Estimation of People Density to Reduce Coronavirus Propagation

Mouad Tantaoui, My Driss Laanaoui, and Mustapha Kabil

Abstract  Today, the spread of coronavirus has become the number one concern of countries as it threatens human life and economics; therefore, the scientific community tries hard to discover the treatment to deal with this virus or at least find out a method to reduce its propagation. In this context, our concern is to estimate density of people being inside all different places of interest in the city in the purpose of distributing users of our application to different places by avoiding congestions of people. The use of big data is very important for data treatment for fast execution. In this time, oldest relational database technologies cannot anymore handle the enormous data created by various application sources, in this direction big data tools permit us to handle this enormous data in the purpose to mine important data from the voluminous data, if we rely on the support of big data tools, the treatment will be complicated to administer. In this chapter, firstly, we define a system that calculates number of people in diverse city’s areas; it will help to inform users’ best places to alleviate areas where there are congestions and redirect people to other place with low density. And secondly, we keep trace in a database all people contacts that have been near each other to prevent people who have been close to the positive cases.

Keywords  Big data · Density management · Congestions handling · Spread of coronavirus

M. Tantaoui (✉) · M. Kabil
Hassan II University, Casablanca, Morocco
e-mail: tantaoui.mouad@gmail.com

M. Kabil
e-mail: kabilfstm@gmail.com

M. D. Laanaoui
Cadi Ayyad University, Marrakesh, Morocco
e-mail: d.laanaoui@uca.ma

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021
A. Joshi et al. (eds.), Sustainable Intelligent Systems, Advances in Sustainability Science and Technology, https://doi.org/10.1007/978-981-33-4901-8_7
1 Introduction

Today, combating covid-19 seems to be the first priority of scientific researchers, so here, we intend to develop management of people density all around the city; this domain of research catches the attention of researchers of different fields linked to either social and economic problems or citizen security. In another shutter, many generators of data generate enormous data characterized by its various genders and velocity that makes the process very complicated; big data technologies are a good alternative to process this voluminous data in real time. In our mechanism of computing people density, we rely on the help of big data tools advantages to process the huge data. Places or services desired in the whole city contain too many people with big waiting lines. The strong growth in waiting lines has engendered many environmental and social issues in towns mainly at present of the spread of covid-19, so fighting this viral disease is introducing the interest research field that concerns the researchers in diverse domains as a result of the problems found. In this chapter, we intend to discover a way for calculating number of users on different places through city, and a way of combatting covid-19 propagation; this permits us thereafter to engender a type of equilibrium on all services where there is a high congestion. To achieve this goal, we have defined a method of estimating density utilizing huge data gathered. In other words, our method intends to equilibrate people density in all locations, and it is a form of redirecting people to different places so that we obtain a medium density; this will reduce the risk of coronavirus propagation. After this introduction, the rest of our chapter is presented as following. Section 2 presents related works. Section 3 proposes the detailed proposed methodology with the experiment results. Finally, Sect. 4 concludes our chapter.

2 Related Works

Lambda architecture is an architecture that allows to process the huge data, and we are inspired on this architecture [1] characteristics for constructing our system; it consists of data processing that permits to process voluminous data. The architecture is composed of different elements: serving, batch, and speed layer.

In paper [2], authors review the evolution history of data controlling systems. After that, they point the state of the art of big data controlling mechanism concerning storage, modeling and request engines of enormous data managing systems. In addition to that, authors analyze the data and characteristics of the data mass controlling systems, big data controlling systems need to be more developed and analyzed because it progresses with a fast manner.

Demography is linked to statistic domain in citizen science. The rise of population is engendering many problems to monitor the huge mass of demographic data (with other words ‘big data’). The help of science development progression and informatics progression, much architecture has been produced to control the mass
of data. Almost architectures rely on old relational database, yet it cannot monitor big data management effectively. In paper [3], Bukhari et al. construct a system based on big data controlling system with the support of apache Hadoop platform to solve issues related to demography great increase data processing. The architecture is composed of Sqoop, HBase, and Hive. The system permits to export data from all RDBMS to apache Hadoop HBase. In addition to that, utilizing HBase and Hive seem to be a better alternative to store and request data to process accurately. The experience proves that the architecture monitors very well the mass data for future development. The system constructed will advance one step forward in monitoring complexes issues of big data and will reinforce processing in various domains such as economic or social domain. Big data is a domain which is becoming linked to our habits recently and engendering several data and solutions to develop and devoting on research field and new technologies of taking good solution in various fields like business and engineering. Data is no more such as supplementary element; big data gives the possibility to be an important piece of puzzle for taking good decision. In paper [4], authors utilized Guizhou as a solution to engender big data processing system depending on time and space relying GIS bussing technique, and after, authors discuss the results, and they concluded to construct technical assistance for the transformation of tourism by the specialists of Guizhou and efficient indicators.

