Using Ontologies for Proposing Adequate Geovisual Analytics Solutions in the Analysis of Trajectories

Gabriel Vatin, Aldo Napoli

To cite this version:
Using Ontologies for Proposing Adequate Geovisual Analytics Solutions in the Analysis of Trajectories

Gabriel Vatin, Aldo Napoli
MINES ParisTech
CRC - Centre for research on Risks and Crises
Sophia Antipolis, France
{gabriel.vatin, aldo.napoli}@mines-paristech.fr

Abstract— This paper presents an original approach for supporting the use of geovisual analytics solutions. Many models have been proposed to characterize information visualization methods, but few have been integrated to an intelligent process for supporting user in geo-information usage. Moreover, several new solutions are continuously proposed by research, but few of them are really used in operational world. For instance, the maritime surveillance systems could gain much more identification capabilities of ship behaviors with adequate geovisual analytics solutions. Therefore, we investigated the use of geovisual methods for the analysis of mobility data, such as ship trajectories. We propose a knowledge-based system using ontologies and rules. These allow modeling the domain of geovisual analytics solutions, and their capacities in the exploration and the analysis of trajectories. This system would be used to support users in geovisual analytics of movement, based on their context of use.

Keywords - geovisual analytics; ontologies; rules; maritime domain awareness; analysis support

I. INTRODUCTION

This last decade, visualization of information has been a major issue. This issue has evolved very rapidly along new computation technologies, and the creation of many new captors for acquiring data. The use of “smart” technologies for geo-locating people, cars, ships, or airplanes, has motivated a breakthrough in the amount of reachable information and in the speed for acquiring movement data. This has resulted in the development of massive geographic databases. For instance, the use of RADAR systems and AIS (Automatic Identification System) for detecting ship locations at sea allows 24/7 monitoring of maritime traffic, within the area covered by the antennas.

The analysis of these data can be led a posteriori, for historical analysis of movement, in near real-time for surveillance purposes, or a fortiori for predictive studies. In each of these cases, the visualization of geo-information, known as geovisualization (GeoViz), allows using several methods for exploring, analyzing or predicting movement information, with the use of visual and highly interactive interfaces [1]. When using several representations of the data into dynamically linked views, information visualization supports the analytical process led by users: this led to the definition of Visual Analytics [2], [3]. When applied to geo-information, this concept is also known as Geovisual Analytics (GeoVA). This use of visualization has a specific focus on research problems dealing with space and time information, both including geographic and non-geographic data sources.

In this paper, we call GeoVA environments the interfaces that allow a simultaneous use of different visualization spaces for the visual analysis of spatial, temporal and/or semantic information. This is based on the principle given by Peuquet [4], for the representation of temporal dynamics. These are environments that make the exploration and analysis of information possible, through various linked views. These linked views improve the comprehension of potential patterns, or the discovery of new knowledge.

GeoVA environments for the analysis of movement and trajectories have been a major field of research for the visualization and cartography communities, as we can see in recent publications and conferences [5]–[7]. However, these numerous propositions for new visualization environments raise major questions about their usage. In front of such productive communities, how can a single user decide of the visualization to use, while facing a very specific analytical problem? These many visualization environments have been developed for specific data sets, or certain analytical questions, and for identified users. Today, a major issue in research is not to propose new solutions for the visualization of movement data, but to document all possible solutions and their evaluations, in order to identify:

- **What can / cannot we visualize with each?**
- **How can / cannot we interact with each?**
- **Who can / cannot use each?**

As Fabrikant has identified the problem, there is still an important need of proposing task-oriented taxonomies for geovisualization [8]. This would allow developing a new type of meta-model of GeoVA environments, for identifying how analytics can be supported with these solutions.

In this paper, we propose a knowledge-based system using ontologies and rules, for modeling methods of geovisual analytics, and their contribution to the exploration and the analysis of trajectories. Contrary to existing models for visualization, this model is not only based on data types. It also takes into account the whole context of use: user’s profile, tasks and data. This system would be used to support
users in geovisual analytics of movement, by proposing adequate visualization methods of the information.

