Visual Analysis of Predictive Suffix Trees for Discovering Movement Patterns and Behaviors

Antonio José M. Leite Júnior
Virtual University Institute
Federal University of Ceará - Fortaleza, Brazil
Email: melojr@virtual.ufc.br

Emanuele Santos, Creto A. Vidal, Jose A. F. de Macêdo
Department of Computing
Federal University of Ceará - Fortaleza, Brazil
Email: {emanuele, cvidal, jose.macedo}@dc.ufc.br

***Abstract*—The use of GPS-equipped devices has allowed generating and storing data related to massive amounts of moving objects, promoting many solutions to movement prediction problems. Movement prediction became essential to perform tasks in several areas ranging from analysis of the popularity of geographic regions; and management of traffic and transportation; to recommendations in location-based social networks. To explore this type of data is a complex task because one must deal simultaneously with space, time and probability. In this work, we apply the branching time concept to visual analytics, proposing an approach that supports movement prediction using Probabilistic Suffix Trees. We try to substitute the traditional evaluation method, based on reading texts, by an interactive visual solution. To validate the proposed solution, we developed and tested a visualization tool using a real dataset. It assisted experts to quickly identify where a person lives, where she works and to recognize some of her movement patterns and probable behaviors.**

I. Introduction

In recent years, the widespread use of devices with built-in geolocation receivers made it possible to record a huge amount of trajectory data generated by moving objects. As a result, applications were devised to consume that type of data. The applications that deal with movement prediction are particularly important to solve problems ranging from analyzing the popularity of geographic regions [1], traffic and transportation management [2] to recommendations in location-based social networks [3].

However, analyzing datasets of individual trajectories for movement prediction is a complex task because one needs to handle, simultaneously, space, time and related probabilities [4]. The available visualization techniques usually fall short of combining those components properly with geographical maps [5].

In a recent work, Rocha et al. [6] proposed a predictive model to forecast when an object will leave its current location to reach the next predicted one. They used a special data structure – Probabilistic Suffix Tree (PST) [7], [8] – that is capable of representing space, time and probability at the same time.

In this paper, based on PSTs and the predictive model presented by Rocha et al. [6], we propose a visual analytics solution that helps users to solve complex problems, like investigating where and when a person probably lives, works or studies, for example. This work is relevant, because, without PSTs, it is hard to investigate in a large trajectory dataset where and when the main occurrences are and which trajectories need to be observed in more detail [9].

Although there are some techniques for visual analysis of different data structures, none have been specifically designed to deal with PSTs. The main contribution of this work is trying to substitute the traditional PST evaluation method, based on reading texts to identify space, time and probabilities (Figure 6-a), by an interactive visual solution. Therefore, the proposed solution focuses on visualizing specific data and the relationship between them, considering user interaction and the complexity of extracting, storing and processing the PSTs data.

In our approach, we apply the branching time concept, which allows a probabilistic event to be a branching point from which two or more events originate. That endows our interactive data visualization technique with the capability of supporting the discovery of movement patterns (Figure 1). The implemented solution allows analysts to explore the dynamics of space-time combination (space, time and space versus time relations) and evaluate their associated probabilities simultaneously.

The rest of the paper is organized as follows. Section II discusses related works. Basic concepts are presented in Section III. Section IV describes the proposed solution. An application developed using the proposed solution is shown in Section V and validated in Section VI. Section VII shows general observations and Section VIII concludes the paper.

II. Related Work

The focus of our work is to help users to analyze PST data to solve problems involving space, time and probability simultaneously. To the best of our knowledge, there is no work in the literature proposing a visualization analysis technique that exploits the probabilistic aspect inherent in such problems. However, some visualization techniques also try to deal with the interpretation of spatiotemporal events.

The most common approaches consist in graphically representing trajectory sets, considering space and time simultaneously, using different data structures. Crnovrsanin et al. [10] propose a two-dimensional representation that associates a time graph with a map, displaying distances to selected places. Because users can see more than one distance at once, authors

