Time context based analysis for improving predictions of user movements using GPS data

André Alcântara[#1], Fabiana Toledo[#2]

[#]*Instituto de Pesquisas Eldorado*
Av. Alan Turing, 275 – Campinas/SP – Brazil
[1]andre.alcantara@eldorado.org.br
[2]fabiana.toledo@eldorado.org.br

***Abstract*— Human behavior is strongly connected to time-based events during day activities. Predicting user movements can be improved with timed context events. If context-rich systems can predict where a user is going to, it can make useful recommendations regarding to actions user does not want to forget. Users would be likely advised to go to place B while traveling from place A to C. Context-rich system knows user must leave a borrowed book at library B and B is in the way to C, where the user is probably going to next time. User movements' prediction can be improved by using day-time relationship. Here we show accurate predictions while considering morning, afternoon, evening and weekends when predicting next place user is going to. We found the accuracy is affected by user's behavior related to time context analysis since humans act in patterns during weekends, business days, business hours and late night or daybreak. When Markov Chains are fed with information about time, its accuracy raises as it is more probably going to lunch before afternoon and going to a club at weekend night. Our results demonstrates strong predictions when comparing the same data before time-based analysis and after. Introducing time context information to the Markov Chains achieved reliable predictions. We anticipate our technique to be an effective way for predicting user movements from a mobile device while running a context rich system. The present method is lightweight for mobile devices' current technology according to the previous research papers plus the accuracy obtained by the time-based analysis. This paper clarifies why time context information is useful to improve prediction quality of users' movements and implied activities.**

***Keywords*— Context-rich, Movement prediction, GPS, Places, Locations, Activities**

I. Introduction

While looking for the future of the computational intelligence regarding to context-rich systems and mobile devices or Personal Digital Assistants, we are taking care of the handheld side. With the present research we are looking for an alternative way for a lightweight software to detect locations, determine places and predict users' movements considering day and time.

We started this study by analyzing a set of GPS coordinates plotted in 3D graphs. By looking at the layers in the 3D graphs we found a hypothesis where user's history behavior is closely related to periods during each day. Business days, weekends and holidays may interfere in users' habits and behavior changes could be found also at the level of daytime sections like morning, afternoon and evening. Since it directly influence users' actions, we can use such information for accurately predicting the next place they are going to.

Based on the hypothesis of date and time influence in the user's daily activity and following the notion obtained by the 3D graphs of GPS coordinates, a Markov Chain were established for calculating probabilities regarding to the next place users are going to, related to day of week and day's period.

This approach is not explored by the previous related work, present at section II, as an easy way to improve prediction accuracy at low cost calculations, allowing mobile devices doing so.

We started an assay with 1,600 hundred GPS coordinates from January 2011 to October 2016 in a total of 67 megabytes of information obtained by Google Takeout [11].

This paper contains sections as follows: Section II presents previous related work regarding to place discovering, labeling and movements prediction. Section III is about to the initial study for introducing the concept of time influence on daily basis activities. Section IV demonstrates the techniques used for reproducing previous work improving it with the new concept of time-based analysis for predictions on mobile devices. Section V shows the experiment results related to developed software and finally, section VI is a conclusion of this work and related future work.

II. Related Work

Our current work aims to improve predictions in a lightweight solution for resource-limited mobile devices by scrubbing the best achievement of each previous work plus our new approach of introducing time-based information to the Markov Prediction Model.

Previous researchers didn't have current sophisticated environment for creating the knowledge related to identify places in locations. Most of their problems are related to GPS devices with low duration batteries and map frameworks difficult to operate. Nowadays every mobile phone has a GPS reporting history location integrated to Google Maps through Google Location History [11] and also GeoPy [19], a Python toolbox for geocoding. So many previous problems are no more obstacles to advance in research of location-aware systems.

Location-based prediction scenario is current covered by Jong K. et. al. [1], [2] which contributed with extracting places from locations. A place is somewhere with a semantic meaning to the users, as their work, home, gym, etc. They provided a useful time-based clustering algorithm simple enough to run on a resource-limited mobile device and able to work determining important places on the go. Their algorithm is robust enough to determine places while removing outlier coordinates. Other advantage of their approach is related to its parameters can be dynamically changed for working with places and sub-places as a university as a place and its departments as a sub-places.

Many other researchers are looking for improving places detection, movements prediction and plus activity inference. Thierry B. et. al. [10] proposed a kernel-based algorithm with resilience to GPS noise.

Lin L. et. al. [3], [4] did the most complete research for doing place extraction, activity discovery and labeling. Their method takes care of inaccurate place detection regarding to there is no threshold for permanent satisfying the discovery of places and sub-places at the same time. In the way of solving detection and provide high accuracy, they labeled places inferring activities according to their model. They started using relational Markov Networks [12] and Conditional Random Fields [13], [14], [15], [16], [17].

In the other side of location based systems, Daniel A. et. al. [5], [6] focused on predicting users' movements regarding to probabilities for transitions in first and second order Markov Models [8], [9], [12]. Their study inspired the observation of users' movements as a pattern frequent enough for doing strong based predictions at this current work. They also ran their experiment for multiple users environment allowing cross classifications of meaningful places.

Finally, we have Parth B. et. al. [7] with a modern apparatus with benefits of previous works above in the way of help users get in time on their schedule. To reach this target they use predictions for calculating travel times and keep users aware of their meetings considering previous research from Inferring Calendar Attendance [18].

A location based prediction system must have elements such as extracting significant and meaningful places from locations in raw GPS data, predict user movements in between places and infer user activities, so knowing user actions in their context is the base for feeding context-rich systems.

As presented, researchers are exploring this vast subject with many different approaches, each one with pros and cons. Our main advantage, not yet explored by anyone, is the use of periods of daytime while predicting user movements regarding to business days, weekends and holidays. This way we have better predictions and accurate ratings, but keeping low cost when processing GPS series of data.