In Section II, we present a state-of-the-art for modeling the visualization process and methods. After identifying the limits of these models, we propose an ontological model for GeoVA environments and the use of SWRL rules in Section III. Section IV gives an example for using this system, applied to Maritime Domain Awareness. We also present the proof of concept for its use. Finally, we conclude on our contribution to visualization in Section V.

II. RELATED WORK

After the international workshop GeoViz’13, Schiewe proposed a definition of geovisual analytics as a linkage of visual and computational methods and tools for extracting hypotheses and information from spatial data [9]. As we can see, it is essential to have in mind that in decision-making, hypothesis and conclusions cannot be drawn from simple spatial data: it is the very combination of spatial information, temporal information and other attributes of the studied objects, that allows extracting patterns of interest [4], [10]. In order to propose adequate visualizations for specific movement exploration tasks, these three views of the information must be considered. Using such combination of various visual representations and interactions with the information is very common in literature concerning this problem [11], [12]. Yet, it is fundamental to know how these tools can be used in the process of visual analytics, for the exploration and analysis of movement information.

Andrienko & Andrienko proposed four stages for exploring and analyzing movement data [13], and for classifying visualization methods. These stages are, in the order: looking at trajectories, looking inside trajectories, bird’s-eye view on movement, and finally investigating movement in context. These stages are sorted from a general overview of movement with classes and clusters, to detailed investigation of the data, and their explanation based on context information. The more we go into details, and the more expert knowledge is needed to interpret the behavior of a single object; thus the less automation is required.

These many works show there is not a single geovisual analytics environment that can answer all questions, which are related to movement exploration and analysis. In order to compare existing solutions, it is required to adopt a common model to formalize what the visualized data are, how they are visualized and how we can interact with it.

When modeling the visualization process from raw data to visualization spaces, it has been widely recommended to make a clear distinction between the data models and the visual space models [14]. This allows refinement, correction and adaptation of these two aspects, along with the domain of application. This concept is known as the Reference Model pattern [15], which is illustrated on Fig. 1. This model proposes five classes: the data source that creates the data set, the view space and its controls, and the visualization being a specific projection of a certain data set. Following this framework since the 90’s, much research has been producing models for all these concepts.

The historical data model was introduced by French cartographer Jacques Bertin [16], who proposed a taxonomy of visual variables and their use, according to the type of raw data. This visual dictionary is still the most used reference in visualization. For modeling the data types (DataSet), Shneiderman also proposed a taxonomy based on the dimension and the structure of data [17]. These data types came along with a taxonomy of tasks in visualization, inspired by the “Visual Information Seeking mantra”: Overview, Zoom, Filter, Details-on-demand, Relate, History and Extract. These control types are expanded with a complete state-of-the-art, which was led by Roth [18]. The author proposed three main groups of controls: objective-based (tasks and purposes of use), operator-based (HCI controls) and operand-based taxonomies (transformation of visual properties). These models for data sets, views and controls can be used in further research on visualization, in order to describe which data can be visualized, their visual form and the possible interactions with it. Yet, the very principle of visualization is a projection of information towards a visual space, therefore requiring a transformation of the raw data into visual information [19].

Therefore, models have been proposed to describe the successive steps of data transformation into visual variables, within visualization spaces. For instance, Card & Mackinlay [14] proposed a characterization of visualization with the use of a table. Lines are used for listing the various information that are visualized, and columns show the process of data filtering, transformation, projection in space, the visual variable that is used, and possible interactions with it. This model was then extended by Hurter & Converse to include a description of dynamic displays in the model, such as animation [20].

A second work of interest has been proposed by Chi [21] with the Data State Reference Model (DSRM). Like the C&M model and the Reference Model pattern, it describes each step in data projection into visual space. These
successive stages are: Value, Analytical abstraction, Visualization abstraction and finally View space (see Fig. 2). In each of these data spaces, users can interact with information thanks to controls, such as filter, representation means, etc. DSRM was applied to the domains of InfoVis and scientific visualization, where data do not always have concrete and pre-defined representations – contrary to geographical information.

