



The Algorithm of the Snail: An Example to Grasp the Window of Opportunity to Boost Big Data

Jean Louis Monino^a and Soraya Sedkaoui^b

^aDepartment of Economics, TRIS Laboratory, University of Montpellier France

^bDepartment of Economics, University of Khemis Miliana Algeria/ TRIS Laboratory

monino@wanadoo.fr, soraya.sedkaoui@gmail.com

KEYWORD

Algorithm of the snail; Localization; Big data; Real time analysis

ABSTRACT

This work explores a new application which can effectively meet different localization accuracy requirements of most data location services studying the interactions between customers and suppliers. It helps to have the status or position of what is sought with respect to an address that summarizes thus a reference point which is the point of research. This proposal explains what snail algorithm is and how we can benefit from using it for the localization of information for business applications especially in the field of analytics. A business application using our algorithm has been developed by the Autour.com company (located in the department of Hérault, Montpellier city) to illustrate its feasibility and availability. The results show that our algorithm can improve the localization accuracy.

1. Introduction

The age of big data is now coming. But the traditional data analytics may not be able to handle such large quantities of data. Therefore, there is a need to be able to recognize the appropriate analytics technique to use for the data and business problem. However, these companies are interested in applications that enable a range of data (coordinates) relative to a reference point, for example, the situation of a manufacturer to offer a good or a service in relation to applicant of this type of good or service. The challenge identified is to organize the flow of data from their sources, process them and make them available to different users, but as applicant information is needed still more. To locate the closest ads you need a fast and efficient way which allows for the status or position of what is sought with respect to an address that sums up to a reference point which is the point of research. So it is more than necessary to plan and implement modern applications with sufficient details and observations covering the points between.

This work explores a new application which can effectively meet different localization accuracy requirements of most data location services studying the interactions between customers and suppliers. It helps to have the status or position of what is sought with respect to an address that summarizes thus a reference point which is the point of research. This proposal explains what snail algorithm is and how we can benefit from using it for the localization of information for business applications especially in the



field of analytics. To show the importance of this algorithm we will use the department of Hérault (Montpellier-France), although specifically, through the company "Autour.com" that handles every day huge files of its customers. The application of our algorithm by this company shows that people living next door (applicants and suppliers) can be put into connection; is a kind of creation of a district social networking. The remainder of this paper is organized as follows. The authors in section 2 discuss an integrative literature review on the localization algorithms. Our algorithm is explained in section 3. Finally, section 4 concludes this paper.

2. Literature Review

For several years, localization issues become more and more a challenging subject to our dynamic era. For this, a wide range of studies was performed to characterize the performance of these systems in different environments. Going from presentation of research capabilities of an interest point to the positioning of the user and its true location on the map. That presents an overview of tools and documentation to better understand the city either in whole, or in its environment through its districts.

Before starting the description of this algorithm, it's necessary to identify the context in which the idea arose. Therefore may be mentioned three essential issues (points)

- The first is the arrival of so-called Petabyte age or "Big data" characterized by the 3Vs: Volume, Variety and Velocity and its challenge that lies in the way of how analysis and combine this amount of data to extract value. The area of big data has compelled the community to pay serious attention to development of algorithms and applications for intelligent data analysis.
- The second issue is related to Fibonacci numbers. Fibonacci applied his sequence to a problem involving the breeding of rabbits, and he mapped out the family tree of a group of rabbits that initially started with only two members.

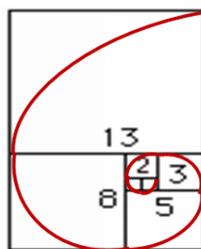


Figure 1: Fibonacci spiral.

For us the most important of these, is the Fibonacci spiral which drawn starting the corner of the first rectangle, all the way to the corner of the rectangle of the side.

- The last point on which the idea of the algorithm is based can be bounded to the shell of the snail. So, it's not about its speed but rather the form, because of this it's called "algorithm of the snail".

