



Celebrity Location Identification for Paparazzi (C.L.I.P.)

W201 Group Project 1

2019 October 15th

Authors: John Lee // Alyssa Augsburger //

Derek Topper

Audience: CEO/COO of Paparazzi Agencies

Overview

Domain

- Entertainment Industry
- Celebrity Photography

Organization

- Paparazzi agencies:
MB Pictures, Bauer-Griffin, X17,
Splash News, TMZ

Stakeholders

- CEO/COO of company
- Photographers - Freelance,
Employees
- News Agencies / Consumers



Brief History

- “Gold Rush” Era vs. Great Recession and Rise of Online Media
- Freelance to Multinational Agencies
- Citizen Photography



Business Problem and Research Question

Core Problem: To Most Efficiently Utilize Paparazzi's Time to Capture Valuable Celebrity Photos

Time

- Idling time - waiting and traveling
- Number of other paparazzi or citizen photographers in same area
- **Where should the paparazzi be and at what time?**

Valuable Content

- High value photos depend on speed at which they can be distributed
- Celebrity popularity
- **How to include photo value into celebrity location models?**

Decision and Decision Maker

Decision: Should a paparazzi agency employ data science technology/methods in order to utilize paparazzi most efficiently?

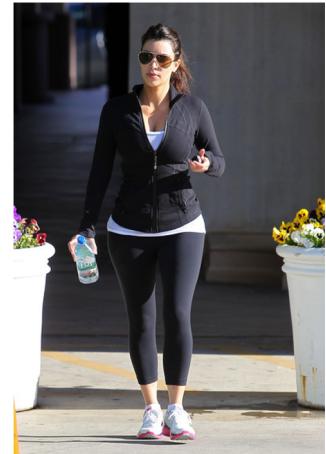
- Potential benefits of increased efficiency:
 - Reduced costs (fewer paparazzi, less idling time)
 - Increased revenue (producing more quality photos)

Decision Maker: CEO/COO of paparazzi agency



Quantitative Data Collection

- Past photo data
 - Existing first-party database
 - Web-scrape to find photos from other companies
- Large volume of location data
 - Instagram photos
 - Natural language processing of tweets and articles
 - Filming locations
 - Celebrity websites



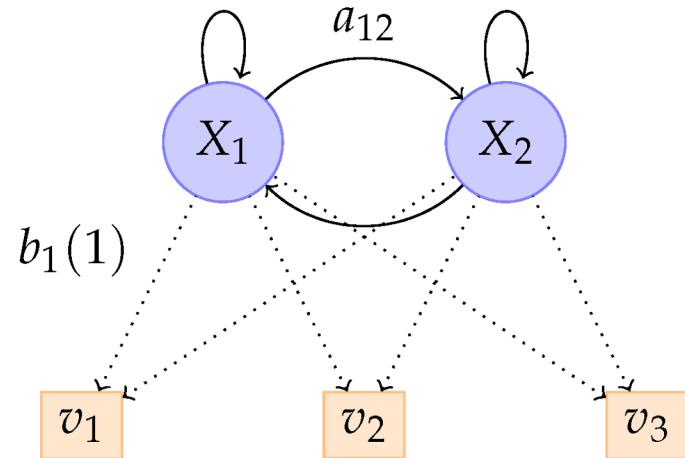
Qualitative Data Collection

- Interview paparazzi
 - What makes a profitable photo?
 - How to determine popular and unpopular locations?
 - How to determine popularity of celebrities?
 - What has enabled them to be successful in the past?
 - What has hindered their success in the past?