III. Using GPS raw data for 3D Graphs analysis

A previous understanding of users' habits is desired for better predicting users' behavior regarding to their pattern of movements. An easy way to check patterns to formulate a first hypothesis is to observe the traces of travels users do. Mapping users' location history as a heat map or even plotting the entire raw history is a good way to check whether the users are spending their time.

We started testing many ways to visualize users' paths. By using Octave [21], we read entire Google

Location History of a user to see where he has been. Observing the traces, it is easy to formulate a hypothesis where the bold lines are the most common paths and the relevant places are seen larger than lines. First map produced with standard shape files [20] denoted paths and places marked inside red circles as seen in Fig. 1 Paths and places' sizes are relative since some ways are more relevant than others regarding to the bold circles and the tiny circles.

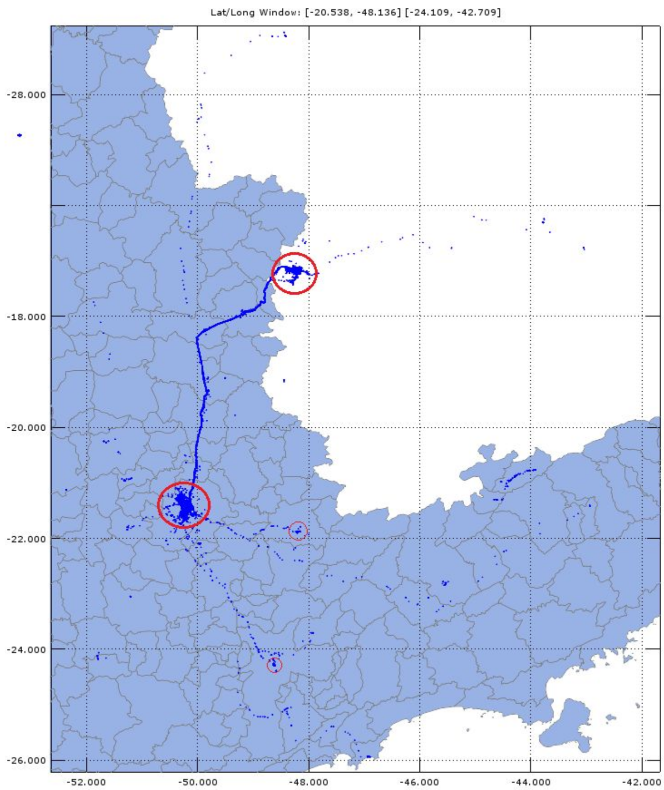

Fig. 1 Shapefile with raw GPS data plotted by Octave.

To ensure there are patterns in users' behavior, we test k-means [22] algorithm in the raw GPS data considering latitude, longitude and timestamp. Timestamp can be grouped in many different ways; its number does not fit any grouping criteria since it is a single number. By converting timestamp to year, month, day, hour, minute and second, we can group a set of GPS coordinates filtering business days, weekends and day's periods like morning, afternoon and night. As a result, for the second test, Fig. 2 displays a color flat map with user's movements. The clustering process separate groups as red, green and yellow as more visible due to its density. Red dots are present in most of time at the end of map while green and yellow are along the ways user moves.

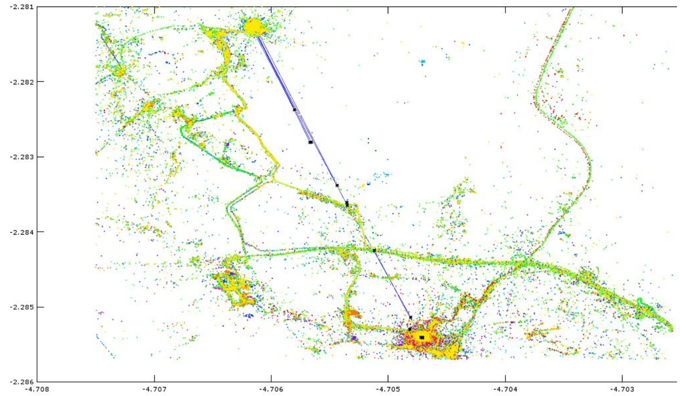

Fig. 2 A set of k-means clustering GPS coordinates from Campinas-SP.

Once we know the user's behavior, it is easy to calibrate k-means algorithm for better clustering results. In the Fig. 3 there are clusters well defined for home as yellow, work as red and lunch as purple. In addition, the way from home to work is cyan and the way back to home is green. Of course, even in a flat map the notion of superposition is clear since yellow dots comes first, than cyan, red, purple and green.

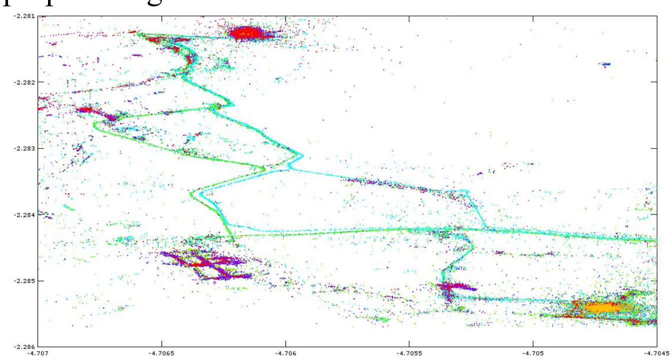

Fig. 3 A set of k-means clustering GPS coordinates with zoom.