The benefits of grid mapping are well known as the readability of the gridded map enables detailed analysis of intra-urban dynamics. If these benefits are common to all types of grid, two families can be distinguished: the grid resulting from a breakdown of data collected in any zoning (zonal data transfer) and the one resulting from the automatic addressing (transfer point data). The interpretation of aerial pho-

tographs from 1986, based on a grid of 125 m side, allowed Dayre and Mazurek (1988) to analyze the land use on the urban district Montpellier. This method offers many possibilities for automated processing of data, especially regarding the dynamics of built areas.

As part of a mapping grid, Lajoie and Langlois (1998) studied the disintegration of zonal variables, returning on the assumption of the spatial distribution of populations. The aim is to approach differentiated rules disaggregation variables by merging morphological nature of information plans, topological and environmental. Modeling of Antoni (2001) appears as an interesting tool to better understand urban sprawl and to simulate various useful development scenarios for development. The methodological approach presented by the author combines three steps. Each step corresponds to a model:

- The first quantifies the spreading,
- The second locates it
- The third differentiates it.

The three stages are associated with a spatiotemporal database that uses the grid to store all the necessary information in a single GIS layer.

In recent years, many researchers have proposed solutions to the positioning system implementation based on different technologies, such as systems based on GPS (Zhou, and Shi, 2008), the ultrasonic sensor (Addlesee et al, 2001), video cameras (Hoey et al,2006) and the systems based on radio frequency identification. These different approaches and technologies offer different ways to address the problem of locating and monitoring in real time. Localization is a broad and active research area, and a diverse set of solutions have been proposed. Significant research has been done using location data of mobile users. Some were fundamental research such as Song et al.'s (2010) work on identifying patterns of human mobility. Some focused on building new services that may have great public or business potentials (Zhang et al, 2010), such as modeling city living neighborhood (Cranshaw et al, 2012) and recommending friends and locations (Zheng et al, 2011).

Our work also benefited from excellent works analyzing and making use of INSEE on the signs of the Toulouse Diversity (2008), but this study do not focus on the usual zoning (town, district) but on a continuous territorial coverage formed tiles. For us the interesting applications are those that contain devices allowing the user to enjoy all the benefits of positioning and location services. Recently, the realization of a location model involves:

- The availability of an infrastructure containing all the data needed to acquire the necessary information.
- Determining a reference against the position to locate.
- The treatment of acquired data and extracting the necessary information to determine position.

Our algorithm is inspired by related work in geo-localization and big data analytics. It offers solutions that can be deployed to large population and used for mobile proximity marketing or social networking services or why not in the field of health. The details of this method are discussed in the previous sections.

3. Presentation of "the snail's algorithm"

3.1. Description

In this section we will present our localization algorithm, which consists of three phases:

