained from RouteBoxer) with a negative sentiment lower than the safety threshold, then SocRoutes may add a waypoint in the close waypoint around the negative box to avoid the potentially dangerous region at the cost of a small detour.  $w_{current}$  is current number of waypoints and  $w_{max}$  is maximum limit of waypoints.

---

```

while  $w_{current} < w_{max}$  do
  for all route boxes do
    if sentiment < threshold then
      for west north and east south waypoint candidates do
        if point is on the path then
          continue
        end
        else
          add waypoint
        end
      end
    end
  end
end
end

```

---

As displayed in Algorithm 1, SocRoutes adds waypoints depends on sentiment threshold but also on the total number of detours that may cause increase in travel time and distance. Maximum number of waypoints  $w_{max}$  may be specified by the user. We have set default value as 5, but a user can adjust it from 1 to 8 on the system according to preference – less number of waypoints may returns less detour but also lower safety score. After the new route has been suggested, a user may compare the default route with the suggested route based on distance, time of travel, as well as safety score; and also reconfigure the safety and time thresholds to find a more appropriate path.

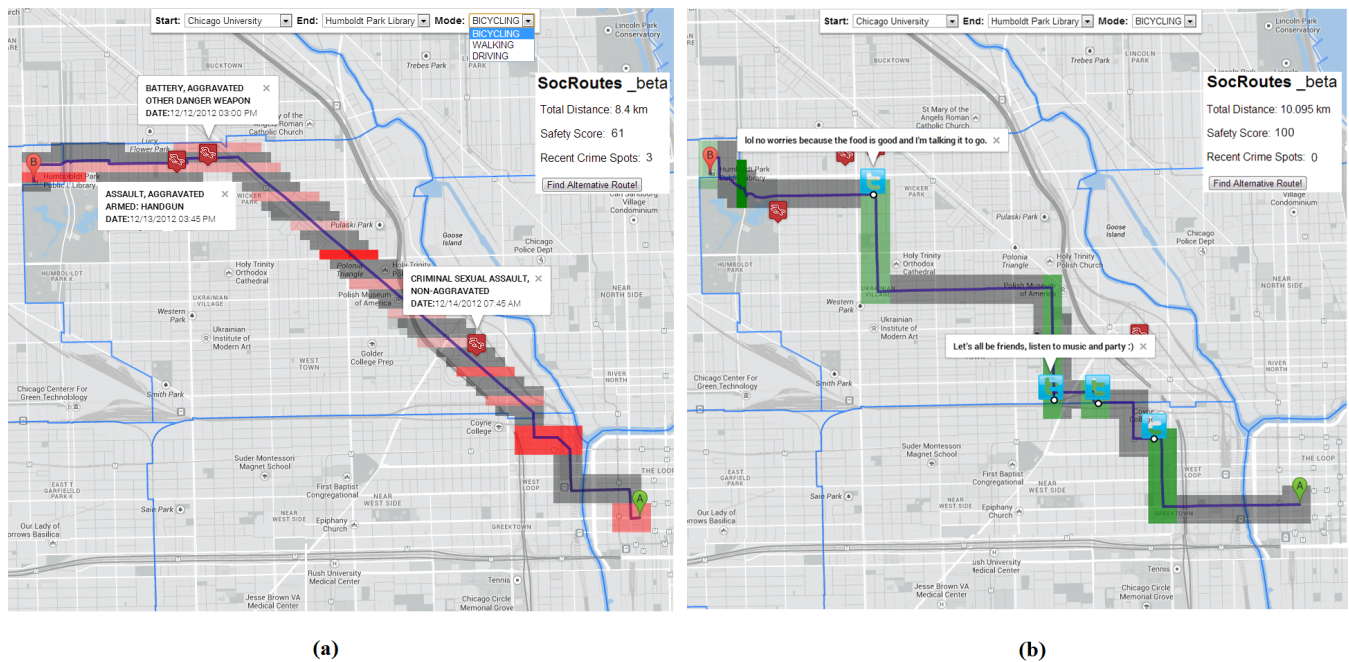
### 4. EVALUATION

We tested our demo system using 100 unique routes between 20 hand-picked endpoints in Chicago<sup>5</sup>. We performed the test under bicycle travelling mode and picked up the route returned first without iterations. Both sentiment threshold and maximum number of waypoints were kept at their their default values (-0.7 and 5).

New routes with less crime were discovered in 24 cases while increasing the distance by about 20% on average<sup>6</sup>. An example of a search result obtained from SocRoutes is shown in Figure 2,

<sup>5</sup>See <http://mia.kaist.ac.kr/project/socroutes/> for more details on endpoints chosen

<sup>6</sup>the current straw-man algorithm just picks random points at a fixed distance away from identified high-risk areas so we are working on improving these results by actively seeking out low-risk areas to re-route to



**Figure 2: Hypothetical example demonstrating the benefits of the SocRoutes system: (a) Example of default (shortest) bicycling route from Chicago University to Humboldt Park Library based on the shortest path algorithm. Red colored boxes mean regions with high degree of tweets with negative sentiments and the example showcases three violent crimes that occurred recently. (b) Alternative safe route suggested by SocRoutes, demonstrating that utilizing social media streams can lead to route with more positive public sentiments along the path and potentially avoid crime hotspots.**

which demonstrates that utilizing social media streams to reroute to areas with less highly negative public sentiments along the path can potentially avoid crime hotspots.

## 5. CONCLUSION

In this paper, we focused on one of the most deployed location based services, the personal navigation system, and suggested a new way to find routes based on regional context or sentiments about places. We started with the premise that fine-grained geo-tagged data from microblogs could provide valuable information about the sentiments people feel towards places, thus allowing one to find a safer and more pleasant route. Subsequently, we found that there was good correlation between sentiments and crime in a use case with data from the City of Chicago. Based on this idea and motivating use case, we proposed the SocRoutes system that finds routes for any given source and destination pair by avoiding extremely negative sentiments. Given a source and a destination, SocRoutes finds potentially safer routes with marginal increase in total distance compared to the shortest path. The demo presented shows that SocRoutes could avoid crime hotspots simply by detouring areas with negative tweet sentiments.

In future work we plan to explore the temporal correlation of crime and tweet posting content further. In addition, we expect that performance of the system can be improved by utilizing more sophisticated route searching algorithms such as actively seeking of low-risk areas. We also want to expand our system into a personal route recommender system that leverages positive sentiment and preference matching tweets in order to find more enjoyable paths, and evaluate it in user studies.

## Acknowledgements

We would like to thank Gisell Youn and Wonjae Lee for their help with this paper. This work was funded by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (2011-0012988).

## 6. REFERENCES

- [1] D. M. Austin, L. A. Furr, and M. Spine. The effects of neighborhood conditions on perceptions of safety. *Elsevier Journal of Criminal Justice*, 2002.
- [2] K. Z. Bertrand, M. Bialik, K. Virdee, A. Gros, and Y. Bar-Yam. Sentiment in new york city: A high resolution spatial and temporal view. *Tech report arXiv:1308.5010*, 2013.
- [3] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha. Comparing and combining sentiment analysis methods. In *ACM COSN*, 2013.
- [4] N. Powdhavee. Unhappiness and Crime: Evidence from South Africa. *Economica*, 2005.
- [5] P. Salesses, K. Schechtner, and C. A. Hidalgo. The collaborative image of the city: mapping the inequality of urban perception. *PloS One*, 2013.