University of California, San Diego



GoSafe: Designing the Fastest and Safest Path Throughout San Diego

Emmanuel Diaz; PID: A13548378

Isabel Li; PID: A14542428

Jesse Kim; PID: A13483476

Parker Addison; PID: A14852334

Roshan Fernando; PID: A14535553

COGS 9: Intro to Data Science

Professor Bradley Voytek, PhD.

15 December 2017

GoSafe: Designing the Fastest and Safest Path Throughout San Diego

**Introduction**:

GoSafe is an app which would work to improve safety in the streets by providing pedestrians with the shortest and safest path to their destinations. Our app would take into account both time and safety while calculating a path. This app would be valuable and cost-effective because the chance of a person being a victim of a violent crime in San Diego is 1 in 246, a statistic which is considered a “rate higher than in most communities of all population sizes”.[1] According to this article, violent crimes consist of considered robbery, assault, rape or murder.[1] What interested our team about these crimes specifically is that the victims of these crimes are often pedestrians rather than people inside cars. Thus, finding safer walking routes for pedestrians would most likely decrease the number of such crimes, so this app will be specifically for pedestrian routes.

The dataset we will reference comes from a website called SANDAG (San Diego Association of Governments), which is run by a committee of directors from the 19 different local governments in the San Diego region.[2] The data is in the form of a CSV file and includes crime reports from the last 180 days of San Diego crime data and is updated every Wednesday.[2] The website notes that some data may be missing, as certain incidents have yet to be approved as official reports. Other crimes may be left out because of their sensitive nature, including crimes involving vulnerable subjects.[2] The reports for such crimes are unavailable to protect the privacy of the subject.[2]

Some applications that take into account the safety factor while designing routes for pedestrian users have already been created. However, there are several differences between our idea and comparable ideas preceding ours. One such app, SketchFactor, was criticized for using crowdsourced comments which led to racism.[1] The perceived racism came from people designating certain ethnic neighborhoods as “unsafe” due to stereotypes and false perceptions, while in reality, the neighborhoods were generally safe. SketchFactor also does not rate the severities of different crimes. This means that something relatively harmless, such as a pickpocketing, could be weighed the same as a more serious crime, such as murder, in the SketchFactor app.[1] Our app would rely mainly on the police reported crime data, as these are cases which are important enough to include—and have a baseline crime severity—and we would show the relative risk of areas in different colors to show the varying severity for the users to see.

Another comparable app is currently being created by Boston University and University of Pittsburgh researchers, but their app only uses big data.[1] Since our app is specialized to the San Diego area, we can to ask locals as well as users of the app for their opinions of how safe they feel in certain areas of San Diego. By weighing the ratings of certain areas based on how close a local lives to the area, we would have more informed and less discriminatory knowledge to aid the big data—residents of a certain area may be more aware of the area’s current state. This is because locals likely have additional knowledge of an area that a tourist or visitor may not have. With this approach, both big data and local knowledge would be used to determine the safest route. Also, user feedback would be able to help us improve the app by seeing which routes tend to be rated the highest, and would give us additional data to feed into our algorithm.

In 2012, Microsoft obtained a patent for a GPS app called Pedestrian Route Production that takes into account harsh temperatures as well as unsafe routes.[3] Because Microsoft is also taking harsh temperatures into account, the data will have a different focus than our app, which focuses on safety and distance. It was also suggested that the Microsoft patent would include advertiser influence. Specifically, users would be deliberately taken past certain stores and advertisements.[4] This would skew the data and take even more focus off safety. Our app will stay nonpartisan, which will keep our data more pure and accurate. In addition, our rating system differentiates us from Microsoft’s app.

**Hypothesis**:

Can violent crime in the streets be reduced by designing a walking route that considers safety in addition to time? If the GoSafe app fulfills its function by designing short and safe paths for pedestrians, we expect the violent crime rate in the area of San Diego to decrease. Our routes throughout San Diego would be safer, more efficient, and overall better than paths created by routing programs which do not take into account crime hotspots. This prediction is justified because our app takes into account the safety factor as well as the route time, whereas most other routing apps just take into account travel time.

**Methods**:

1. Data Collection

Our data on specific crime attributes comes from the SANDAG web-accessible CSV which is updated weekly (as mentioned above).[2] No cleaning needs to be done on the already organized data. We can continuously keep our algorithms running on updated data by pulling the file from the web and adding it to our databases. We will make sure to not read old data and will only take the newly uploaded caes. This way, our database will continue to grow.

Using Google’s open API, we will extract predicted path duration and traffic along a street from Google Maps. This is reliable and commercially acceptable to use.

We will use crowdsourcing to collect additional data in the form of safety and speed ratings for certain paths, as well as local ratings of the safety of particular streets and intersections. This crowdsourced data is discussed in section IV. Crowdsourcing.