Big data offers the opportunity of monitoring the huge datasets which were very complicated to process before. Actually, traditional relational database does not have in its power to administer the huge dataset which permits distributed database NoSQL to progress through the time. In paper [5], authors engendered and constructed a new distributed big data management system (DBDMS), which run under apache Hadoop technology and NoSQL mechanisms; it allows collecting the voluminous data, searching and infinite storage and showed by the results; DBDMS improves the treatment quality of huge dataset, and it is good for huge log backup and recovery, voluminous network packet capture and process, and other many different domains.

In [6], Matallah et al. establish a mechanism to well manage metadata of Hadoop technology to improve consistency and ameliorating performance and scalability of metadata by proposing a hybrid alternative on centering and apportionment of metadata to improve the quality and development of the mechanism.

Today, information technologies engender huge quantity of data every hour and every minute from various sources. Those quantities of data seem to be very complicated for the processing capability of traditional data treatment methods to monitor and manage in such a required limited time. This voluminous amount of dataset characterizes the big data. It is facing many problems in various processing on data such as capturing, treatment, searching, and filtering. Hadoop technology offers adaptive tools for various companies for big data monitoring mechanism. Hadoop concentrates on the concretization of various industry methods. Therefore, for handling Hadoop technology, it is obligatory to overcome how to master the operation of Hadoop ecosystem and Hadoop architecture. In this direction, the paper [6], Raj et al. discussed the Hadoop architecture and Hadoop ecosystem. All these methods lack to fast real-time data processing, because today, we need solution that handles
and process the data immediately, and also we must adapt the solution to different situations.

In this age of high technology, innovation, production creativity and arranging have been mainly transformed toward open collaboration ecosystems. People organize and with an open manner their works concerning data treatment and collaborate on exploring new solutions or dataset transmission for lot of systems. Big data is an emerging tool to monitor the complexes systems of open collaboration, and community of research needs to work with open manner with collaboration researchers with new data that expose different aspects, and new numerical models and techniques. So in paper [7], authors focused on problems that is based on numerical collaboration and constructed the dataset analytical issues which have to be showed to remedy the essential research challenges.

In paper [8], a survey is elaborated to analyze the propagation of the coronavirus. They specify how the virus is spreading in cities; also authors clarify vision of the main indicators which affect the propagation of the virus, where most of indicators attached to the infection spread are taken into consideration. It is proved that the difference in terms of distance to epicentrum is greatly affecting indicator that impact the spread of covid-19. The propagation of the virus needs to be more investigated if we want to combat it because it still looks blurry on the way of its spread. In paper [9], several appearances of the covid-19 infection are showed; authors show a global vision of the spread of this virus, and exhibit some tools of data analytics on the virus. Firstly, they point a literature study on covid-19 specifying lot of factors like the origin, resemblance with old coronaviruses, the transfer power, symptoms, and more indicators. In the second part, they employed dataset diagnosis tools on a dataset of Hopkins University to know very well how this harmful virus spread.

In paper [11], authors proposed a case study of utilizing composite Monte Carlo which is supported by machine learning mechanism and fuzzy rule induction to obtain good stochastic vision concerning coronavirus progression. In paper [12], authors show some examples of technologies which are supported by artificial intelligence (AI) that enhances autonomous process, better technology, and good making decision, in the purpose of combatting the covid-19 and saving lives. These AI technologies are good but they need to be more developed because this virus is new that is why we need some novel tools to combat it.