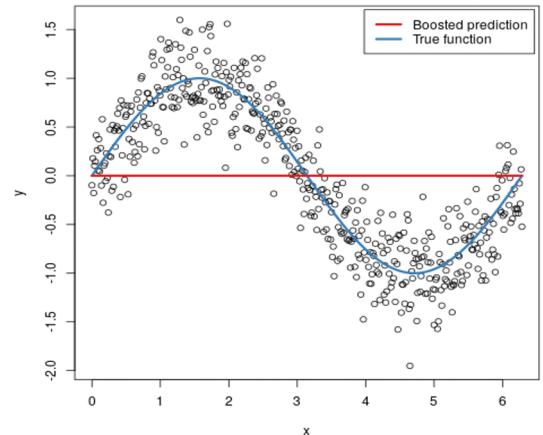
Modeling Locations

- Our plan is to use a multi-pronged approach to generating our models.
- Considering our data collection processes and past research, we plan on employing the following techniques:
 - K-means clustering algorithm
 - Group different types of locations based on their innate characteristics such as type of establishment, neighborhood, number of previous visits, etc.
 - Hidden Markov Model
 - Model of where celebrity is at a given time, as it allows us to project the spatio-temporal positions of celebrities.
 - Sub-Trajectory Synthesis
 - Converts historical location trends into “synthesized location trajectories” and will allow us to consider various patterns of movement as well as maintain the privacy of our data, in case of a company data leak, as it uses fewer location points than a traditional data model.



Modeling Where To Go Based On Celebrity Value

- We also plan to model the value of seeing a particular celebrity.
- To better place our photographers, we need to understand which celebrities should be targeted by our paparazzi.
- To tackle this, we'll combine the following methods:
 - K-means clustering of different types of celebrities.
 - Formalize celebrity groupings and develop clusters of celebrity quality for each celebrity, based on their popularity and fame.
 - Gradient Boosting Machine Learning Models
 - A statistical method that uses decision trees to provide a regression based approach to predicting a value, in our case, how much a photo of a celebrity is worth.
- Using the aforementioned clusters, and historical data on company photo sales, we can predict exactly how much a potential future photo will sell for and thus, let us know where to send photographers to maximize profits.



Decision

- Perform a 3 month trial
 - Deploy half of the paparazzi to locations based on what the model recommends
 - Deploy half of the paparazzi using the existing strategies
- Compare the number of photos taken and revenue generated per day between the 2 groups
- If the paparazzi are able to obtain more photos when deployed using the model's recommendations, then we would suggest continuing to use the model
- Using a trial and comparing results in order to make the decision, will help to avoid any bias in the decision making process

The Impact and Response of Data Science for Paparazzi

Organizational Impact:

- More accurately deploy paparazzi for higher probability of celebrity sighting
- Increase paparazzo efficiency - better utilize their time
- Better content
- Increase potential for more high quality photos - diversification

Organizational Response:

- Increase job satisfaction & job performance
- More customers & larger pool of consumers
- Mitigating risks
- Increase revenue and profits



Challenges to Overcome

Celebrity Behavior Correlated to Avoid Paparazzi

- Dynamic modeling
- Limit method publicity

Ecological Fallacy

- Avoid by using variety of data

Ethics

- Establish boundaries
- Study past failures (i.e. Gawker Map)
- Ensure privacy of celebrity location history through aforementioned Sub-Trajectory Synthesis modeling

What's Next?

- Data science to help paparazzi agencies
- Benefits to agencies, paparazzi, and consumers
- Data collection methods take advantage of available free data
- Data analysis methods are proven based on other applications
- Decisions and impact are outlined
- Challenges can be overcome

Will you help disrupt the entertainment industry?

Questions?

Citations

- Baram, Marcus. "How Celebrity Paparazzi Use Similar Risk Strategies as Wall Street Traders." *Fast Company*, Fast Company, 8 Apr. 2019, www.fastcompany.com/90331761/how-celebrity-paparazzi-use-similar-risk-strategies-as-wall-street-traders.
- Mcnamara, Kim. "The Paparazzi Industry and New Media: The Evolving Production and Consumption of Celebrity News and Gossip Websites." *International Journal of Cultural Studies*, vol. 14, no. 5, 2011, pp. 515–530., doi:10.1177/1367877910394567.
- Pant, Neelabh, and Ramez Elmasri. *Detecting Meaningful Places and Predicting Locations Using Varied K-Means and Hidden Markov Model*. 2017, doogkong.github.io/2017/papers/paper7.pdf.
- Valdes, Robert, and Cristen Conger. "How Paparazzi Work." *HowStuffWorks*, HowStuffWorks, 16 Sept. 2004, entertainment.howstuffworks.com/paparazzi3.htm.
- Xue, Andy Yuan, et al. "Destination Prediction by Sub-Trajectory Synthesis and Privacy Protection Against Such Prediction." Microsoft Research Asia, 2016.
- Jessica. "Introducing Gawker Stalker Maps." *Gawker*, 14 Mar. 2016, <https://gawker.com/160338/introducing-gawker-stalker-maps>.