An adaptive agent for Google Place crawling

Domenico Monaco

April 28, 2019

An adaptive agent for Google Place crawling

Domenico Monaco[1]

d.monaco2@campus.unimib.it, monaco.d@gmail.com

Abstract

This agent was developed over the question "how to crawl all google place of an urban area?". To solve this i developed a data crawler for Google Place as and adaptive agent ([Guessoum(2004)]). This was used to crawling data in the Master Thesis "Geospatial data analysis for Urban informatics applications: the case of the Google Place of the City of Milan" [MONACO(2018)].

In this context the urban area is a real envrioment while the Google Place API is a digital representaion of this, both have some limitations and rules to access it. In this case our envrioment is Google Place API while the agent goal is capturing all Places of an area with a minimum input. [Google(2017)]

This agent in completely autonomy, whit a minimum input, capture all Place by center position to a maximum diameter, both specified by user. This is realized over a spiral movement inspired on my Roomba spiral-pattern [Jones and Mass(2013)]. In this case the cells where the agent moves are hexagons, why this are best approximation of circle and has the property to have same distance by neighboring hexagons. [Hales(2001)]

The agent work on three steps:

- Planning of track with cells of crawling;
- Process of crawling over cells;
- Checking of results and if necessary replanning of track of crawling in more fine cells;

The minimun user-input are composed by: center of crawling, default size of cells and finally the number of spirals of crawling.

The algorithm choose when use more smaller cells and where, and, if there are some problem, where re-planning cells. So the core of algorithm is the adaptation on some envrioemental details of his planned track and granularity tessellation previously planned.

Keywords: adaptive agent, data crawling, geospatial data, social media, urban analysis.

1 Introduction

Digital environments are become more a more pervasive in every aspect of our life, just think to more and more social networks that are vital in our daly activity. In this way the complexity of real world is progressive replicated in digital environments. Therefore, this are not a simplified representation of real world but in more case are compose by a huge quantity of data, more than a human can consider in real world.

By this considerations, the idea is that if you want explore, crawl or analyze digital envrioment you must consider the digital envrioment as a real complex envriomant in a same way of real world.

In other word, an simple crawler that has a goal to acquire all place of specified area by Google Place API, must became and intelligent crawler that work as an agent in real world, with a set of Actions, Behaviours, Adaptation and Fault tolerance skills; all of this with minimum user-input.

2 Envrioment

In this work we have two level of envrioments, the real world and his -partial- digital representation.

2.1 Real envrioment

The real environment in this case is a urban area with many areas, places, points of interest and roads classically representend over a Geospatial Social Networks like a Google Map, OpenStreetMap and more. These are the our digitial envrioments, as digitial represention of real world.

2.2 Digital envrioment

Google Map, OpenStreetMap and more are the our digitial envrioments as digitial represention of real world. They have some rule of access and limitations, sometimes similar to real world, sometime very different.

In our case the digital environment is Google Place API, that enable people to accessing to places, areas and roads located on earth surface by Web API.

2.2.1 Rule of access

With a simplifications of case we can consider the environment accessible by a single Web API the give us vision on envrioment limited on circular shape.

This rule reqquire us to acces to envrioment by GPS coordinate and the radius of area that we want to access..

2.2.2 Limitations

The limitations, with same simplification of case, is 60 maximum Places that we can have as response by environment (Google Place API).

We not consider the other technical limitations such as the number of requests per hour or the paging of replies.

3 Adaptive Agent

The our adaptive agent have goal to acquire all of places located in a such area of earth (in this case Milan city) in a respect of previous environmental rules and limitations of environment access. With a minimum user-input and in completely autonomy.

3.1 Agent envrioment representation

For first, we must define a secondary digital representation of envrioment, specific to the agent. In this case, according to environment rules, limitations and agent goal the agent representation is discrete tessellation that able the agent to cover all space of an area with a minumum overlapping considering a circle shape accessing.