Therefore, plotting this map in 3D is useful to confirm each group is being clustered according to a relation between latitude, longitude and time. Our first observation in 3D graph were made with a 7 cluster k-means set. Having the knowledge where is the user's home we can see at the bottom right corner of Fig. 4 two early morning clusters as green and yellow, next is morning when user leaves home in the way to his work, clustered as red. This view puts all points together in the same cluster so it is not possible to know the direction of flow. Since this is an observation of known habits, to confirm their patterns, is clear that cyan color is happening between afternoon and evening, where user goes back to home. Purple indicates night period with greater density at home and in a shopping mall.

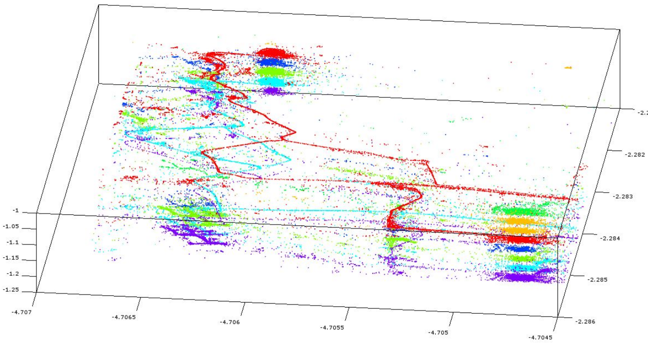

Fig. 4 A 3D set of k-means clustering with GPS coordinates.

Road tracks are always present at the top and the bottom of 3D graphs. When k-means groups clusters based on time, it puts work place together home. There is no k-means settings for automatically group roads and places based on periods.

A. Looking for an Ideal Clusterization Settings

Clustering GPS data with good settings is difficult due to several constraints. Each of the following graphs demonstrate different cluster groups. Even when you find a good set of parameters to describe groups that put it all together, it always has some kind of 'leak', keeping uncovered groups outside the patterns and standards.

1) *Work versus Rest*: In a 3D spatial distribution, groups are defined in terms of periods. Adopting the correct scale while grouping hours is the key for achieving the best set. Unfortunately, cycles during daytime are not regular and measuring working hours must consider lunchtime. By this rule, working periods are split in two groups: before and after lunch. Also, there must be a prior group for morning, before work and a group for late afternoon or early evening after work. For a good example, in Fig. 5, there are red clusters for working hours. These k-means configurations are good to see early morning in green, breakfast in blue, end work in purple and finally back to home or dinner in light green. The main problem of this approach is finding lunch, which may be considered also as a working hour.

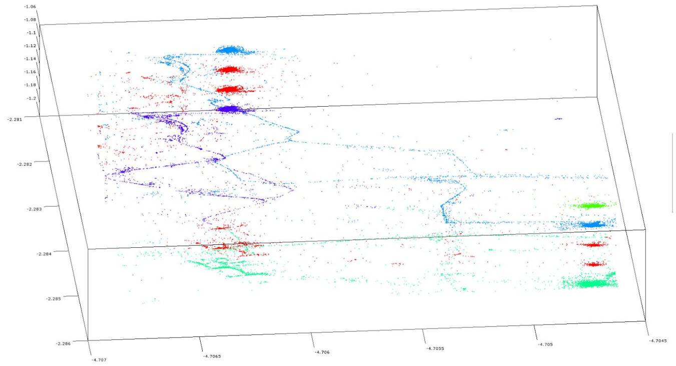

Fig. 5 A 3D filtered set of k-means clustering focusing work time.

When changing settings for another approach, as in Fig. 6, even when the working time is split in red and green, lunchtime does not appear in the graph as well. Lunchtime belongs to the red group but its density is low and just a few dots are visible in the graph.

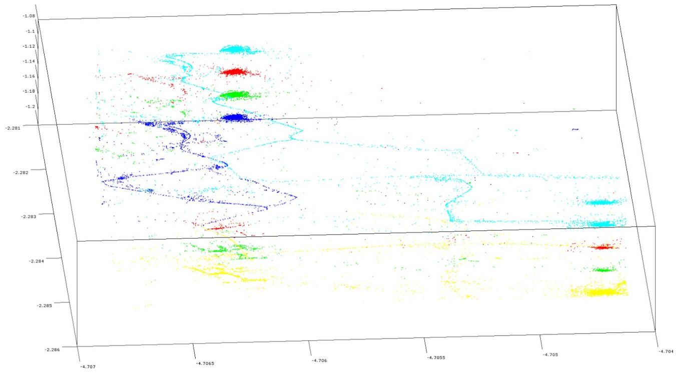

Fig. 6 A 3D filtered set of k-means clustering splitting work time.

2) *Places versus Tracks*: Users spend their times in places longer than 30 minutes which leads to use at least a whole our for cluster size. However, traveling from one place to another does not take, in general, more than 40 minutes. By noticing the track back to home, Fig. 6 in blue, it becomes yellow in the half way. Trying to minimize and isolate work cluster from its following track, in Fig. 7, we see new clusters just with tracks in yellow and purple. Even when k-means clustering is capable to isolate tracks from places, there is no way to differentiate between them and it does not work all the time since, in Fig. 7, we see tracks from early morning are still clustered together home and work morning clusters.

3) *Weekend versus Business day*: Another relevant point regarding to context behavior classification is that users intended to act different while working or resting, according to Fig. 8 which considers only GPS data timestamp from weekends and holidays. There is no trace of work clusters at the graph. This user behavior also changes regarding to lunchtime, since lunch is usually clustered together lime color group if a business day. By the weekends and holidays it is clustered a little late in the green cluster. Dinner is also noticed as yellow.

Separate business days apart from weekends and holidays is an effective filter. According to Fig. 9, workplace is correctly grouped in four clusters as early morning, morning, afternoon and evening while home place does not have any cluster but early morning and night.

