The Power of Social Media in Supporting Warehouse Location Decisions for Online Retailers Using GIS

Emergent Research Forum papers

Husam Lahza
Claremont Graduate University
husam.lahza@cgu.edu

Yaser Alhasawi
Claremont Graduate University
yaser.alhasawi@cgu.edu

Andrew Marx
Claremont Graduate University
Andrew.Marx@cgu.edu

Abstract

While online shopping is one of the fastest growing sectors in the U.S. economy and is quickly surpassing traditional retailers (Enright 2014), shopper demand data used to place warehouses is either proprietary or expensive. To address this, we present an alternative approach to identifying where online shopping demand occurs in Los Angeles County and therefore where to most efficiently place warehouses for online retailers. Twitter data was harvested identifying the location of tweets about online retailers such as Amazon or eBay. This information was used as a proxy to model location of online shoppers. When compared with U.S. Census population data for ages 18 to 40, the Twitter-derived data was found to be a much more effective means to model the location of online shoppers and more efficiently place online warehouses of goods.

Keywords

GIS, social media, Twitter, warehouse, online retailers, online shopping.

Introduction

Online shopping is expected to reach $414 billion by 2018 (Enright 2014). While California is one of the best places for doing online business in the U.S., choosing a good site for a warehouse is both critical and expensive, preventing small online retailers from starting a business in the state. Warehouse location is determined by customer, competitor and manufacturing locations and is complicated by an ever-shifting and expanding population. Additional factors include transportation, rental and other logistic concerns. When placing an online business, the most efficient warehouse location brings tremendous gains to an online business. Geographic Information Systems (GIS) provide a means to address these questions offering complex procedures for analyzing and modeling location-explicit datasets.

This research demonstrates that Twitter data can be successfully used as a proxy to model locations of online shoppers which provide an inexpensive approach to help new online businesses decide on warehouse location. Using data from Los Angeles County, we demonstrate how Twitter users tweeting about their online shopping experiences offers a cost-effective approach to modeling the location of online shoppers. When compared with traditional demand-density datasets derived from U.S. Census data of 18 to 40 year olds living in Los Angeles County, the Twitter-derived data was found to offer a more effective and cost-efficient means to place online warehouses.

Literature Review

The literature has little contribution to e-retailer's warehouse location-allocation decisions. While researchers have attempted to study traditional geospatial techniques for warehousing in general, the connection between social media and warehouse location-allocation decisions are not well understood.

Vlachopoulou and Manthou (2001) reveal how GIS can support the decision for the warehouse site selection. They use quantitative and qualitative analysis to classify alternative warehouses, and they integrated GIS and Decision Support System (DSS) for the evaluation of a warehouse location. Eldin and Eldrandaly (2003) propose a component that helps in designing a DSS for industrial site selection. The tools used in this paper's DSS were Expert Systems (ES), GIS, Multicriteria decision-making (MCDM), and analytic hierarchy process (AHP).

In social media researchers are making tremendous progress in tracking and locating specific events based on users' tweets. Guille and Favre (2015) developed a method that relies on tweets and uses the creation frequency of dynamic links that users insert in tweets to detect significant events and estimate the degree of their impact over the crowd. Lampos, Bie, and Cristianini (2010) harnessed Twitter by using it with a machine-learning algorithm to track the prevalence of flu in several regions of the United Kingdom.

None of these studies however combined the use of Twitter with GIS to support warehouse allocation decision-making. This study makes the unique contribution to the evolving body of MIS research by utilizing the data generated from a social media platform (Twitter) for the purpose of selecting warehouse locations using GIS and spatial analysis techniques. The novelty of our contribution resides in combining the analysis of social media data with modern GIS practices to expand research in warehouse location-allocation decisions.

Research Objective

This research creates a framework for choosing online shopping warehouses based on online shoppers' density. The research benefits MIS researchers interested in GIS and social media along with practitioners and online retailers who are interested in finding more efficient, cost-effective techniques for warehouse location-allocation. We believe that finding the optimum warehouse location has various benefits including low-cost shipping, fast customer service and higher level of customer satisfaction. Our primary objective is to answer if GIS techniques incorporating social media data improve e-retailer's warehouse location-allocation decisions.