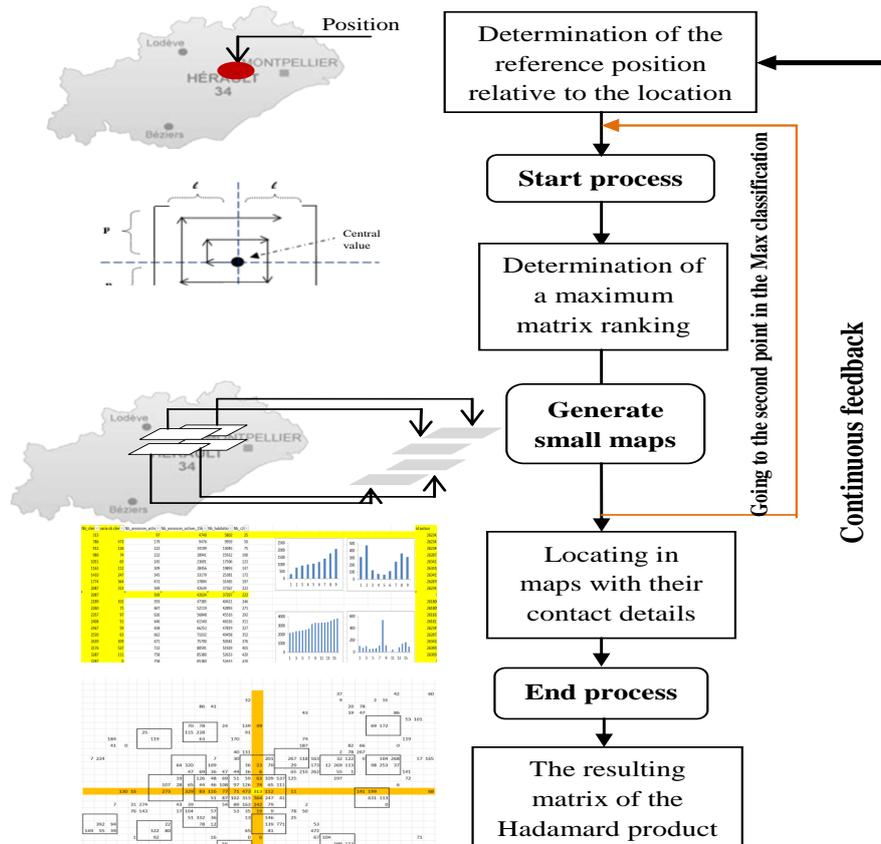


Figure 2: Snail algorithm process.

We are interested to the presentation of a point defined by a position on a plan or map that could be helpful to the user. And the most effective way for a retail designer map is to define the geographic coordinates in the coordinate system (X,Y). It should be noted that the situation of a point is expressed as coordinates in a reference geo-localization system in terms of territorial coverage and the quality of resulting data.

3.2. Data processing

First we must cut out the Local Area Map 3 km away into square (piece).



Figure 3: Cutting out the map of Herault.

Then, to start searching it must first define the position. We start from the point which is the central value of the indicated positions, around which rotates on a radius of 3 km range.

The first step is to determine a ranking of maximum matrix, it is not necessary that the starting point or the central value is a Max card, and then determine a point (x: for example) than 1000 people, in a square of 200 m on both sides, which makes a first square matrix, and each time there is a superimposed array. All that to find the median matrix " $M_{2p+1, 2p+1}$ ".

In which the target region (map Herault department) is split into small grids. On a central point (max) removing the map is going to have a "mille feuille" on geo-location we will find small maps localized that bring together applicants and suppliers.

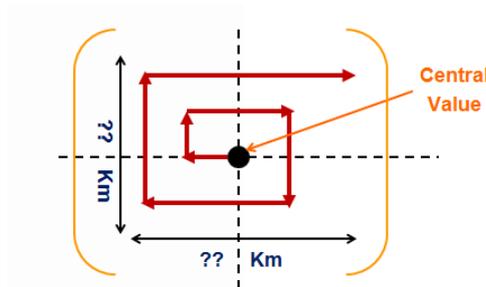


Figure 4: Graphic presentation of the snail algorithm.

Then, the same operation each time passing the second point in the ranking of Maxs, but there will be some cases where the second matrix in the ranking of maximum points will be included in the first round. It should be noted that used the Hadamard matrix product, which represents a binary operation for two matrices of the same dimensions, combines another matrix of the same size where each coefficient is the term by term product of the two matrices. Just like the addition, the Hadamard product is defined only for the matrices of the same dimensions, and the product of two matrices A and B is defined by:

$$A*B = [a_{i,j} * b_{i,j}] = \begin{pmatrix} a_{1,1}*b_{1,1} & a_{1,2}*b_{1,2} & \dots & a_{1,j}*b_{1,j} \\ a_{2,1}*b_{2,1} & a_{2,2}*b_{2,2} & \dots & a_{2,j}*b_{2,j} \\ \dots & \dots & \dots & \dots \\ a_{i,1}*b_{i,1} & a_{i,2}*b_{i,2} & \dots & a_{i,j}*b_{i,j} \end{pmatrix}$$