3 Methodology and Experiment

In this chapter, firstly, we spot places where grouping of people are dense such as places of interest, we took different banks, all supermarkets, cashplus, and sales services; different places of interest should have equipment such as cellular detection sensor to log in a database all cellphones and compute number of phones in the location, and then we deduct number of people on all places within the city (sensor of cellphone detection is located at the entrance of the area to log the identifier number of the all entrances and exits of phones). When someone is entering to a particular
location, he will connect to our system and he can select kind of area where he
desire to attain, so the system will pick the place where the density is low; like this
way, the system equilibrates load dispensation of all areas of interest in the sense
of number of people, and so, reducing the covid-19 chance of propagation. Firstly,
the methodology tends to build the database: every place of interest log incoming
and outgoing of smart phones in the sensor of the area, and register them on a its
own database for two purposes: we calculate smart phones number lying the place
indoor, and in second hand, the system record the trace of cellphones which have
been near to each other. This approach is described in Fig. 1, then the place system
disseminates every 5 min the data and the number id to the application to append
the data to the central base of all sites of services within the city, and the database
is refreshed after each 5 min; therefore, the central system possess a total control on
places densities located indoor in real time and cellphones which are near to each
other. So, when an user search a wanted destination, the application selects the best
place of service described by a lower number of people located there based on the
base constructed previously, and therefore, transmits to the consumer the optimal

![Fig. 1](image-url)  Different steps of building the database having status of all area of interest
itinerary to arrive safely its desired place. This approach is described in Fig. 2. The usefulness of establishing a database of all cellphones that have been near to each other is very useful, when someone catches a positive test for covid-19; after that, the application warns users who were near to the positive test diagnosis. The suggested system is summarized in Fig. 3. Our application allows picking the best place of service of a person and choosing an itinerary to arrive the best destination by picking the area of service that contains low number of people, then warning users who were near to positive covid-19 test relying on utilizing big data tools to have a fast processing in real time. For the experiment, we catch Casablanca City card, Morocco containing various areas; we took one thousand stores, one thousand banks branches, and one thousands of each different supermarket, wafacash, and teshihat services. And we simulated the experience in our computer characterized by: Processor: Intel Core i7. RAM: 16 GB of DDR4 2400 Hz RAM, and Nvidia GeForce GTX 1070 8 GB in Graphics card.
Fig. 3 Different steps of the methodology
For the experiment below, we selected some users who picked the same bank as destination; after that, the application will distribute them to banks having lower number of people; Table 1 exhibits the assignment of peoples to the bank assigned by the application relied on the database of places of number of people. Table 2 points the number of people degree of various banks and their degree after 15 min. At 12:00, bank of id eleven is characterized by a high density, so incoming people will be sent by the application to the bank of identifier twelve so the bank eleven is alleviated after some time, so the application can redirect next.

person to it. Table 3 indicates users who were diagnosed positive for covid-19 and users who were near to them. User of identifier one was near to person of id 7, 10, 14, 31, and then these users are warned to make their diagnostic of the infection; in case of these persons are diagnosed positive, we look for the entry on the database that have identifier $x$ of the infected user to warn users who were near to user $x$. Sending users to various banks with a manner to balance the number of people in different locations will permit to lighten people density who were near to infected ones, and warning users earlier will alleviate the spread of the infection and anticipate the treatment earlier to recover rapidly.

On account of the recurrent modification in the locations state within the city, there is lot of modifications on the database instantly. The suggested application offers a tool to control congestions and to limit the probability to have a high number of people on a such area of service, and it permits to spot positive tests betimes by diagnosing users who were near to positive person; it also permits to have as a gender of equilibrium of loads in the sense of our application fill up areas by users, and when it starts to be full of people, it redirects the coming peoples to a location of the same kind which is less filled than the others, and the application promotes this impact of equilibrium of load in an automatic way. Thanks to this application, we catch nearly all potential connection between users, and then we handle and limit the propagation of covid-19. These several modifications within the database allow our system more precise about the considering density of people on various locations and to distribute

| Timestamp | User identifier | Bank identifier |
|-----------|-----------------|----------------|
| 12:00     | 11,12,13        | 12             |
| 12:10     | 14,15           | 13             |
| 13:00     | $n, n + 1$      | $k$            |

| Timestamp | Bank identifier | Density ($t$) | Density ($t + 15$ min) |
|-----------|-----------------|---------------|------------------------|
| 12:00     | 11              | High          | Medium                 |
| 12:00     | 12              | Low           | Medium                 |
| 12:10     | 13              | Low           | Medium                 |
| 13:00     | $K$             | Low           | Low                    |
them to different place of services. The architecture is organized of central batch dataset storage and treatment techniques and a dispatching data storage technique for instantaneous processing to handle and treat this huge dataset.