Spatial Datasets

There are few data sources available which represent online shoppers. Most online retailers, such as Amazon and eBay, consider transaction data proprietary, so it is difficult to attain live, accurate information. In recent years, Twitter has proved to be a promising source of information by providing a stream of live and accurate data about individuals' daily life. This data has helped researchers and practitioners to track different types of events (Guille and Favre 2015, Lampos, Bie, and Cristianini, 2010). Our research uses data streamed from Twitter to extract geospatial information about potential online consumers. This is projected using modern GIS tools to aid warehouse site location-allocation decisions. For the purpose of this study, a proxy has been proposed, in which data was collected from Twitter to represent the density of online shoppers. This data was developed using a program code written in Python. The resulting data was a stream of tweets from Nov 9th to Nov 21st 2015. This only included tweets that contained specific keywords related to online shopping and tweets that have geographical coordinates. Tweets from 1,023 unique accounts met these criteria for the study period within Los Angeles County.

Ancillary data included census tract cartographic boundary shapefiles of the Los Angeles County, summarized from the U.S. Census Bureau's MAF/TIGER geographic.

Research Site

Los Angeles County was chosen because it is the most populated county in the U.S. whose 2010 population is just under 10 million residents. Additionally, the large geographical area of LA County, 4,058 square miles of land, makes good study area for a variety of potential warehouse locations. Most importantly, according to some business reports, Los Angeles city has the biggest twitter population in the U.S. following New York City, and it is considered the fastest growing city in social media.

Research Design

The study was designed to help in decide the most suitable warehouse locations for online stores using spatial analysis tools provided by ArcGIS. Since, not enough datasets were available for scholars about online shopping demographics; tweets from twitter has been used as a proxy to address the potential density of online shoppers in the LA County. Twitter is one of the most popular social networking websites, based on the concept of micro-blogging. Micro-blogging is where users share information about their daily activities in the digital world. Scholars have shown that twitter can be used as a data source to track specific events in timely manner (Guille and Favre, 2015, Lampos, Bie, and Cristianini, 2010).

This study was conducted in five consecutive phases. The first phase was data collection in which tweets were streamed using a program code written in Python. The code tracked specific tweets in LA County that contained keywords related to online shopping such as Amazon, AmazonPrime, Groupon, eBay, AmazonLocal and CyperMonday. For the study period of 9th to 21 November 21st 2015, tweets from 1,023 accounts were captured which met our criteria. Tweet fields captured included the text from each tweet, the longitude and the latitude.

Amazon	EBay	Alibaba	Zulily	Zappos	AmazonLocal
Walmart.com	CyberMonday	beezid	boohoo	StyleBob	BonaDrag
UBid	HappyBidDay	Craigslist	DealDash	Groupon	SiteToStore
Amazon	EBay	Alibaba	Zulily	Zappos	AmazonLocal

Table 1. Keywords used in the search criteria for Twitter streaming.

The second phase was creating a geo-database linked to the ArcGIS spatial database engine. Twitter data was imported into the database and was projected onto LA County census tract map as points (Fig. 1).

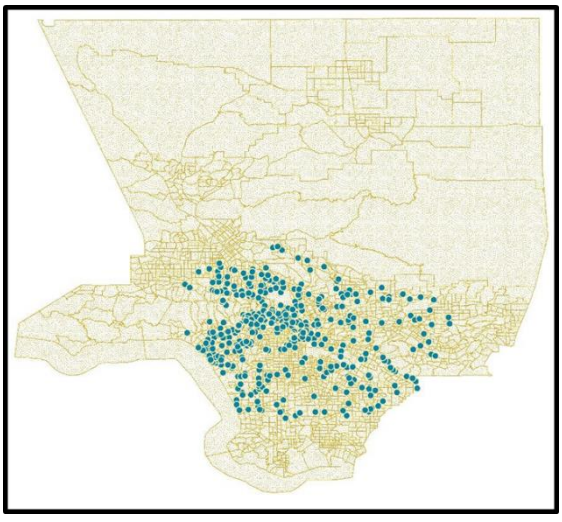

Figure 1: Projected Twitter point data on 2010 LA County Census Tract Map.