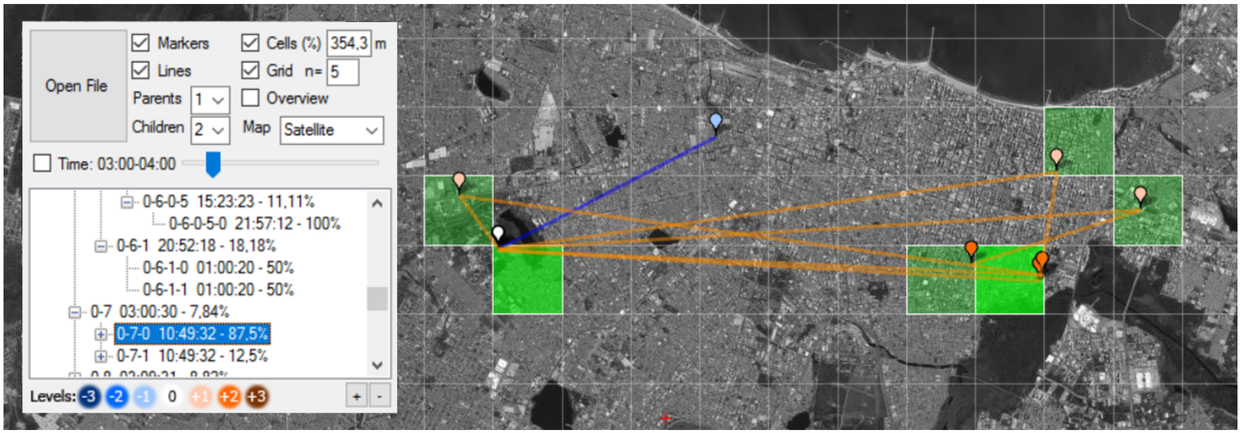


Fig. 1. Visualizing a Predictive Suffix Tree (PST). The selected node (in white) and its related nodes are displayed using colored markers. Ancestors are shown in the shade of blue and descendants in the shade of orange. The color saturation indicates the level of the node concerning the selected node. Colored edges are drawn to show node relations. The nodes are displayed on a grid where each cell is colored in green and proportionally saturated by the number of markers present in that cell. The user can control all the visualization parameters by using the control panel on the left.

use proximity PCA (Principal Component Analysis) to avoid path overlapping. Spretke et al. [11] use color segments to display many trajectories on a map simultaneously. With that, although it is possible to analyze each trajectory individually, their system shows certain attributes in aggregated fashion. The overlapping of trajectories is not fully avoided which makes visualization sometimes cumbersome. Kincaid and Lam [12] and Matkovic et al. [13] combine time graph and maps to show multiple trajectories using particular time band displays to avoid overplotting. Wang et al. [14] analyze traffic jams associating maps, visual propagation graphs, and table-like pixel-based visualization simultaneously. Scellato et al. [15] use GPS data frequency maps to identify significant locations on geographic maps.

Some approaches focus more on interactivity. Ferreira et al. [16] propose the visual query (origin-destination) maps, which support interaction and show hidden details using overlaid heat maps. Zeng et al. [17] present the, so called, interchange *circos* diagram to show interchange patterns in interactive applications.

But trajectory clutter on a 2D map makes visual analysis difficult, and some authors try to address that issue by visualizing trajectories in 3D space. Kraak [18] presents the space-time cube model, in which space is represented by two geographical dimensions defining a plane map, and the third dimension corresponds to time. A trajectory is displayed inside the cube, combining time and space components. In the same way, the trajectory wall model extends the idea of a space-time cube for the simultaneous presentation of several individual paths [19]. Many trajectories are juxtaposed vertically over a map on a three-dimensional space. An interactive interface allows the analyst to modify the order in which the trajectories are displayed; to select and view some of the trajectories' specific attributes such as direction, and maximum instantaneous or average velocities, for example.

For the problem we are addressing, a Probabilistic Suffix Tree (PST) defines a set of possible trajectories. Therefore, it is not practical to try to define beforehand the whole trajectory to follow, just to be able to apply the traditional trajectory visualization techniques. Displaying the entire set of possible trajectories of a PST at once would certainly make visual analysis and comprehension almost impossible [20].

Moreover, a PST considers not only space and time components but also their associated probabilities. Therefore, it is necessary to define a solution that allows the accurate visualization of the three associated components (time, space and probability) and the exploration of all related data using interactive tools.

In this way, some of the cited techniques can be adapted to the visual analysis of PSTs. However, they may suffer from problems of visualization of specific data and the relationship between them, low interactivity or even performance, due to the complexity of storing and processing the PSTs.

III. Basic Concepts

In this section, we briefly discuss the three core concepts involved in our work: Visual analytics, Movement prediction and Branching time.

A. Visual Analytics

Originally, Adrienko et al. [21] consider geovisual analytics as the computer-assisted human activity where useful results are obtained through the processing of spatiotemporal data into generated images from which knowledge is extracted. In a parallel and more broadly manner, Keim et al. [22] define visual analytics as a set of methods, technologies, and practices that combine data processing capabilities of humans and machines. Visual analytics is different from other investigation approaches because it relies on the interactive visual representations that amplify the human capacity of detecting patterns, of establishing relationships and of creating inferences [23]. Its main goal, therefore, is to allow analysts to go beyond their natural interpretation capabilities. Visual analysis is supposed to be more efficient than automatic procedures because an analyst uses the interactive visualization resource to elaborate new questions and search for new answers according to changes in observed events and contexts [23]. Obviously, visual analytics can also include automatic tools that help users to do a more efficient investigation.