These models for the formalization of visualization methods have allowed major progress in the automation of visualization process, based on the input data. For instance, software such as Show Me [22] or AutoVis [23] would base their analysis of data structure on these models, and propose automatic visualization of the information. But these automatic processes are only based on visualization theory, meaning they do not take into account the reality of user’s habits, or the work tasks they have to perform to explore a certain data set.

Some similar research works have proposed systems for intelligent visualization. Gotz & Wen have based their research on the behavior of user, to propose behavior-driven visualization recommendation [24], instead of the usual data-driven or task-driven approaches. This recommendation is based on the current use of visualization by users, which is analyzed to infer future analysis. Similar works have been achieved by Hipp et al. [25], with the use of a semantic representation of process, information and context for user-adapted visualizations. This work on semantic allows using adequate process information for business processes. But such work has not yet been applied to the domain of geovisual analytics, to recommend adequate methods in the exploration of geographic information.

Therefore, it is fundamental to explore such an intelligent system in geovisual analytics, to take into account the whole context of use. We base our work on the DSRM by Chi for describing existing geovisual analytics environments. Whereas the C&M model is well adapted for comparing visual processes with tables, in a clearer way, DSRM is best adapted for the use of classes and relations, in a knowledge-based system. Next section presents the development of a formal model based on DSRM in a knowledge base, using ontologies and rules.

III. A KNOWLEDGE-BASED APPROACH FOR THE SUPPORT OF GEOVISUAL ANALYTICS

As a start to this process, we define a Situation as the context of use, composed of: a given user, a single task or a collection of ordered tasks, the available data and the time period of studied data (e.g., real-time, cyclic pattern, linear). Users are defined by their profile, which gives indication of interest for further study of their skills: job and place of work (i.e., the place where they use the visualization), experience with the data, scientific education and technological abilities (according to the available hardware, or user’s own experience with information visualization). To limit the possible values of these last three attributes, we use a scale from 1 to 3; 1 being Basic and 3 being Expert. These simple values will be used later for writing the rules for the use of visualization methods.

Our framework is based on four main components to support the analysis of information: (1) the model of the current situation, (2) the model of GeoVA methods, (3) the rules that allow matching the situation to the various methods, and (4) the reasoner that applies these rules. For representation of knowledge in GeoVA, and intelligent machine reasoning, a formal model was meant to be used. For this reason, we chose ontologies for the knowledge representation, using the formal language OWL, based on RDF. Ontologies are based on a formal and explicit specification of a shared conceptualization [26], making this representation of knowledge re-usable by other communities. Moreover, the use of ontologies also allows setting up rules with the developed concepts, using SWRL formalism. The rules are based on “if, then” statements with the existing concepts and individuals in the ontologies.

To formalize GeoVA environments, we chose the most used visualization methods in current visualization systems and GIS functionalities for the analysis of movement. We also chose to take into account recent works that were achieved by GeoViz community for advanced visualization of mobile objects and their own limits. For instance: space-time cube [27], trajectory wall [13] or self-organizing map [28]. This allows proposing new ways to visualize the information and to discover new knowledge. These solutions depend on the current context of use, and on the limits of these methods: amount and time of data, user’s profile and acceptance of technology, etc.

Fig. 3 shows how these four components are used, and at which stages the rules are employed. Users define a situation of use, with their own profile and the context of work. Choosing the main goal to be performed will ask for certain data defined by the ontology, and a list of analytical tasks. These analytical tasks were identified with users of maritime domain, based on their own experience. According to the successive stages that are identified, an automatic selection of the possible visualizations is done. This allows finding: which visualizations can represent these data, and which visualizations can propose these task stages?

From this first selection of visualization methods, based on pure visualization theory, a second selection is done. This selection results from the user’s profile, which may require simple visualization methods: for instance no use of 3D or no complex statistics. The time of the data (real-time, historical) also acts as a filter after the first selection, as many visualization methods cannot be applied to real-time data. The second selection of visualization methods provide adequate solutions for the data, the user’s profile, the context of use, and the exploration task(s) to be led.