3.3. Data processing

The previous phase will generate little maps that facilitate the research. So, instead of working on full maps we can have small maps. These small cards will be identified quickly with their contact information. But every time we finish analyzing a sub card the algorithm is spirit to pit the matrix (we replace by zero). The resulting matrix of the Hadamard product between two same-sized matrices contains the result of a multiplication element by element. The navigation data are determined by setting a plane tangent to a fixed point on the map of Herault. This method is a technique for mapping a set of points in a multi-dimensional space, using a matrix of distances calculated in the departure space. Items placed on the borders of central calculate their coordinates iteratively until it converges to the Max of Maxs, and in this state there is performed the rotation around this point to calculate the product. That is to say, one seeks the Max throne the matrix "A" and removes the matrix "B".

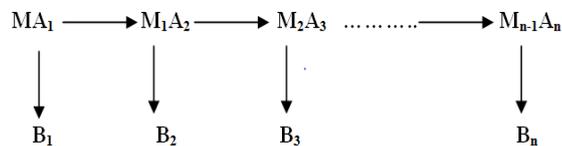


Figure 5: Transition from the first to the last matrix.

Such as:

A1 is the first matrix with enthroned extraction B1, An is the first matrix with enthroned extraction Bn. Note that "An" that represents the last enthroned matrix is a zero matrix because it always replaces the zero value as follows:

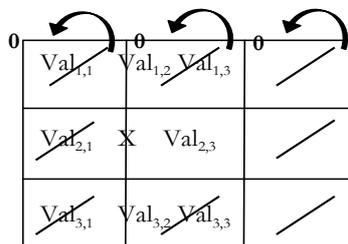


Figure 6: the last matrix form.

The great quest data is primarily aimed at understanding what customers' need, what works, what could be improved. Concretely it is to quickly collect raw data, explore and analyze, translate data into actionable information, and therefore globally to reduce the time between the discovery of relevant facts, the characterization of business opportunity and the outbreak of shares. The use of this algorithm reduces the information and develops a methodology based on the interaction data from several sources to provide reliable positioning and determining a navigation method. The Big data is becoming a tool to help companies explore new territories (Monino et al, 2014). For these gold mines can reveal intrinsic value which was not necessarily expected to start the analysis.

This algorithm was applied by the company "Autour.com" (see appendix). Know and meet their neighbors, helping each other, exchanging objects or services or to make some money by renting or selling property is what Autour.com offers to its members through use of algorithm.

But it is also an opportunity to promote a different type of consumption, sustainable and for social cohesion, focusing on rental, lending or exchange of goods or services. Autour.com today has over 15,000 profiles mainly in Montpellier but also throughout France as the national network is now

With using and application of the algorithm autour.com try to help their client to:

- Know and meet they neighbors
- Exchanging objects or services etc

In general this enterprise aims to mixing ads or mails for individuals, local shops, public institution and associations to better know the resources of their neighborhood (environment) the aim is to create a local social network to make a person who leaves next in contact