B. Movement Prediction

A sequence of space-time data, that is, specific positions in space established at certain time points, defines a moving object's trajectory. When analyzing the trajectories of moving individuals, it is often possible to detect space-time data that indicate some repetitive behavior, known as trajectory patterns. Those patterns can support location-based popularity analysis [24], travel route planning [25], and path evaluation for traffic optimization [26], for example.

The goal of movement prediction is described, in general terms, as the challenge of predicting the next location of a moving object based on historical spatiotemporal data encompassing its trajectory patterns [6].

Through the analysis of trajectory patterns, it is possible to predict one or more future locations of a moving object according to its latest move changes. Those predictions are accomplished with the use of the so-called movement functions, from which the speed and direction of recent changes are computed. However, that type of calculation usually is useful only for short-term predictions, being ineffective for long-term predictions [27].

A more reliable alternative for the long-term prediction of a future location of a moving object is based on the analysis of that object's most visited places. This analysis can create inference rules [28] or well-defined sequences of trajectory patterns [29] associating the most frequented areas. This approach, moreover, tends to generate greater precision for longer time intervals [30]. However, both the inference rules and the pattern sequences may be biased if associated temporal data are not considered [31].

For predicting motion with greater precision, one can use a PST, which stores spatial and temporal data simultaneously [6]. Such solution establishes direct inference relationships between recorded space-time data, and, simultaneously, stores and represents a significant number of movement patterns, directly associating trajectory patterns with probability data [4]. Through a quick inspection of a PST's different levels, one clearly sees the information about the object's movement. Spatial, temporal and spatiotemporal data are associated with probabilities in a very compact and intuitive way, which allows them to be used quickly and directly.

Figure 2 illustrates a single PST that represents possible daily movements of a person, considering space, time and probabilities. In this example, the individual stays at home until 8:30 a.m. Later, there is a 50% probability that he will be in school at 8:45 a.m. and a 50% probability that he will be at work at 9:00 a.m.

C. Branching Time

Originally, the term branching time was related to the classification of narrative structures that were present in some literary works, when a given event could originate two or more distinct events. Later, considering communication and media evolution, the branching time concept was applied to movies, theater plays and video games [32]. However, nowadays, branching time has become an important concept for evaluating scenarios, and it is critical for planning activities.

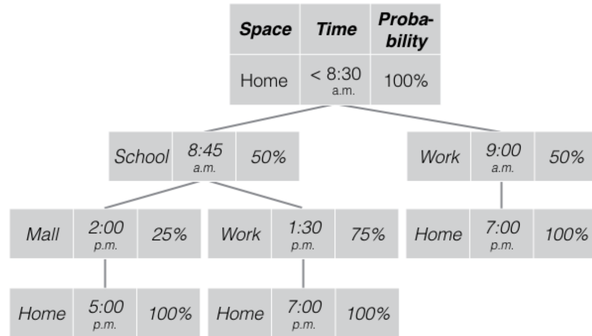


Fig. 2. Graphical depiction of a Probabilistic Suffix Tree (PST). Each node contains the place, the time and the probability of the person to be in that place at that time.

Planning may use scenarios where sequences of actions are foreseen, often with multiple alternatives, depending on future decisions. Branching time is regularly used to study the implications of a current action on the future decisions or future reactions to external influences (Figure 3). The scenarios represent possible future states, only one of which will be actually reached [33]. So, branching time can often be used to analyze decision-making and resulting actions, in which multiple outcomes represent possible scenarios to be interpreted and investigated [5].

IV. Proposed Solution

Identifying significant occurrences in trajectory trees such as PSTs is a complicated task, because of the complexity of aggregating space, time, and attributes (PST probabilities) simultaneously [9]. A PST defines a complete set of possible trajectories, and, therefore, it is not practical to display the whole set of possible trajectories at once, since the occlusion of space-time data will certainly hinder visualization, interactive analysis, and comprehension of the represented information [20]. Moreover, considering a visual analysis approach, choosing which trajectories to investigate is tough

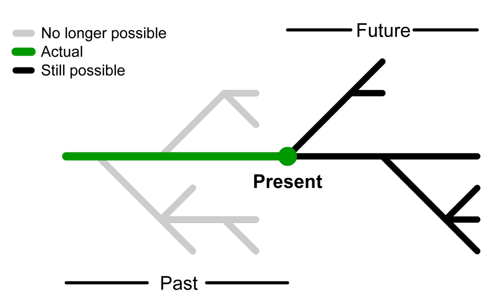


Fig. 3. Branching time graphical representation.

because of the necessity to simultaneously respect the multiple and complex PST node and edge relations.