The third phase was the first step in the spatial analysis where a density surface was used to take the projected points from twitter data and create a layer consisting of continuous values that represents the density in a raster format. The tool used to create the density map was Point Density, from the Spatial Analysis toolbox in ArcGIS 10.1, which suggests output cell size, "1.68346240000005E-03", and radius, "0.014029". The tool was used without any modification (Fig. 2)

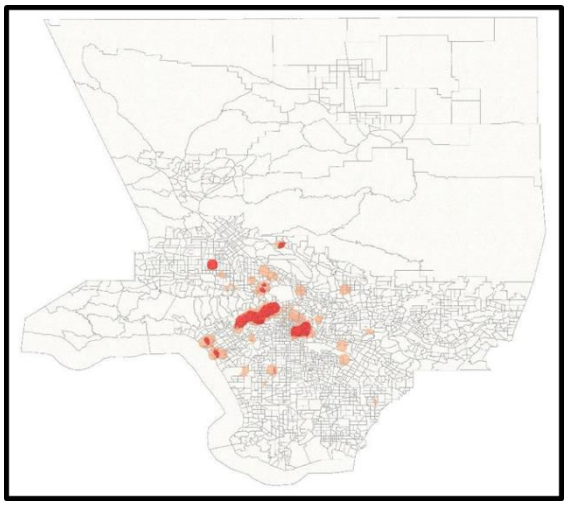

Figure 2: Twitter population density on 2010 LA County Census Tract map.

The fourth phase found the optimum location for the warehouse based on the density of the tweets. Figure1 shows the scattered tweets as blue dots all over LA County. Figure 2 however reveals a highly dense area in the middle of the map, specifically the area around Beverly Hills and Downtown LA. This gives a clear idea of the potential warehouse location. In order to select the site empirically, the "Mean Center Tool" which is part of the Spatial Statistics toolbox in ArcGIS 10.1 is used. The Mean Center tool calculates the center of distribution of the features, in this case tweets, to locate the best site to service all online shoppers (Fig. 3). Nonetheless, for the intent of comparison, two more spots have been placed manually in the map (locations A and B) due to their proximity to the dense area (Fig. 4).