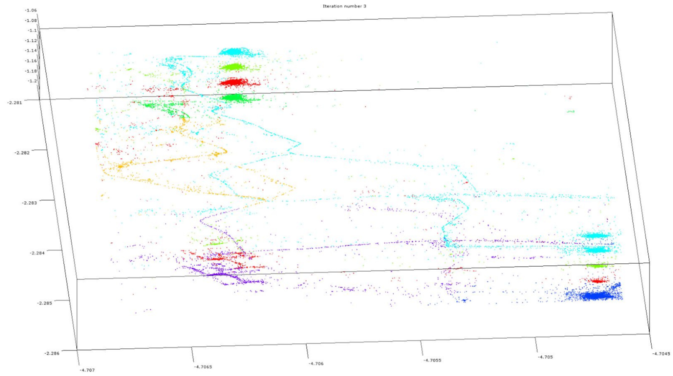

Fig. 7 A 3D set of k-means clustering with separated tracks and places.

Comparing Fig. 10 to Fig. 9 is possible to see k-means clustering is failing in day's edges. While Fig. 9 mix morning home and work clusters in blue, Fig. 10 mix late evening work and home clusters in purple.

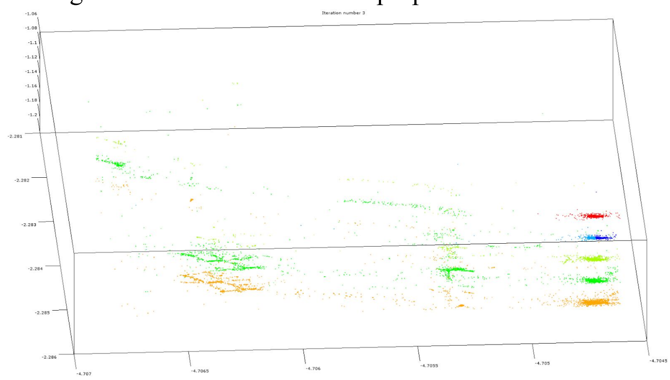

Fig. 8 A 3D set of k-means clusters with just weekend and holidays GPS data.

Even when telling k-means to use the correct number of clusters, the result is complex to handle, demanding a pipeline for processing the result, which adds cost of calculations. Even if other approaches with Mean Shift [23], [24] or X-Means [25], [26] are used, when is not needed to provide number of clusters, they fail to get the ideal number of clusters. Many colored clusters can be found at Fig. 11 and the extended number of clusters do not help in predictions indeed.

B. Looking for predictions related to lunch-time

Time for lunch is a recurring problem mentioned in previous sections. Lunch clusters are small in comparison to place clusters. Applying Principal Component Analysis in latitude and longitude, over the time, we have Fig. 12 describing where users is during a period. Horizontal lines at the top of the graph represent the home location and horizontal lines at de bottom are representing work location.

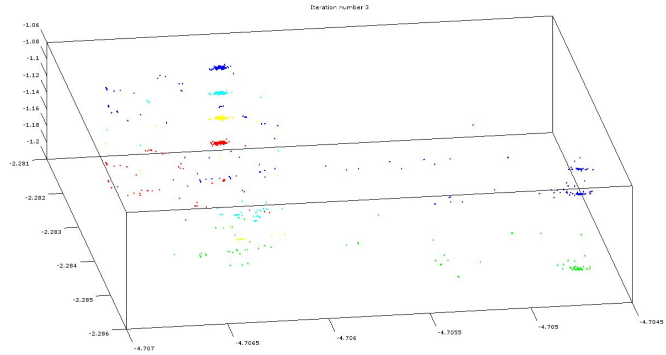

Fig. 9 A 3D set of k-means clusters with just business days GPS data.

While at home, we notice that there is no interruption in the location, but work line representation has a variation in the middle of day, that represents user's movements regarding to lunch.

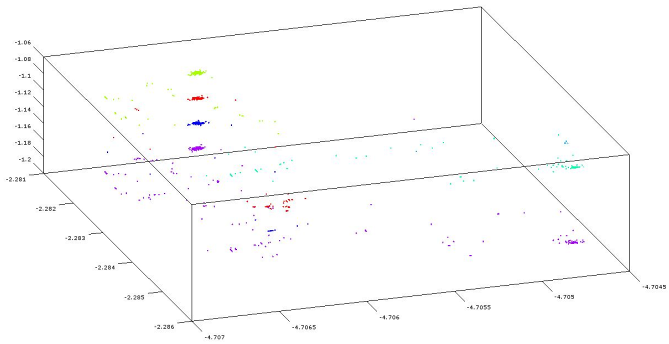

Fig. 10 A 3D k-means cluster set with business days with mixed last cluster.

All those graphs are suggesting users' movements are directly related to the environment on their around. A PCA analysis [28] in Fig. 13 shows user is not at his work sometime around 12 pm usually.

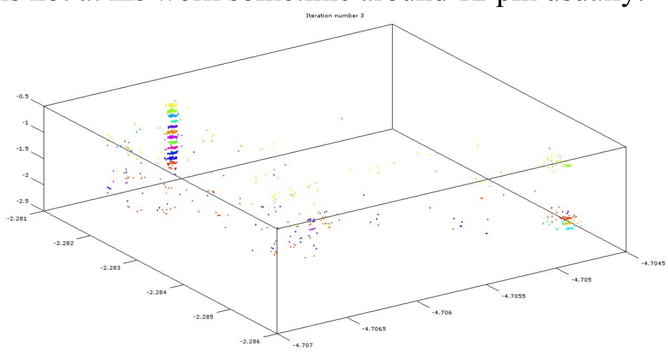

Fig. 11 A 3D set of k-means with many clusters of work GPS coordinates.

As seen in Fig. 12, there is a single linear view of a streamed user's behavior and plotting them all together, we have Fig. 13 where Z axis represents each day of month. In this view, there is no density enough to see relevant changes in lunch-time. It seems outliers or noise.