So, for PST interactive graphical representation, we propose to apply the branching time concept in association with a two-dimensional model, in which every PST node can be a branching point of one or more of its child nodes. Due to the space-time nature of that model, it is possible to consider that each node is a distinct temporal event in a particular location. Thus, such association can be used to describe the behavior of a moving object considering the whole set of its possible trajectories.

Therefore, the underlying assumptions of the visual analytics for discovering movement patterns, proposed in this work, are:

1) Each PST node represents a temporal and individual event, and its ancestors and descendants mean respectively previous and later occurrences;
2) Each PST node is properly positioned in its respective geographic coordinates on a map. Edges should be visible and clearly show both its descending and ascending nodes;
3) It is possible to visualize the entire space-time model as well as its respective probability values (current probabilities associated with each node) independently or together;
4) It is possible to examine distinct portions of the space-time model. So, specialists can view the data that define time intervals, geographical locations, or individual probability values.

Despite those design choices, both the tree size and the map size are challenging issues in a visual analytics process. If the number of PST nodes (or other graphical elements) is too large, it may impact the hardware and software performances. Moreover, even if it is possible to display the entire graphical representation, visibility and usability problems may occur, such as visual occlusion or increased navigation complexity [34]. Thus, to avoid such problems, we adopted the direct support for the following interactive features [34]:

- **Panning**, change the visualization both in space and in time, maintaining space-time correlation;
- **Zooming**, zoom in or out in space and in time, automatically displaying results of such operations;
- **Filtering**, choose a set of PST nodes (time) or a particular map portion (space), selecting one or more trajectories;
- **Clustering**, group data by some characteristics, such as the number of events represented in a certain map portion, for example.

Those interactive features allow better use of available processing and storage resources because temporal and spatial partitioning ignore superfluous data, limiting the computer access to what is needed at the moment. Also, by reducing the number of displayed elements, specialists can avoid visual cluttering, and perform a more focused evaluation, observing events in more detail [35].

Moreover, all those features support the discovery of distinct probable behaviors represented in the PST. For example, an analyst can evaluate the PST's nodes representing a person's home, and respective map position, to check her usual departure time to work. Then, considering the individuals workplace nodes, the analyst can verify whether the person returns home at noon or goes out for lunch at a restaurant. Also, considering the probabilities stored in the PST, the analyst can deepen the observation, identifying movement patterns and consequently foreseeing events. So, it is possible to differentiate daily routine at workdays and during the weekend, for example.

V. Visualization Tool

We developed a visualization tool (Figure 1) to test the proposed solution. It simultaneously relates the three analysis components involved (space, time and probability), allowing an expert user to explore the PST data for identifying events and making discoveries.

The tool was implemented in C# and uses data produced by Eaí? (http://www.appeai.com), a mobile application that provides information (daily menu, food quality, and so forth) on the Federal University of Ceará's cafeteria to approximately 30,000 users in the city of Fortaleza, Brazil.

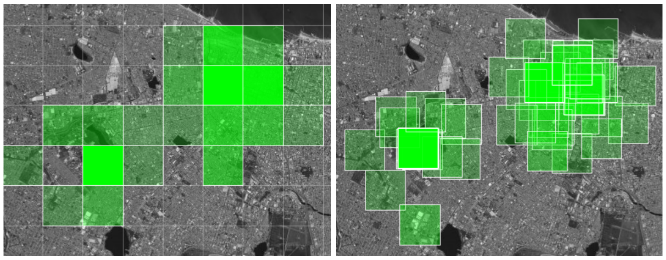
The users' GPS positions and their associated time acquisition captured by the mobile devices are processed and organized into PSTs, which are stored in JSON files. For privacy reasons, the users are not identified.

The developed tool reads the PSTs, extracts data (geographic coordinates, time stamps, and probabilities), represents data nodes inside a tree list, preserving tree levels, and also draws the nodes on a map. Analysts can select a PST node by clicking directly on the respective tree list line. When this occurs, the application displays a marker on a map according to the geographic coordinates of the selected node. If it has ancestors or descendants, their markers can also be shown on the map, considering their respective PST levels.

Markers indicating parent and child nodes are colored using a complementary chromatic scale, ranging from navy blue, on the third level parent node, to dark orange, on the third level child node (Levels scale in Figure 1). Concerning to ordered time, this procedure allows the correct identification of situations, including the occurrences that happen before and after the one selected in the tree list line. To simplify the analysis process, it is also possible to visualize semitransparent edges that connect the selected node to its parent and child nodes, following the same chromatic scale (Figure 1). So, using the branching time approach, it is easy to see the spatial and temporal ancestors and descendants of the node.

The reverse process is also possible, and when the analyst clicks on a map's marker, all related data (space, time and probability) is presented directly on the map, and the respective tree list line is automatically selected.