The main tools utilized on our experiment, we employed apache storm to process data in real time and dispatching and rapid flow processing. And for handling the huge dataset, we employed Hadoop MapReduce which builds the database that is used by speed layer characterized by a fast treatment layer instantaneously. The use of storm is very important; it allowed us to process the data in real time, and it is the key to succeed and achieve the objectives of our method. We concluded according to the experimental result that the suggested application model is a good alternative that relies on big data tools, allowing disseminating data in real-time processing for smart affecting control of users to handle the number of people in various locations through city.

4 Conclusion

Places of interest engender huge dataset that is very difficult to monitor, and it implies that employing big data tools is compulsory to mine interesting data from the huge data. The propagation of covid-19 reveals a peril to the safety of our life and to find an alternative to the problem of controlling the danger of high people density; this chapter develops a techniques of big data to ameliorate redirecting handling, and we managed to build an instantaneous number of people detection system with parallel data treatment, which permit to process rapidly in terms of time. The application allows logging precisely the people density being in each place of service, and it permits to overcome the density and conduct people to the safest place desired to attain and the establish the route to reach their destination, which lead to restrict congestion in several places and avoid the danger of covid-19 spread. The results of our simulation demonstrate that the application limit congestion accurately and so control covid-19 spread and not only this; but, the results show good latency and better precision. Utilizing the database of people number and merging this database with machine learning mechanism would be our next research focus to overcome and limiting people number of all locations and showing good results. The results of our method are good, but it need more better accuracy; we are working on the next paper on a method that will ameliorate this methodology by adding more factors in the purpose of improving latency and having a good insight over density in city to handle the distribution of people in different places.
References

1. M. Nathan, J. Warren, Big Data: Principles and Best Practices of Scalable Real-Time Data Systems (Manning Publications Co., New York, 2015)
2. X.L. Wei, Z. Feng, History, current status and future of big data management systems. J. Softw. 30(1), 127–141 (2019). https://doi.org/10.13328/j.cnki.jos.005644
3. S.S. Bukhari, J. Park, D.R. Shin, Hadoop based Demography Big Data Management System. in Presented at 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (2018). https://doi.org/10.1109/snpd.2018.8441032
4. S. Lin, B. Luo, J. Luo, Z. Wang, Y. Wei, Guizhou Big Data Management System and Key Technology. in 2018 26th International Conference on Geoinformatics (IEEE, 2018), pp. 1–7. https://doi.org/10.1109/GEOINFORMATICS.2018.8557109
5. H.Y. Chen, Design and Realization of a Distributed Big Data Management System. in Advanced Materials Research, vol. 1030 (Trans Tech Publications Ltd), pp. 1900–1904. https://doi.org/10.4028/www.scientific.net/amr.1030-1032.1900
6. H. Matallah, G. Belalem, K. Bouamrane, Towards a new model of storage and access to data in big data and cloud computing. Int. J. Ambient Comput. Intel. (IJACI) 8(4), 31–44 (2017)
7. A. Raj, R. D’Souza, A Review on Hadoop Eco System for Big Data. (2019). https://doi.org/10.32628/CSEIT195172
8. S. Brunswicker, E. Bertino, S. Matei, Big data for open digital innovation—a research roadmap. Big Data Res. 2(2), 53–58 (2015). https://doi.org/10.1016/j.bdr.2015.01.008
9. L. Liu, Emerging Study on the Transmission of the Novel Coronavirus (COVID-19) from Urban Perspective: Evidence from China. Cities, 102759
10. M.R.H. Mondal, S. Bharati, P. Podder, P. Podder, Data analytics for novel coronavirus disease. Inf. Med. Unlocked 100374 (2020)
11. S.J. Fong, G. Li, N. Dey, R.G. Crespo, E. Herrera-Viedma, Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl. Soft Comput. 1, 106–282 (2020). https://doi.org/10.1016/j.asoc.2020.106282
12. S.J. Fong, N. Dey, J. Chaki, AI-Enabled Technologies that Fight the Coronavirus Outbreak. in Artificial Intelligence for Coronavirus Outbreak (Springer, Singapore, 2020), pp. 23–45