1. Traffic Data

Prior studies have shown that on average, fewer crimes occur on heavily populated streets due to a greater number of potential witnesses.[6] Using data from Google Maps, we will devise a scalar representation, *traffic*, on the interval (0, 1] to represent how much traffic a particular street has. We can consider our traffic data as an adequate representation of how populated a street is.

1. Statistical Weighting

To calculate the total amount of risk on a given path, we will consider *risk* as a combination of the risks of each block the path is made of. For each block, the individual risk is equal to a scalar representation of the *severity* of the crime on the interval [0, 1]—with more violent crimes weighted towards 1 and issues such as noise complaints weighted towards 0—divided by the *recency* of the crime where recency is how old the crime is in number of days prior to the current day. A resulting formula is as follows:

This way, more severe crimes and more recent crimes are weighted more heavily.

1. Crowdsourcing

At the end of each GoSafe journey, the user will be prompted to respond to two ratings. Both on a scale of 5 stars—1 being worst, 5 being best—the user will be asked how safe their walk felt, and how speedy their walk felt. These ratings are a form of crowdsourcing that we will later use in our path choosing algorithm in the form of *safety rating* and *speed rating*.

Additionally, we will collect data from locals who know the city innately and who may have more accurate information on the city. Crime data may miss out on key information, such as if a certain street is always avoided due to crime in the past, or perhaps there is no recent data even though the street may be unsafe. Locals in the area may have additional information about residents and future happenings in the part of town as well. Using Amazon Turk and Survey Monkey, we will ask locals how safe they would rate certain streets, intersections, and neighborhoods within a half-mile radius from where they live. Slight bias is inherent in crowdsourcing, but we can adjust the weight of these ratings, see V. Function Minimization, and hope that using a high sample size will result in fairly accurate results.

In order to promote accuracy, we will weight individual’s ratings on interval [0.5, 1] as a function of the distance between where they live and the street they are rating. The function we will use will remain at full weight for the first quarter-mile, and then deteriorate linearly to 0.5 by the edge of their half-mile radius.

Because these surveys are anonymous, we expect that the majority of people will be accurate with their responses and comfortable revealing the nearest intersection they live on.

1. Function Minimization

When choosing the optimal path, we have a few aims in mind and use Occam’s razor to create a function. We want to find a path that: minimizes risk, minimizes time, maximizes safety rating, maximizes speed rating, and maximizes traffic. A resulting function that can be applied to each potential path is:

We can then find the paths that minimize this output.

In our initial simulations, we will set all weights equal such that they sum to 1.0. If rounding beyond 2 decimal places needs to occur, then we will favor the weight for safety. In subsequent simulations, we will observe how the suggested paths change with different weights—see section VII. Visualization. We will decide which weights to use based on input from local advisors, and we can always adjust the weights further if deemed necessary.

1. Machine Learning

Some basic machine learning algorithms will be at play behind the scenes. We will have algorithms to analyze times and locations where and which crime is more likely to occur. Attributes fed into this algorithm would include the day of the week, the time of day, the category of day—holidays, sporting events, seasons—and key locations.

Closer analysis of the results of these algorithms may reveal important trends that can be factored into future path decision making. For example, there may be more crime on a street next to a baseball stadium in the hour after a baseball game ends. Our path choosing process will then modify the risk of specific streets based on these observed trends between time, date, and location.

1. Visualization

Visualization of the risk in different areas can be represented by a heat map overlay of the city map where areas with higher risk are displayed towards orange and areas with lower risk are displayed towards blue. Smoothing with a low radius of 1 block will be applied to this map in order to prevent harsh boundaries but to still allow for visualization of sharp changes in risk in close proximity.

Three suggested paths provided by GoSafe will be displayed as an overlay similar to other direction applications, same with the other navigational aspects of the application—current location, next set of directions, etc. The optimal path will be selected by default, with the other suggest paths displaying their relative safety and relative speed in percent compared to the optimal path.

Basic visualization—i.e. Plotting—will be carried out on potential trends spotted by the machine learning algorithm in order to provide logical confirmation of the association before it is ultimately factored into our path decisions.

**Discussion**:

To see if our app works, we can look at whether crime rates and locations have changed after the public begins using our app. We can look to see if crimes have moved to different areas and whether violent crime rates involving pedestrians on foot have decreased. We would want to see if the crimes have shifted because if they are not being committed in the same designated “dangerous areas” as before, but instead along our suggested routes, we would be endangering the users of our app. Also, a shift in the locations of violent crimes could be as a result of our app, either through our users committing crime or through crimes following the users because certain areas that might have been less populated before become more populated and vice versa. We also want violent crime rates for pedestrians to decrease because it would show that we are keeping more people safe by removing them from the areas where they are more likely to be victims of crimes.