Optionally, analysts can investigate time intervals using a slider bar, selecting a specific time. When it happens, the application shows on the map all markers whose time is inside the respective time interval defined.



(a) (b)

Fig. 4. Visualization of presence information using cells. (a) The grid is enabled: green tones are proportional to the number of markers present in each cell. (b) The grid is disabled: color represents an individual nodes' probabilities.

To explore space segments, analysts can cluster data, enabling and defining the size of a visual grid on the map. Therefore, the application considers the squared cells whose green tones are proportional to the number of markers present in each cell (Figure 4-a). Analysts can control green tone calculation setting a max limit for the counter of markers. If grid map is disabled, the cell color represents individual node probabilities (Figure 4-b). Notice that, by using a non-smooth probability distribution function, we can emphasize abrupt changes in cell presence. Specialists can use cells to quickly identify points of interest (POIs) and investigate occurrences related to them.

It is also possible to display all the data stored in a PST directly on the map. It presents all possible trajectories and analysts can individually show or hide markers, edges, and cells. Considering all the listed features, Figure 5 summarizes the architecture of the developed application.

VI. Solution Validation

Since the nature of the PST application was very particular, we decided to postpone the use of quantitative methods for validating the developed visualization tool [36] and adopted a qualitative (field observation) approach [37]. According to

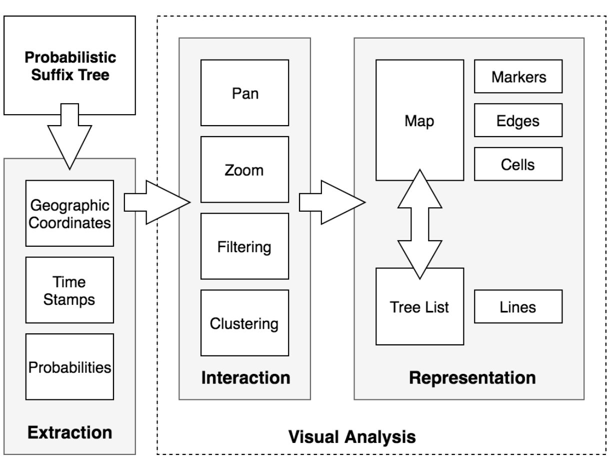


Fig. 5. The system architecture of the developed application.

this methodology, we asked specialist users to perform certain predefined inspection tasks at their workplaces (the *field*). For that, we selected two experts in movement prediction:

- **X1**, male, undergraduate student in Computer Science. He is in charge of the data management of the Eaí? mobile application and frequently identifies some PST nodes using maps;
- **X2**, male, master's degree in Computer Science. He works as a senior developer, and he reads JSON PST files regularly to find events but does not usually check coordinates on the map.

To avoid evaluation problems like casual conversations during the field observation, the person applying the tests did not know the specialists [38].

We analyzed many different JSON files to elaborate inspection tasks and selected three PSTs referring to real data collected on weekdays between 8/27/2015 and 10/1/2015. In this period, there was a strike at the University, and many of the classes were suspended. That was particularly interesting for the test because it changed the University's routine, to which the tested specialists were already familiar, and allowed for the identification of new movement patterns on the three selected PSTs, making the test tasks more stimulating and challenging.

During the test, the participants were asked to follow a specially prepared task script that contained questions to be answered while they used the tool. So, for each selected PST, they had to answer the following questions:

1) Where does the person live?
2) In which course is the person enrolled?
3) Where does the person work?
4) When does the person go to the University's cafeteria?

No previous observation on the three selected files, including period, was provided to the specialists. And, due to data privacy and probabilistic PST characteristics, we were not interested in formal problem resolution. The questions' main purpose was to help the experts to elaborate new queries (hypotheses) and propose solutions (prove or refute hypotheses), according to visual analytics principles [22]. So, specialists could manage available data to suggest new possibilities of information interpretation, extrapolating the original predefined contexts.

We contacted the selected experts separately, explained the research goals, and briefly presented the visual analytics tool they would use for one week in their workplaces. We provided a laptop (Apple MacBook Pro 17" Early 2011 with Intel i7 processor, 8 GB RAM, 240 GB SSD running Microsoft Windows 10 and equipped with an external mouse) containing the developed application and the three selected JSON PST files.

The field observation started with the specialists signing terms of free and informed consent. Then, we began to take notes and to record the session using a mobile phone. Initially, the analysts freely used the developed application so that they could recall its primary functions. After that, they carried out the predefined tasks over the three selected PSTs.

A. Analysis of the first PST

The first PST concerns a person who did not go to the University during the considered period. We observed that experts initially adopted the traditional methodology [4] to investigate PSTs to answer the questions. But after understanding the visualization tool features, they worked in a more efficient way.