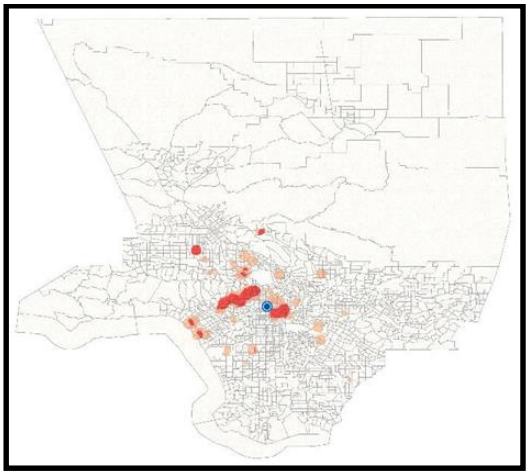

Figure 3: Warehouse location using Mean Center tool

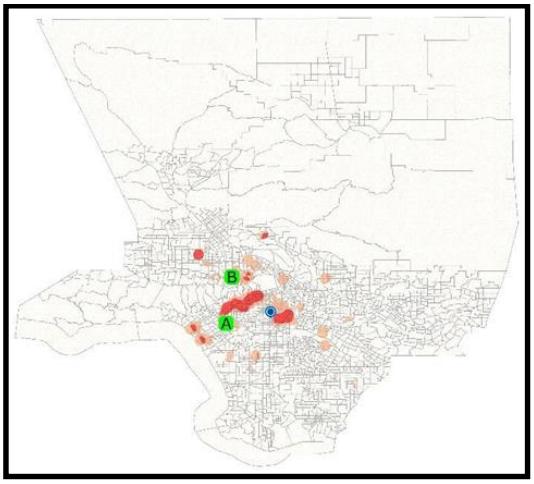

Figure 4: Two additional Warehouse locations picked manually

Although the warehouse locations (Fig. 4) appear relatively close to each other, each is able to provide quality service and fast delivery to a number of customers within a specific zone. Therefore, for phase five, the Multiple Ring Buffer tool from the analysis toolbox ArcGIS 10.1 was used to find features, in this case tweets, within a specified distance inside the buffer rings. A buffer with three rings was created for each of the warehouse sites (Fig. 4). The three rings of 2 miles, 4 miles and 6 miles were chosen to represent speed of delivery. After creating the buffers (Fig. 5), the number of tweets within each ring is calculated. Table 2 shows that the location selected by the mean center could serve total of 464 customers within 6 Miles while the ones that were manually selected (A and B) could serve only 251 and 284 customers respectively. Therefore, the warehouse location selected by the Mean Center on twitter data was more effective than the manually selected points.

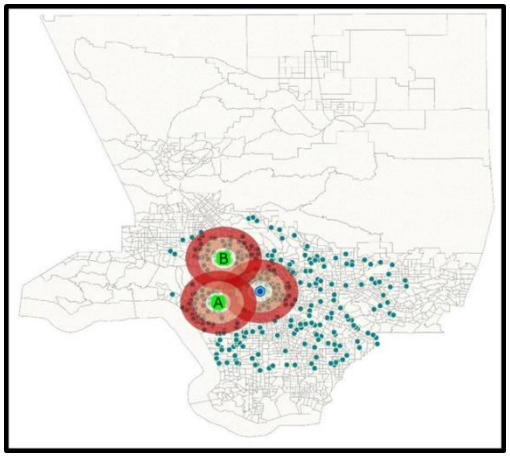

Figure 5: Two additional Warehouse locations picked manually

	2 Miles	4 Miles	6 Miles	Total	%
Mean Center Warehouse	42	315	107	464	45.4%
Manually Selected Site A	30	103	118	251	24.5%
Manually Selected Site B	22	63	199	284	27.8%

Table 2: Keywords used in the search criteria for Twitter streaming.

Results and Discussion

The majority of online shoppers were located in the center of LA County; specifically, the area of Beverly Hills and downtown Los Angeles. The result thus raises the question of why majority of the tweets were generated in the center of the city. It is likely that this concentration represents either heavy twitter traffic, a concentration of online shoppers or perhaps both. Future planned research includes determining why the tweet data was distributed in this manner by examining additional factors including demographics, median income, land value and overall twitter.

Conclusion

This research presents a GIS approach, leveraging Twitter density data, for the identification and selection of the best warehouse location for online retailers. Data collected from Twitter was efficient and useful to determine the location of the storage, and GIS provided an excellent platform to exploit this data. Small online retailers may consider using this approach as a cost-efficient method to place their warehouses, further saving them time and money in the form of distribution costs.

REFERENCES

Eldin, N., Eldrandaly, K., and Sui, D. 2003. "A COM-based Spatial Decision Support System for IndustrialSite Selection," *Journal of Geographic Information and Decision Analysis*, pp. 72–92.

Erickson, W. P., Mcdonald, T. L., and Skinner, R. 1998. "Habitat Selection Using GIS Data: A Case Study," *Journal of Agricultural, Biological, and Environmental Statistics* (3:3), p. 296.

Guille, A., and Favre, C. 2015. "Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach," *Soc. Netw. Anal. Min. Social Network Analysis and Mining* (5:1), pp. 1–18.

Lampos, V., Bie, T. D., and Cristianini, N. 2010. "Flu Detector - Tracking Epidemics on Twitter," *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science*, pp. 599–602.

Vlachopoulou, M., Silleos, G., and Manthou, V. 2001. "Geographic information systems in warehouse site selection decisions," *International Journal of Production Economics* (71:1-3), pp. 205–212.