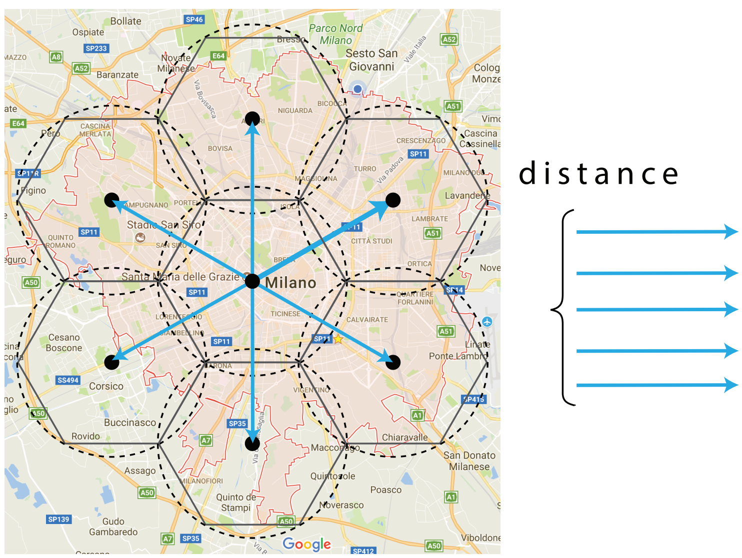

Figure 1: Hexagons tessellation as circle approximation

In this case the agent envrioment represantaion is a hexagon tessellation, that in according to [Hales(2001)] is a best approximation of circle with a minimum overlapping between circles (figure 1). In addition, respect to others tessellation we have same distance by neighboring hexagons.

3.2 Actions

For second, we must define a basics actions that agent is able to have in envrionment his representation of envrioment.

The agent must basically be moving over his discrete representation, so his basics action are the six movements to six neighboring hexagons. All othermovements are directly derived from these as composition or repetions of these.

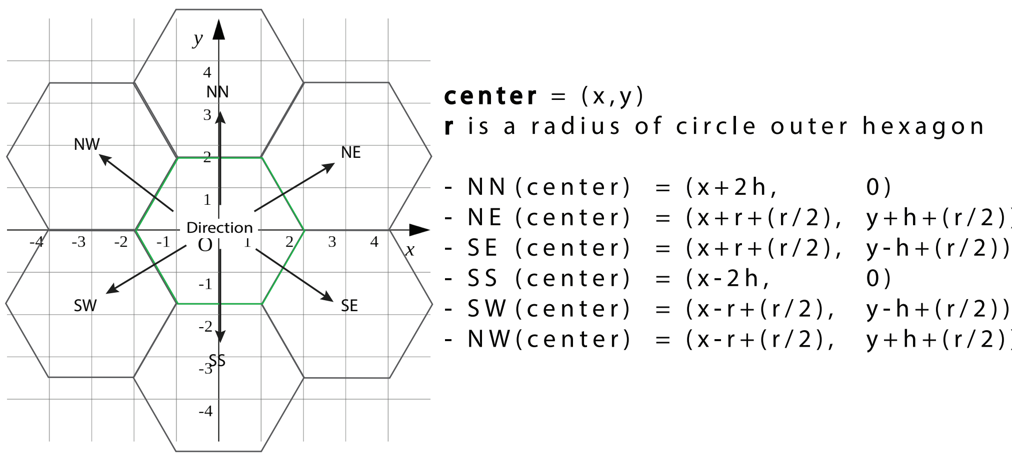

Figure 2: Basic six movements of agent over hexagon tessellation

3.3 Behaviour

In according to environmental rules, limitation and especially for the his goal the basic behaviour is a spiral-pattern movement inspired by Roomba initial exploration movement. [Jones and Mass(2013)]