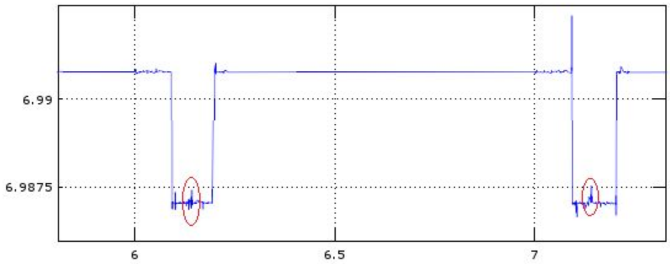

Fig. 12 Principal Component Analysis with lunch-time movement.

Rotating the graph in Fig. 15, demonstrate a little more clear dispersion of dots representing the lunch period. Since there is a restaurant inside user's company, it is very common having lunch and work clusters together.

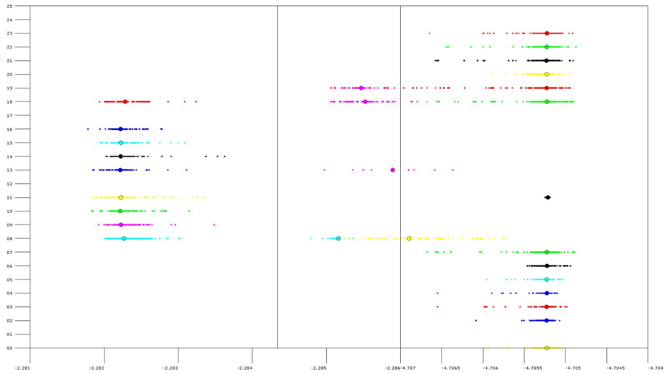

Fig. 13 A 24 hours cluster with locations between work and home.

Based on 3D graph representation of time-based GPS coordinates, this work sustained a hypothesis that time influences directly on predictions.

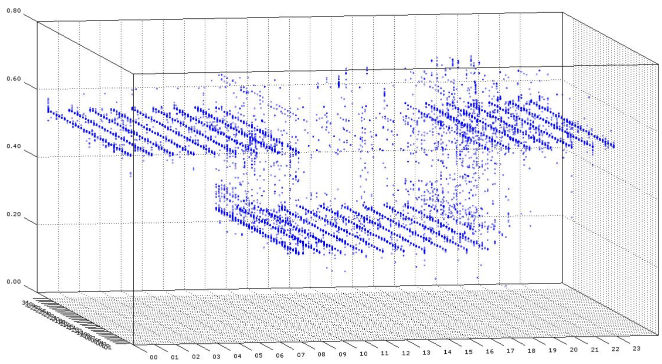

Fig. 14 Daily activity by month view per hour.

So, we started next stage of the research, looking for a way to predict users' movements considering the period of daily time and date.

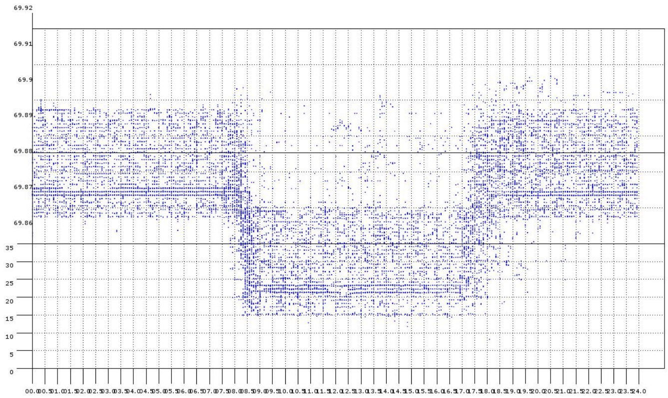

Fig. 15 Daily activity by month flat view per hour .

Since this proposal is aiming to improve previous results from other relevant research, we decided to pick up the better feature present in each previous work to compose our base research platform. According to the next section, we present the way we achieve our results.

C. Detecting Places and Predicting Users' Movements

According to the previous work, from section II, there are many ways to look for places. Google Maps [27] does this as well. Going to visited menu, under Maps Android Application, there is a list as in Fig. 16 with each recent places users visited.

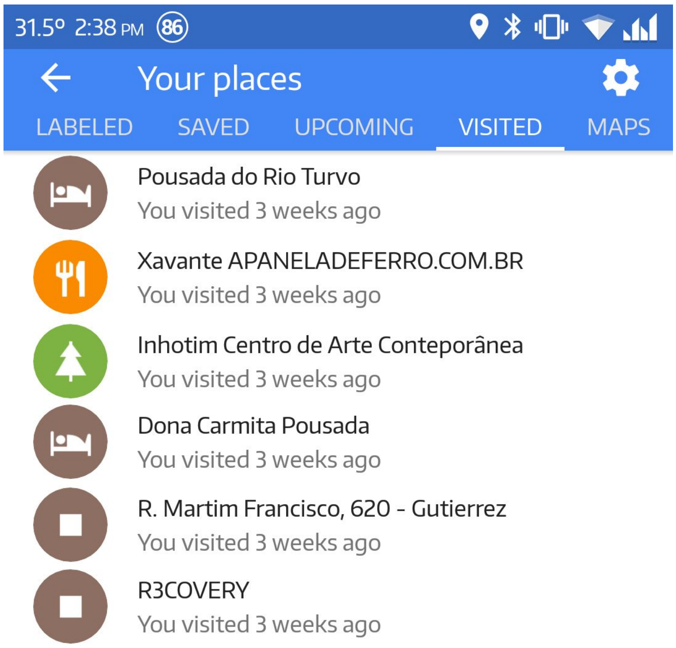

Fig. 16 Google Maps Android application with a list of visited places.