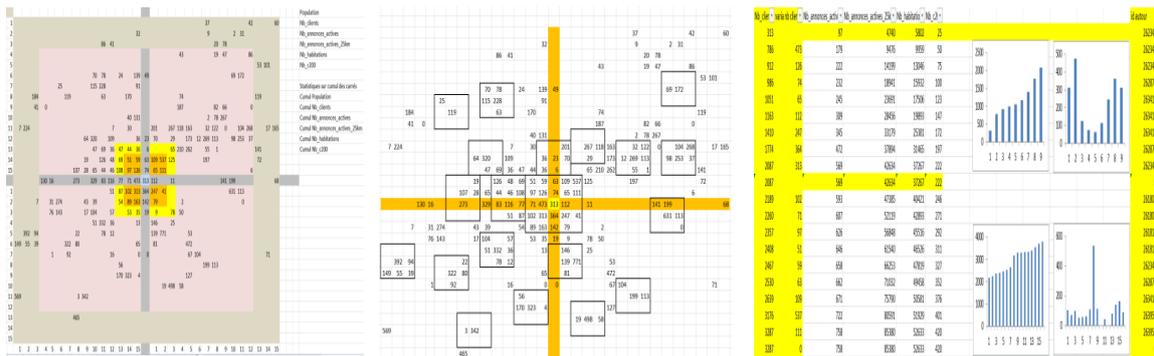
4. Conclusion

The algorithm just presented is only one example enabling companies to monetize their data. In this paper, we open a new application area for people living in the same area or district. This research provides a possible solution to the growing quantity of data. This work has immediate business implications with autour.com. The results of the application of our algorithm with this company show that our algorithm has a better performance, but it needs to be developed, with integration of many other operations.

This algorithm is a means that enables the company to better monetize Big data is a matter of understanding the model it wants to follow to define its way into the big data. For that currently have data on its customers or in the market is a competitive advantage. Finally, we wish that we can develop our algorithm and extend its advantages, in the field of health for example, recently we had a contact with a startup company in Morocco in order to develop this method in the field of health. This small company wants to create "moving parts" for kidney dialysis, with the integration of all necessary the tools and techniques in semi-trailers and go to the patient.

5. Appendices

Sample Application with “Autour.com”



Acknowledgements

The authors would like to thank to gratefully and sincerely Mr. Abdelhalim Mekbel (English Trainer at Anadarko company) for his critical reading, revising and language corrections of this article. This accomplishment would not have been possible without him. We appreciate all his interest and friendship.

6. References

Addlesee M., Curwen R., Hodges S. et al, Implementing a Sentient Computing System, IEEE Computer, vol. 34, no. 8, pp. 50-56, 2001.

Antoni J.P., Urban sprawl modelling : A methodological approach, Published in Cybergeog, European Journal of Geography, 2002, Topics, 12th European Colloquium on Quantitative and Theoretical Geography.St-Valery-en-Caux, France, September 7-11, 2001. St-Valery-en-Caux, France, September 7-11, 2001.

Cranshaw, J., Schwartz, R., Hong, J. I, and Sadeh, N., The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In Proceedings of ICWSM 2012, 2012.

Hoey J., Chantier M. J., Trucco E. et al, Tracking using flock of features, with application to assisted handwashing, British Machine Vision Conference BMVC, pp. 367-376, 2006.

Langlois, P. et Lajoie G., Cartographie par carroyage et précision spatiale, Mappemonde, vol. 49, no 1, p. 20-23, 1998.

Mazurek H. et Dayre P., Analyse de l'utilisation du sol par la méthode du carroyage: le District Urbain de Montpellier, Mappemonde, n°88/3, GIP. RECULS, pp. 27-30, 1988.

Mcnulty E., Big data: the seven V's, available at: <http://dataconomy.com/seven-vs-big-data/>, 2014.

Monino L.L., Sedkaoui S., Matouk J., Big data, éthique des données, et entreprises, Les Cahiers du CEDIMES, Dossier “Economie et gouvernance”, vol. 8, no. 2, available at: http://cedimes.com/images/documents/cahiers_cedimes/cahier_cedimes_vol_8_N2_2014.pdf, 2014.



- Song, C., Qu, Z., Blumm, N., and Barabási, AL, Limits of Predictability in Human Mobility, *Science* 327, no. 5968, 1018-1021, 2010.
- Zhang, D., Guo, B., Li, B., and Yu, Z. Extracting Social and Community Intelligence from Digital Footprints: An Emerging Research Area, *Ubiquitous Intelligence and Computing*, 4-18, 2010.
- Zheng, Y., Zhang, L., Ma, Z., Xie, X., and Ma, W. Y, Recommending Friends and Locations Based on Individual Location History, *ACM Transactions on the Web (TWEB)* 5, no. 1, 5, 2011.
- Zhou J., and Shi J., RFID localization algorithms and applications—a review, *Journal of Intelligent Manufacturing*, vol. 20, no. 6, pp. 695-707, 2008.