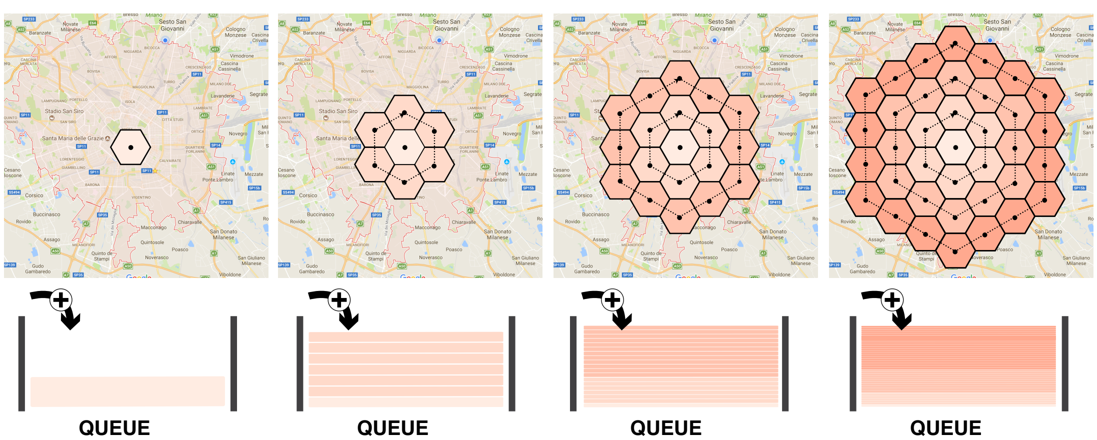

Figure 3: Spiral-pattern movement over hexagons tessellation

Because, in this case, the agent evrioment representation is an hexagon tessealation, the basic behaviour is implemented with a FIFO Queue that is filled with hexagon center position. From the first hexagon in initial position, and progressively with the other hexagon following a spiral-pattern. (Figure 3)

In other words the agent, starting from the center, calculate the position of every hexagons ring-by-ring. Where the intial hexagon for the i-esimo is the one positioned in top of ring.

3.4 Adaptation

The adaptations of agent is directly derived from the limitation of 60 maximum Place for request. Because, we can't know the density of every area in according to the previous limitation.

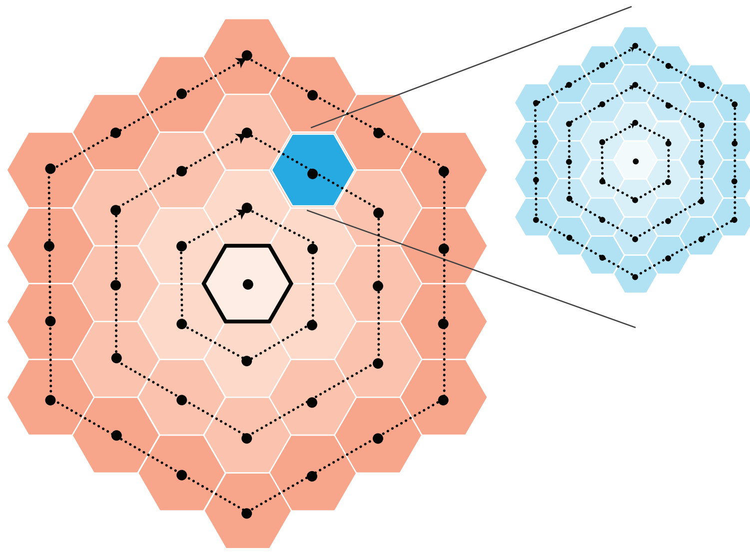

Figure 4: Fine grained tessellation of an hexagon

For this reason the agent, in a case of dection of 60 places in a single hexagon, he plan a new small spiral with more fine grained hexagon that overlap the hexagon where are detected a 60 Place limitaion. (Figure 4)

3.5 Fault tolerance

The fault tolerance is derived from the technology used by Google Place API. In fact when we using the Web API is possible that we have some problem such as comunication errors, wrongs API parameters and some other problem derived of internet protocols communications or API access policy.

In this case the agent actuate a conduct aimed at correcting the error implemented in a simple re-planning in a FIFO Queue the visiting of a specifc hexagon where is identified a problem.

3.6 Minimum user-input

Therefore, by giving these basics feature of adaptive agent we must define the minimum user-input to take a initial control of the agent.

The minimum user-input are:

- intial center position of crawling, in this case in GPS coordinate;
- the radius of default circle that is approfimated to the hexagon, in this case in meters;
- the numbers of ring of spiral-pattern;

- division ratio of fine grained hexagons tessellation for the adaptation behaviour;

By these inputs the adaptive agent can crawl the totally Places over any area of earth, in this case using a Google Place API, without losing place in a some area with high density o caused by error.