Based on the PST data overview (by displaying all edges at once), both specialists quickly identified where the person probably lived. Their evaluation mainly consisted of observing the most important points of origin and the higher frequencies of movement. They searched for the greater amounts of overlapping lines in the data overview. Whenever necessary, analyst identified regions of interest using cells to cluster data visually [39] (Figure 6-b).

After identifying the main POIs on the map, the specialists initiated a more detailed exploration by also interacting with the tree list in the control panel, observing time, presence probabilities and places on the map simultaneously. Controlled investigation of parent and child nodes, at various levels, helped them to detect movement patterns, analyzing positions, time and probabilities.

X2 noted that the person remained in a particular place late at night until dawn, and concluded that the place has a high probability of being the person's home. X1 also used the same features to recognize where the person lived. Besides, X1 noticed that the individual frequently went to a church and a few more places in the neighborhood (Figure 6-c).Both specialists observed that the individual did not visit the University during that period.

X1 assumed the person did not work because X1 did not find a significant movement pattern on the map. X2, on the other hand, analyzed the tree list more deeply and detected the person was constantly going to a particular location. He, then, supposed that to be the person's workplace. However, by not paying much attention to the map, he did not realize the designated area was the church mentioned by X1. Due to data privacy, it is not possible to know whether the person worked at the church or just visited it during that period. Nevertheless, we noticed that the specialists accurately distinguished possible movement patterns and made valid inferences about the person's behaviors.

B. Analysis of the second PST

The next step consisted in analyzing the second PST. Again, starting with the data overview and later interacting with the tree list and the map simultaneously, the specialists quickly identified the person's possible home.

X1 stated that the second person attended courses at the University during the considered period and that he probably studied Biology at night. Such information, however, initially seemed not to be true because the indicated course runs during the daytime. However, due to the faculty strike, it is possible that the person went to the University at night. X2, on the other hand, said that the individual probably studied Computer Science or another course offered at the Science Center, which includes the Biology course. However, X2 did not indicate at what time the person was at the University.

X1 noticed a repetitive movement pattern between 2 p.m. and 5 p.m. connecting a particular region with the University. So, X1 reported that location to be a probable working place of the individual. X2 also found the same pattern, but built a complete narrative, stating that the person traveled from home or University to the possible workplace indicated by X1, arrived there late in the morning, possibly for lunch, and continued there until late in the afternoon, and then returned home.

According to X1, the person did not stop by the University's cafeteria during that period, since he identified no movement pattern on the map that would lead to such a conclusion. X2, on the other hand, intensified the search in the tree list and concluded that the person might have had lunch a few days at the University's cafeteria, generally arriving between 11:45 a.m. and noon.

C. Analysis of the third PST

In this step, the specialists explored the third PST. Once again they used data overview and connections between the tree list and the map to distinguish a few POIs. X1 indicated the person's possible home and pointed out bus stops frequently used for trips to and from the University. X2 identified the equivalent probable home and also noted the individual's recurring movement to the same two points, one near home and the other near the University. However, because X2 did not look for those points on the map, he did not realize they were bus stops and said that the points could be the individual's family or friend homes.

X1 inferred that the person probably attended Computer Science in the afternoon. X2 reached the same conclusion and identified a corresponding movement pattern, where the person moved early in the evening to a building in the University's neighborhood twice a week. The individual spent about two hours there.

X1 did not detect the person's workplace, but X2 indicated that the individual probably worked at that building near the University.

As X1 did not observe a distinct movement pattern, he suggested the person probably did not stop by the University's cafeteria during that period. X2, on the contrary, reported that the individual sometimes had lunch in the University's cafeteria, arriving there between noon and 12:10 p.m.

D. Test Evaluation

After testing, we transcribed all the meetings' audios, recorded with the mobile phone, adding field notes and emphasizing the specialists' quotes. We evaluated the results using the content analysis methodology [40], by systematically organizing data into a structured format (time required to execute each task, identification of events, significant specialists behaviors, notable quotes, and so forth).

(a) (b) (c)

Fig. 6. A user discovering an interesting movement pattern by (a) the traditional method, examining a PST stored as a JSON File; (b) using the developed visualization tool, first performing a data overview to identify the most frequent places (white markers and green cells) and their relationships (green edges), and (c) later zooming in to perform a more detailed exploration of the dashed region in (b).

The total operation time was about 12 minutes for each specialist. Both experts considered the developed application to be pleasant and easy to use, providing an adequate way of analyzing PSTs.

Qualitatively, the developed application is satisfactory for visual analytics. In general, the specialists found distinct movement patterns using map-tree list association and made inferences about a person's probable behavior easily.

It can be seen, however, that both experts presented different analysis profiles: X1 concentrated more on the map, while X2 concentrated more on the tree list. X1, for example, identified the first person's visits to a church and detected the bus stops that the third individual probably used. Likewise, X2 suggested sporadic lunches at the cafeteria for the second and third individuals, and he identified complex movement patterns, specifying possible workplaces for all the subjects.