A possible issue with crowdsourcing for ratings of efficiency and safety is that each person’s definition of safety and efficiency are different. Humans would be more accurate and capable at deciding safety than an algorithm, but someone who has lived a sheltered life might think walking through a park at night is dangerous while someone who works graveyard shifts might think it is safe. People may also be impatient or more sensitive to time, so someone in a rush to get to work might think that adding five minutes to their walk is not worth travelling a safer route, while someone else may prefer a safer route no matter how much time it adds to their path.

While the app may have a positive aspect of guiding users to their destination safely, there are many other implications that may arise. The ethics of conducting our experiment may be questionable as we have people travel through areas classified as dangerous to us. We are responsible for the safety of our users, so although our algorithm should find a safer route, any incident from taking a route our algorithm generated is deemed our fault. One indirect implication would be the impact on businesses in crime-filled areas. Since the app guides users to avoid such areas, these businesses will ultimately see less traffic. This is a huge factor to take into consideration as this app moves forward since the algorithm uses weights and priority analysis to choose routes, those of which are seemingly biased to safer travel areas. These areas would see an increase in foot traffic which may cause a huge economical shift as businesses and advertisers work towards moving their ventures to these locations. Neighborhood economies improve as community businesses thrive, so we could be impeding the growth of neighborhood economies if we decrease the traffic certain businesses receive. Another very dangerous implication would be the increase in vulnerability of the users. A person using the GoSafe app has a sense of safety and security as we provide the safest route, which means that they would not expect any crime to occur along their journey. This can be taken advantage of by criminals who misuse the app to find “routes that unsuspecting victims take”. This has been seen in apps that use geolocation such as Pokemon Go, where crime perpetrators were “able to anticipate the location and level of seclusion of unwitting victims.”[5]

When we ask locals to give their opinion about the streets there could be a possible bias because they are accustomed to the area so they would respond more positively to questions about safety.

The impact our app would have on society is that it would help decrease crime rates and ensure the safety of pedestrians. Additionally, our app can help tourists visiting San Diego and those directionally challenged, because oftentimes Google maps only calculates the shortest route in terms of traffic and distance without considering the safety of the paths they are leading pedestrians on or the areas they are taking them through. Tourists do not know the area they are visiting and often blindly follow the given route, assuming that it is safe. We would change that assumption into a near guarantee and thus encourage urban tourism in San Diego. Many people are afraid to travel on their own because of fears of safety and unknown areas. Having this app would allow them to explore the city safely and help encourage exploration of unknown places.

**Contribution**:

We all came up with the idea and the hypothesis. Jesse wrote the introduction and hypothesis. Parker wrote the methods with additional input from Roshan. Emmanuel and Isabel wrote the discussion. We all edited the paper after we finished writing it. Roshan edited and polished the final paper. We all contributed ideas to each other’s sections.

**Bibliography**:

1. [Laskow, Sarah. “When Big Data Maps Your Safest, Shortest Walk Home.” Next City, 2 Sept. 2014, nextcity.org/daily/entry/pedestrian-open-data-safest-shortest-walk-map.](https://nextcity.org/daily/entry/pedestrian-open-data-safest-shortest-walk-map)
2. [Scanlon, Pam. “Public Crime Data Extract.” SANDAG :: PROJECTS :: San Diego's Regional Planning Agency, SANDAG, Dec. 2017,](http://www.sandag.org/index.asp?classid=14&subclassid=21&projectid=446&fuseaction=projects.detail) [www.sandag.org/index.asp?classid](http://www.sandag.org/index.asp?classid)

[=14&subclassid=21&projectid=446&fuseaction=projects.detail.](http://www.sandag.org/index.asp?classid=14&subclassid=21&projectid=446&fuseaction=projects.detail)

1. Tashev, Ivan J. *Pedestrian Route Production*. 3 Jan. 2012.
2. Matyszczyk, Chris. “The Joy of Microsoft's 'Avoid Ghetto' GPS Patent.” *CNET*, CNET, 7 Jan. 2012, www.cnet.com/news/the-joy-of-microsofts-avoid-ghetto-gps-patent/.
3. Yuhas, Alan. "Pokémon Go: Armed Robbers Use Mobile Game to Lure Players into Trap." *The Guardian*. Guardian News and Media, 11 July 2016. https://www.theguardian.com/technology/2016/jul/10/pokemon-go-armed-robbers-dead-body.
4. McArdle, Megan. “Density and Crime.” *The Atlantic*, Atlantic Media Company, 16 May 2011, www.theatlantic.com/national/archive/2011/05/density-and-crime/238944/.