To make this research possible, we adopted Jong K. et. al. [1], [2] method for extracting places from traces of locations. In their research, a time-based clustering algorithm was developed to run on a resource-limited mobile device. A pseudocode is presented in Fig. 17 and their deep explanation is found at researchers' papers as [1] and [2].

```
cluster(loc)
input: measured location loc
State: current cluster cl,
       pending locations plocs,
       significant places Places

if distance(cl, loc) < d then
   add loc to cl
   clear plocs
else
   if plocs.length > l then
      if duration(cl) > t then
         add cl to Places
      clear cl
      add plocs.end to cl
      clear plocs
      if distance(cl, loc) < d then
         add loc to c
         clear plocs
      else
         add loc to plocs
   else
      add loc to plocs
```

Fig. 17 Time-based clustering algorithm from Jong K. et. al. [1], [2].

Although Lin L. et. al. [3], [4] should be the robust way for extracting places from GPS traces, we found satisfactory results while using Jong K. [1], [2] presented in Fig. 18, and after developing the pseudo code presented in Fig. 17, using GeoPy [19], we reach a fair set of visited places. Those results composed our baseline for feeding the Markov Chain model as described by Daniel A. et. al. [6], [7], which have also their own method for extracting places in their article. The Markov model described in [6], [7] uses information from transitions from a place to another. For predicting where user is going to, Markov decision is taken by a relative frequency, calculated by all transitions that comes from a place A to B divided by all transitions that comes from A to anywhere. Another second calculus is made considering also where users were before. With information from second order, prediction results are improved, so user was in place A before going to place B and based on this we predict next place C.

Previous work from [3], [4] just uses transitions from one place to another for predicting where users are going to. In this present work, we add information helper to this model by our understanding time does matter.

Fig. 18 A 3D set of k-means clustering with GPS coordinates.

Assuming place A is a shopping mall and user is at home B, user' probability to leave place B going to place A depends on the period of day. Users are more intended to do the same kind of things at the same periods, e.g. having lunch, dinner or breakfast and those cycles also are affected by business days or weekends and holidays.

While working with context-rich systems, time is an important context, since some actions are time-related, so we can assume the reason for going to place A, from place B, in the morning, may be a breakfast. Going from Home to a Shopping Mall, at night, maybe diner, cinema, or even both and this event is more likely to happen at the weekends or holidays. Considering context, a transition between home and shopping has a heavyweight at Saturday night than at Monday morning.

IV. Time-Based Improvements

Since 2003 technology has changed a lot. Daniel. A. et. al. [6], [7] faced difficulties while collecting data due to their apparatus in the Zürich study. Nowadays mobile phones are integrated with GPS and they collect location data every time, everywhere. Our database is full of coordinates, one from 2011 to 2016 and another from 2015 to 2016, which means we do not have a natural limitation in the quantity of data available for analysis and the number of second order transitions is not relatively small.

Then we started implementing the Markov model described in [6], [7] to reproduce their assay. With GeoPy [19] we are fully able to set unique ID's for each place, including their address and real name, if available, such as Building Name.

The source code, in Python, walked through the entire GPS data, from begin to end in the timestamp ascending order generating a table with pairs and triads of places, representing first and second order transitions, then transitions were calculated to obtain the relative frequency.

With this first Markov model, we have a baseline for comparison and we can do predictions in the same way of previous work.

Once their model is working in our Python development, we started our model based on their previous. The algorithm still consists the same, but now the unique ID proposed has changed.

First, we calculate transitions from A to B and also calculate transitions from A to B and to C. This way A is a unique place, with its ID. Changing the algorithm consists in treating a place not just the same place all the time, but a new place at each period of daytime.

According to Table I, a place will assume 6 different IDs, e.g. A0 when it is Early Morning, A2 when in the Afternoon and so on. Now, transitions from A to B are not just put all together, having A0 → B0; A1 → B2; A2 → B3 → C4.

TABLE I
MEANING OF PERIODS OF DAY

Number List for Time in Periods		
Code	**Period**	**Meaning**
0	06H to 09H	Early Morning
1	09H to 12H	Morning
2	12H to 15H	Afternoon
3	15H to 18H	Evening
4	18H to 21H	Night
5	21H to 06H	Late Night

As previously explained, the chances of a user going to some place (for breakfast) based on where he is, at the morning, are greater than if user is in the same place but at night. A user can usually go to have breakfast and goes just occasionally for dinner.

This behavior is also affected when it happens in business days or weekends and holidays. Using the same technique, we expanded again the unique ID for having another instance of differentiation. As seen in Table II, weekends and holidays are being separated from business days, then we have a place A in a combination with 6 periods and 2 types in a total of 12 different IDs for each place.

For example, a place A in the morning will assume A11 for a business day and A01 for a weekend or holiday. This notation is made by [Place|Day|Period] and indexes from Table II come first in the sequence, before indexes from Table I.

TABLE II
MEANING OF WORKING DAYS

Number List for Working Days	
Code	**Period**
0	Weekends, Holidays
1	Business Days

We can divide a day in many fractions, as many as we want. In this case, place A is expanded in 12 new cases. By slicing periods less than three hours will explode combinations and maybe cause redundant predictions as the user is not really staying one hour in each place. Slices greater than three hours are also a problem, since user may have a pause for a snack.

The decision for using chunks of 3 hours are based on the traditional way humans already use to classify periods of day, so intuitive as morning, afternoon, evening and night should represent our behavior in fidelity and seems to be effective.

In Table III we see an example of combination with many transitions. From A to B is possible to infer user is going to lunch by transitioning from 1 to 2 in the second number. First number is always 1, indicating a business day. As already discussed, sometimes transitions happens between different periods, which means we are facing the last hour from the previous period and the first hour from the next period. This case is good for diagnosing what is happening. In the second line of Table III we know user is going to D04 at a weekend or holiday and passing the boundary from Evening to Night.