VII. General Observations

We noticed that both specialists started analyzing PSTs in the same way, adopting the traditional methodology: loading the JSON file and inspecting each PST level, trying to isolate data (geographic coordinates, time stamps, and probabilities) (Figure 6-a), identifying POIs and attempting to discover the relationships among them [4].

Using the developed application, specialists also started opening each tree list line, trying to understand data. However, as they perceived available tools and solved the first problems, they adopted a new work methodology. First, to investigate a problem, they used the data overview to identify POIs. Next, they interacted with the tree list to explore data in details, simultaneously hiding or showing markers and edges on the map, making it easy to decide which trajectories should be observed in more detail.

When experts needed to consider probabilities, they used cells to evaluate individual values or cluster data, changing cell size, when needed, to analyze more specific regions. Later they selected markers directly on the map to identify the respective tree list lines. So, for deeper analysis, examining the traversed trajectories, the specialists defined different parent and child node levels. Finally, the analysts could find solutions or formulate new questions to investigate.

We observed this approach allowed experts to explore the dynamics of space-time combination freely and to evaluate their associated probabilities simultaneously. But considering the test results and the observed methodology they used, we can conclude that the analysts profile influences the way they search for the solution and how they propose new problems to be solved, regarding the tree list or the map.

VIII. Conclusions

The present work proposed applying the branching time concept in visual analytics to simplify finding movement patterns and possible person's behavior, based on PSTs, a structure capable of simultaneously representing space, time, and probability. We developed a visualization tool and qualitatively validated it using field observation test. In general, the obtained results demonstrated the feasibility of using the solution to identify movement patterns and a person's probable behaviors. In fact, the tool provides enough freedom for the analyst to explore the represented data in many different ways. We noticed that a particular methodology of use emerged from observed interactions, but we also perceived that the analysts' profiles and their preferences for using either the tree list or the map might influence the way they reach their conclusions.

As future work, we intend to conduct more tests, trying to evaluate precisely how an analyst develops her methodology of use, considering how she balances the simultaneous control of the tree list and the map. Despite, the particular nature of the PST application, we plan to evaluate how non-experts can solve problems using the proposed interactive visual analytics, examining how the use of the tool compares with the traditional investigation process. We also plan to include machine learning techniques to automatically identify the most visited locations, trying to simplify the initial search for POIs. Another future work is to improve the layout of the GUI application, using techniques such as edge-bundling to reduce clutter in tree list and providing annotation features to allow analysts to register and substantiate their findings.

Finally, we plan to work on other datasets to apply the proposed visual analytics solution to other situations, like police investigation and crime prediction, health delivery, and public transport systems.

Thus, we expect to contribute to simplifying PST analysis, creating more efficient tools to assist the evaluation of data that relates space, time and probability.

ACKNOWLEDGMENT

Our work has been partially funded by the MCTI/CNPQ/Universal grant number 459448/2014-5.

REFERENCES

[1] Y.-H. Lin, C.-H. Lai, and P.-R. Lei, "Mining top-k relevant stay regions from historical trajectories," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2014, pp. 293–304.

[2] A. Asahara, K. Maruyama, A. Sato, and K. Seto, "Pedestrian-movement prediction based on mixed markov-chain model," in *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2011, pp. 25–33.

[3] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: a survey," *GeoInformatica*, vol. 19, no. 3, pp. 525–565, 2015.

[4] P.-R. Lei, S.-C. Li, and W.-C. Peng, "Qs-stt: Quadsection clustering and spatial-temporal trajectory model for location prediction," *Distributed and Parallel Databases*, vol. 31, no. 2, pp. 231–258, 2013.

[5] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visual methods for analyzing time-oriented data," *IEEE transactions on visualization and computer graphics*, vol. 14, no. 1, pp. 47–60, 2008.

[6] C. L. Rocha, I. R. Brilhante, F. Lettich, J. A. F. De Macedo, A. Raffaetà, R. Andrade, and S. Orlando, "Tpred: a spatio-temporal location predictor framework," in *Proceedings of the 20th International Database Engineering & Applications Symposium*. ACM, 2016, pp. 34–42.

[7] D. Ron, Y. Singer, and N. Tishby, "Learning probabilistic automata with variable memory length," in *Proceedings of the seventh annual conference on Computational learning theory*. ACM, 1994, pp. 35–46.

[8] G. Bejerano and G. Yona, "Modeling protein families using probabilistic suffix trees," in *Proceedings of the third annual international conference on Computational molecular biology*. ACM, 1999, pp. 15–24.

[9] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski, "Space, time and visual analytics," *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1577–1600, 2010.