4 Conclusion and future works

In conclusion, the agent was tested for more than 6 month on Google Place API over 52 Km of diameter from Milan to near city like Monza and others. It has captured without losses over 290000 Places applying adaptation and fault tolerance behaviours in completely autonomy.

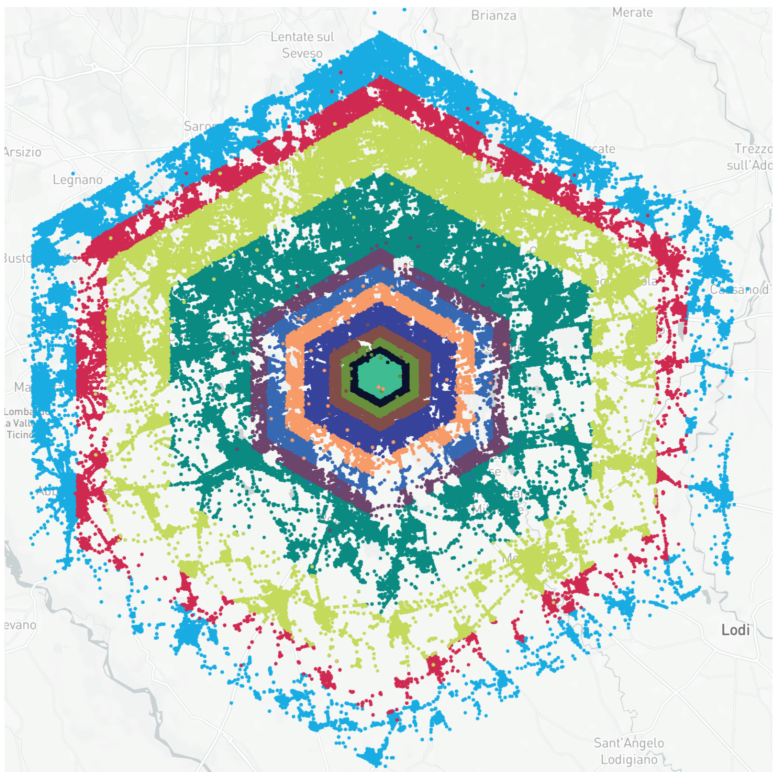

Figure 5: The 290000 Places over Milan area, divided in differents steps of crawling

For simplicity and firts debug using, the agent was equipped with a possibility of execute different portion of a spiral movement, as exaple in firs moment from ring 10 to 20, and in a second momento from 21 to 40. (Figure 5)

4.1 Different behaviours

Different behaviours, in according to the agent goal we can imagine other behaviors base to the six degree bisc movement as a custom track pattern that follow a different goal of this paper. This because the hexagon tessellation is very powerful discrete represenaton of a plane space with the equals movemets in six different direction with a same weight in a term of distance.

4.2 Other application

Obsviously, Google Place API are only a specific case of digital envrioments representation. By starting this we can imagine a other application of this agent as different Geolocated Social Network with same o similar rules of acces.

Or in other case we can generalize the approach to different use case, in example for drones for aerial photo mapping o simply sequential photo capturing over very large spaces where zoom or other factor can be varied during acquisition.

References

[Google(2017)] LLC Google. Google places api — google developers, 2017. URL https://developers.google.com/places/. (2017).

[Guessoum(2004)] Zahia Guessoum. Adaptive agents and multiagent systems. *IEEE Distributed Systems Online*, 5(7), 2004.

[Hales(2001)] Thomas C Hales. The honeycomb conjecture. *Discrete & Computational Geometry*, 25 (1):1–22, 2001.

[Jones and Mass(2013)] Joseph L Jones and Philip R Mass. Method and system for multi-mode coverage for an autonomous robot, June 11 2013. US Patent 8,463,438.

[MONACO(2018)] Domenico MONACO. Analisi di dati geospaziali per applicazioni di urban informatics: il caso dei google place nella citt di milano. Master's thesis, Universit degli Studi Milano-BICOCCA, 10 2018.