TABLE III
PLACES IN TERMS OF TIME-BASED UNIQUE TRANSITIONS

Transitions	Description
A11 → B12 → A12	Last Morning to Afternoon (Lunch)
C03 → D04 → C04	Last Evening to Night (Cinema)
H05 → W10 → L12	Home Sunday to Work Monday to Lunch

These special transitions are meaningful. Note in the third line when user was at Home in a Sunday or Holiday H05 (late night) and then make a transition to Work in a Business Day (early morning).

Depending on GPS updates, sometimes we can have a situation when user transition from 0 (early morning) to 2 (afternoon) without passing from 1 (morning). This is not a problem for predictions since user stays at W. Even though system is blind for a stream of data, these missing events do not change the prediction results regarding to the next place is predicted as afternoon lunch L12.

V. PREDICTION EXPERIMENT RESULTS

Having the two models side by side, we analyzed the same data with both. Due to the original table has thousand lines, we picked up expressive samples for showing here in Table IV.

In a side-by-side comparison, our model filtered the first line of Table IV from 180 to 159 transitions that occurred in a business day early morning, improving prediction from 85.56% to 92.45%.

Another good example is the 402 transitions from “Home ⇒ Shopping Galleria ⇒ Home”, split in four new groups for the new model. While first three groups happened in weekends\holidays, the last one occurred in a business night. Even though we split a single set of transitions to four, the prediction improved from 92.04% to at best 94.59% in the worst case. Another curious observation we can make is when Home and Shopping are together, notice it happens in weekends or holidays according to lines 5 and 7.

TABLE IV
COMPARISON BETWEEN MARKOV MODELS BEFORE AND AFTER TIME-BASED APPROACH

Markov Model without Time-Based Approach			**Markov Model with Time-Based Approach**		
Track	**Relative Freq.**	**Accuracy**	**Track**	**Relative Freq.**	**Accuracy**
Galleria Office => Shopping Dom Pedro => Work	154/180	85,56%	Galleria Office|10 => Shopping Dom Pedro|10 => Work|10	147/159	92,45%
Galleria Office => Shopping Dom Pedro => Home	20/180	11,11%	Galleria Office|14 => Shopping Dom Pedro|14 => Home|14	5/6	83,33%
Home => Galleria Office => Home	135/947	14,26%	Home|02 => Galleria Office|02 => Home|02	12/13	92,31%
			Home|13 => Galleria Office|14 => Home|14	14/17	82,35%
			Home|14 => Galleria Office|14 => Home|14	18/21	85,71%
Home => Panetteria => Home	174/190	91,58%	Home|02 => Panetteria|02 => Home|02	81/83	97,59%
			Home|12 => Panetteria|12 => Home|12	20/20	100,00%
			Home|13 => Panetteria|02 => Home|02	15/18	83,33%
Home => Shopping Galleria => Home	370/402	92,04%	Home|02 => Shopping Galleria|02 => Home|02	217/222	97,75%
			Home|03 => Shopping Galleria|03 => Home|03	32/33	96,97%
			Home|04 => Shopping Galleria|04 => Home|04	16/16	100,00%
			Home|14 => Shopping Galleria|14 => Home|14	35/37	94,59%
Shopping Dom Pedro => Galleria Office => Home	119/149	79,87%	Shopping Dom Pedro|13 => Galleria Office|14 => Home|14	17/19	89,47%
Shopping Galleria => Home => Galleria Office	121/506	23,91%	Shopping Galleria|02 => Home|02 => Shopping Galleria|02	217/229	94,76%
			Shopping Galleria|03 => Home|03 => Shopping Galleria|03	32/37	86,49%
			Shopping Galleria|04 => Home|04 => Shopping Galleria|04	15/18	83,33%
Work => Galleria Office => Home	522/618	84,47%	Work|10 => Galleria Office|14 => Home|14	249/301	82,72%
			Work|12 => Galleria Office|14 => Home|14	18/18	100,00%
Work => Panetteria => Work	46/48	95,83%	Work|12 => Panetteria|12 => Work|12	29/29	100,00%
			Work|10 => Panetteria|12 => Work|12	9/9	100,00%

First number after pipe symbol
0: Weekend, Holidays
1: Business days

Second number after pipe symbol
0: 06H to 09H - Early Morning
1: 09H to 12H - Morning
2: 12H to 15H - Afternoon
3: 15H to 18H - Evening
4: 18H to 21H - Night
5: 21H to 06H - Late Night

The better predictions reflect the sense of user behaves in distinct ways on different dates, even when in the same places. The introduced IDs in a composed key also helps meaningful analysis for high-level interpretations of users' behavior. For context-rich systems, this can be a useful tool in terms of even more than just predicting where users are going to, but inferring what they are doing.

VI. Conclusions

During the progress of this research, we developed many tools using Octave [21] to analyze GPS raw data from Google Location History [11]. The toolset produced in this research generated various data visualization types, including 2D and 3D graphs. Some of techniques applied to the GPS data, like k-Means [22], X-Means [23], [24], MeanShift [25], [26] and PCA [28] helped constructing a high level comprehension of users' behavior, for creating the main hypothesis that time influence directly into the prediction results.

By picking up the best of each previous work, we developed a system able to extract places from traces of locations, on the fly, through a resource-limited mobile device as present in [1], [2]. And after having places found, we added context-time information into the Markov model from [5], [6] to enrich predictions with improved high accuracy.

The main result obtained from this research is the conclusion periods during a day can improve predictions from Markov model, creating a lightweight and reliable system as a tool for using integrated in Context-Rich Systems for predicting, in time-based context, where is the next place user can go.