[10] T. Crnovrsanin, C. Muelder, C. Correa, and K.-L. Ma, "Proximity-based visualization of movement trace data," in *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*. IEEE, 2009, pp. 11–18.

[11] D. Spretke, P. Bak, H. Janetzko, B. Kranstauber, F. Mansmann, and S. Davidson, "Exploration through enrichment: a visual analytics approach for animal movement," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2011, pp. 421–424.

[12] R. Kincaid and H. Lam, "Line graph explorer: scalable display of line graphs using focus+ context," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2006, pp. 404–411.

[13] K. Matkovic, D. Gracanin, Z. Konyha, and H. Hauser, "Color linesview: Anapproach tovisualization of families of function graphs," in *Information Visualization, 2007. IV'07. 11th International Conference*. IEEE, 2007, pp. 59–64.

[14] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168, 2013.

[15] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, "Nextplace: a spatio-temporal prediction framework for pervasive systems," in *International Conference on Pervasive Computing*. Springer, 2011, pp. 152–169.

[16] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.

[17] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, "Visualizing interchange patterns in massive movement data," in *Computer Graphics Forum*, vol. 32, no. 3pt3. Wiley Online Library, 2013, pp. 271–280.

[18] M.-J. Kraak, "The space-time cube revisited from a geovisualization perspective," in *Proc. 21st International Cartographic Conference*, 2003, pp. 1988–1996.

[19] G. Andrienko, N. Andrienko, H. Schumann, and C. Tominski, "Visualization of trajectory attributes in space–time cube and trajectory wall," in *Cartography from Pole to Pole*. Springer, 2014, pp. 157–163.

[20] H. Purchase, "Which aesthetic has the greatest effect on human understanding?" in *International Symposium on Graph Drawing*. Springer, 1997, pp. 248–261.

[21] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M.-J. Kraak, A. MacEachren, and S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," *International Journal of Geographical Information Science*, vol. 21, no. 8, pp. 839–857, 2007.

[22] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," in *Information visualization*. Springer, 2008, pp. 154–175.

[23] S. I. Fabrikant and A. Lobben, "Introduction: Cognitive issues in geographic information visualization," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 44, no. 3, pp. 139–143, 2009.

[24] R. Montoliu and D. Gatica-Perez, "Discovering human places of interest from multimodal mobile phone data," in *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2010, p. 12.

[25] K. Zheng, G. Trajcevski, X. Zhou, and P. Scheuermann, "Probabilistic range queries for uncertain trajectories on road networks," in *Proceedings of the 14th International Conference on Extending Database Technology*. ACM, 2011, pp. 283–294.

[26] C.-T. Lu, P.-R. Lei, W.-C. Peng, and J. Su, "A framework of mining semantic regions from trajectories," in *International Conference on Database Systems for Advanced Applications*. Springer, 2011, pp. 193–207.

[27] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*. Ieee, 2008, pp. 70–79.

[28] M. Morzy, "Mining frequent trajectories of moving objects for location prediction," in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*. Springer, 2007, pp. 667–680.

[29] H. Jeung, H. T. Shen, and X. Zhou, "Mining trajectory patterns using hidden markov models," in *International Conference on Data Warehousing and Knowledge Discovery*. Springer, 2007, pp. 470–480.

[30] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 637–646.

[31] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[32] M. Martínez, "Travelling in branching time," *Disputatio*, vol. 4, no. 31, pp. 271–287, 2011.

[33] A. U. Frank, "Different types of times in gis," *Spatial and temporal reasoning in geographic information systems*, pp. 40–62, 1998.

[34] I. Herman, G. Melançon, and M. S. Marshall, "Graph visualization and navigation in information visualization: A survey," *IEEE Transactions on visualization and computer graphics*, vol. 6, no. 1, pp. 24–43, 2000.

[35] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl, "Trajectorylenses–a set-based filtering and exploration technique for long-term trajectory data," in *Computer Graphics Forum*, vol. 32, no. 3pt4. Wiley Online Library, 2013, pp. 451–460.

[36] J. Nielsen, *Usability engineering*. Elsevier, 1994.

[37] D. E. Polkinghorne, "Language and meaning: Data collection in qualitative research." *Journal of counseling psychology*, vol. 52, no. 2, p. 137, 2005.

[38] A. Holzinger, "Usability engineering methods for software developers," *Communications of the ACM*, vol. 48, no. 1, pp. 71–74, 2005.

[39] A. O. Artero, M. C. F. de Oliveira, and H. Levkowitz, "Uncovering clusters in crowded parallel coordinates visualizations," in *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On*. IEEE, 2004, pp. 81–88.

[40] P. Liamputtong and D. Ezzy, *Qualitative Research Methods*, ser. Qualitative Research Methods. Oxford University Press, 2005.