In order to develop a formal model of GeoVA environments, we followed the DSRM framework. We divide the ontological model for data transformation in three concepts: Data, VisualShape and VisualData. Data concept stands for domain data, i.e., the raw data as they are collected into databases. Each individual of this concept has a name, understandable and sharable by all users of the domain. For instance, within maritime data, we have the concepts: ShipTrajectory, ShipSpeedOnGround, Bathymetry, etc. With the use of visual attributes such as shape or color, these data
are then processed into visual data, which is the Visual Abstraction in DSRM. In terms of ontologies, the concept VisualData is at least one Data object, coupled to a single VisualShape object. This visual data is then used in a visual space for a specific visualization method. As we define GeoVA environments earlier, we use two different concepts in the ontology for their conceptualization: GeoVAE and VisualSpace. Environments being the linked usage of various visualization spaces, we developed object properties such as: hasVisSpace(GeoVAE, VisualSpace) for describing the composition of environments, and linksVisSpace(GeoVAE, VisualSpace) for describing which visual spaces are dynamically linked. Moreover, we use the property hasInteraction(VisualSpace, Interaction) for listing the controls that are allowed in each visualization space, based on existing taxonomies of interaction.

Based on publications on the evaluation of visualization spaces that are included in our ontologies, we used data properties to inform about the limits with the use of visualization methods: for instance the maximum amount of data, the supported type of time for display. This information gives fixed details on each visualization space. This represents theoretical knowledge of what is made possible with each.

As it is illustrated with Fig. 3, a first selection of theoretically adequate solutions is made from these descriptions. A second selection is made from a restriction in these solutions, according to the user’s profile. As they were introduced earlier in this paper, user’s characteristics are processed with SWRL rules to select visualizations that are adapted to user’s expectations (tasks, habits) or that could be easy to understand and to use, according to user’s technological abilities.

In next section, we show how to use this framework, with a behavior of interest, which cannot be interpreted on its own. We also present some geovisual analytics solutions that have been developed within a web GIS (Geographic Information System).

IV. EXPLORING STOPS IN SHIP MOVEMENTS

A. Application to Maritime Domain Awareness

We now give an example of a situation where visual analytics are strongly needed, in the maritime domain. To introduce this, a fundamental and recurrent mantra in visual analytics is: “Detect the expected and discover the unknown” [2]. In terms of the Cynefin model, that classifies events in four categories (simple, complicated, complex and chaotic) [29], we could translate this visualization mantra as: visual analytics provide methods to recognize simple and complicated situations, and to discover relations in complex situations. In this section, we present the application of a situation awareness scenario, where GeoVA is a major issue.

To illustrate the contribution of this system for the visual analytics of risks and threats, we choose to explore the case of Maritime Domain Awareness. In maritime surveillance, the operators are constantly faced to heterogeneous data, coming from AIS or RADAR captors, meteorology, etc. Their goal is to monitor real-time traffic and detect anomalies in maritime traffic. Analysts are also included in the decision-making, as they analyze past information for supporting decisions in the actions. We choose an event that would not be identified with the simple use of current surveillance system, i.e., a complex situation according to the Cynefin classification.

![Figure 3. Use of the ontology in the reasoning process for proposition of geovisual analytics.](image-url)
The schema on Fig. 4 illustrates an event that would not be detected by current surveillance tools, which are limited to real-time display of ship locations. This scenario is defined by two ships in the open sea, stopping at the same place \((X_S, Y_S)\) and having significant time difference between the two passages. The data overload in surveillance tools and the human cognitive workload are major brakes to the visual detection of such an event. However, this could be a security matter such as two ships exchanging illegal goods (drugs, arms), by dropping it at sea.

This type of scenario cannot be anticipated with current technologies in maritime rescue coordination centers (MRCC). Indeed, current surveillance tools do not allow a complete investigation of this situation, for too little analysis tools are available [29]. We remind the definition of a complex situation [30], in which the relation between causes and effects cannot be anticipated, and is only perceived afterwards. In this context, detecting two consecutive stops of different ships is considered as a complex situation, for their detection and interpretation cannot be anticipated. Some information can be identified for characterizing this scenario, to decide whether there is a security matter or not.