A. Future Work

Next research includes a filter with some criteria to clear the Markov table with relative frequencies that are not relevant, in which we freeze some transitions until they have more data to be relevant. In this case, when users go to new places and we found just one or a few transitions.

An error analysis with Bias-Variance Tradeoff [29] will be applied to confirm the filter is improving predictions for this software can be optimized and ported to an Android Device.

Finally, extracting places from traces of location's settings should be automatically adjusted and also dynamically adjusted to decide whether we need zoom in or zoom out on the map for finding locations, sub-locations and predict even movements inside a place like a university campus or factory.

Acknowledgment

Authors of this paper wish to acknowledge Instituto de Pesquisas Eldorado for maintaining their internal Research Facility and Development Laboratory, which supported the accomplish of this work.

References

[1] Kang, Jong Hee, et al. "Extracting places from traces of locations." *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*. ACM, 2004.

[2] Kang, Jong Hee, et al. "Extracting places from traces of locations." *ACM SIGMOBILE Mobile Computing and Communications Review* 9.3 (2005): 58-68.

[3] Liao, Lin, Dieter Fox, and Henry Kautz. "Extracting places and activities from gps traces using hierarchical conditional random fields." *The International Journal of Robotics Research* 26.1 (2007): 119-134.

[4] Liao, Lin. *Location-based activity recognition*. Diss. University of Washington, 2006.

[5] Ashbrook, Daniel, and Thad Starner. "Learning significant locations and predicting user movement with GPS." *Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on*. IEEE, 2002.

[6] Ashbrook, Daniel, and Thad Starner. "Using GPS to learn significant locations and predict movement across multiple users." *Personal and Ubiquitous Computing* 7.5 (2003): 275-286.

[7] Bhawalkar, Parth, et al. "ScheduleNanny: Using GPS to Learn the User's Significant Locations, Travel Times and Schedule." *arXiv preprint cs/0409003* (2004).

[8] Marzolla, Moreno. "The Octave Queueing Package." *International Conference on Quantitative Evaluation of Systems*. Springer International Publishing, 2014.

[9] S. Cordwell. Markov decision processes: Python toolbox [online]. URL: https://github. com/sawcordwell/pymdptoolbox

[10] Thierry, Benoit, Basile Chaix, and Yan Kestens. "Detecting activity locations from raw GPS data: a novel kernel-based algorithm." *International journal of health geographics* 12.1 (2013): 1.

[11] Location History Download from Google Take Out retrieved from: https://takeout.google.com/settings/takeout/custom/location_history?hl=en&gl=US&expflags

[12] L. Liao, D. Fox, H. Kautz, Location-based activity recognition using relational Markov networks, in: *Proc. of the International Joint Conference on Artificial Intelligence* (IJCAI), 2005

[13] Kristjansson, Trausti, et al. "Interactive information extraction with constrained conditional random fields." *AAAI*. Vol. 4. 2004.

[14] Lafferty, John, Andrew McCallum, and Fernando Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data." *Proceedings of the eighteenth international conference on machine learning, ICML*. Vol. 1. 2001.

[15] Peng, F. C., and A. McCallum. "Accurate information extraction from research papers using conditional random fields." *Proceedings of the 2004 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 2004.

[16] Quattoni, Ariadna, Michael Collins, and Trevor Darrell. "Conditional random fields for object recognition." *Advances in neural information processing systems*. 2004.

[17] Sha, Fei, and Fernando Pereira. "Shallow parsing with conditional random fields." *Proceedings of the 2003 Conference of the North*

American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003.
[18] Mynatt, Elizabeth, and Joe Tullio. "Inferring calendar event attendance." *Proceedings of the 6th international conference on Intelligent user interfaces*. ACM, 2001.
[19] Geopy: A Geocoding Toolbox for Python, https://pypi.python.org/pypi/geopy/
[20] Shapefile. (n.d.). Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Shapefile. Accessed Nov. 03, 2016.
[21] John W. Eaton, David Bateman, Søren Hauberg, Rik Wehbring (2014). GNU Octave version 4.0.3 manual: a high-level interactive language for numerical computations. CreateSpace Independent Publishing Platform. ISBN 1441413006, URL http://www.gnu.org/software/octave/doc/interpreter/
[22] Kanungo, Tapas, et al. "An efficient k-means clustering algorithm: Analysis and implementation." *IEEE transactions on pattern analysis and machine intelligence* 24.7 (2002): 881-892.
[23] Comaniciu, Dorin, and Peter Meer. "Mean shift: A robust approach toward feature space analysis." *IEEE Transactions on pattern analysis and machine intelligence* 24.5 (2002): 603-619.
[24] Cheng, Yizong. "Mean shift, mode seeking, and clustering." *IEEE transactions on pattern analysis and machine intelligence* 17.8 (1995): 790-799.
[25] Pelleg, Dan, and Andrew Moore. "Accelerating exact k-means algorithms with geometric reasoning." *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999.
[26] Pelleg, Dan, and Andrew W. Moore. "X-means: Extending K-means with Efficient Estimation of the Number of Clusters." *ICML*. Vol. 1. 2000.
[27] Maps - Navigation & Transit. Android version 9.40.2 [Google Inc.] 1600 Amphitheatre Parkway, Mountain View 94043 (Nov 2016) https://play.google.com/store/apps/details?id=com.google.android.apps.maps
[28] Richardson, Mark. "Principal component analysis." *URL: http://people. maths. ox. ac. uk/richardsonm/SignalProcPCA. pdf (last access: 3.5. 2013). Aleš Hladnik Dr., Ass. Prof., Chair of Information and Graphic Arts Technology, Faculty of Natural Sciences and Engineering, University of Ljubljana, Slovenia ales. hladnik@ ntf. uni-lj. si* (2009).
[29] Fortmann-Roe, Scott. "Understanding the bias-variance tradeoff." (2012).