As we have described above in this paper, the first stage is to define a \(\text{Situation} \) object in the ontology. In this example, the controller does not know if there is such a common stop location. A \(\text{Situation} \) needs to be defined for exploring a possible drug traffic scenario, for instance. It is defined as two small ships stopping at the same place, with a particular time difference. This event should happen when the sea is calm, in a non-busy area. The following axioms are used with pre-defined objects in the ontology:

- \( S\text{-hasUser}(\text{Situation, Controller}) \)
- \( S\text{-hasTime}(\text{Situation, HistoricalTime}) \)
- \( S\text{-hasGoal}(\text{Situation, DrugExchange}) \)

Once the new situation object is added to the ontology, the reasoner analyzes the axioms and applies the rules that are defined. As the goal \(\text{DrugTraffic} \) is already defined, the tasks are automatically added to the situation as follows:

\( S\text{-hasTask}(\text{Situation, GetAreaInfo}) \): Maritime context information such as wave, wind, current, etc.
\( S\text{-hasTask}(\text{Situation, GetStops}) \): Locate stop locations and their duration
\( S\text{-hasTask}(\text{Situation, MeasureSpace}) \): Simple measure of distances, generally on all maps
\( S\text{-hasTask}(\text{Situation, MeasureTimeDiffSpeed}) \): Measure the time difference, based on some characteristics for speed (for instance, between two stops locations, or locations with low speed).

With these inputs that characterize the situation, the reasoner applies rules to identify new links and properties between the ontology’s concepts. This results in a list of visualization methods, which are adequate to the analysis to lead. In this specific case of drug exchange with a controller, we obtain the methods: \(\text{DensityMap}, \text{SpeedGraph}, \text{TrajectoryMap}, \text{StopMap} \) and \(\text{SpeedMap} \). A proper geovisual analytics environment could, for instance, combine these visualization methods to support the exploration of a possible drug exchange. Using this GeoVA environment, users can lead a complete analysis based on their own knowledge and their own interpretation of the visualized information. Refining the input tasks, or user’s profile, leads to a new proposition of visualizations. This exchange between the knowledge-based system and the human operator allows discovering possible links between the visualized data, which may not have been perceived with usual surveillance tools.

\section{Graphical User Interface for Using the Ontology}

A graphical user interface (GUI) was developed to use this knowledge-based system. This interface is a user-friendly application developed with Java OWL API, which allows reading and writing in the ontology. This allows the user to create a profile and to inform about the situation to study. Fig. 5 gives a screenshot of this GUI, as an analyst user could fill it for the analysis of a fishing behavior. The interface makes a selection in all the possible tasks according to the \(\text{Goal} \) that was selected. Once the user’s profile and the situation are created, these data are written into the ontology. Then, the reasoner Pellet is used to apply rules, and complete the new \(\text{Situation} \) object. Visualization methods are then proposed to users in a list of adequate methods, so that they can choose a method.

This GUI is still under construction as a proof-of-concept, and has not yet been tested by maritime users. Ongoing work is planned for creating a link from the selected method, as an icon, to a real and ergonomic geovisual analytics environment. This is the next challenge in our research work, now that adequate methods for geovisual analytics can be proposed.

Several methods for geovisual analytics have been implemented into the ontology and in a web-mapping platform, called FishEye. This web application is used to test visualizations with maritime data, along with several algorithms for the detection of unusual behavior, such as automatic recognition of threats in near real-time with ontologies [31], or data-mining techniques for identifying sensitive areas and unusual movements [32].
AIS data are collected from ships with GPS devices, and displayed in near real-time on a map with FishEye. This information is used for creating, for instance, density maps, common operational pictures, trajectory displays, or for leading trajectory analysis. The visual environment illustrated on Fig. 6 is an example of visualization, which is proposed by the knowledge-based system. This one is adapted for the exploration of any activities that requires investigating the stops of a ship. For instance, drug traffic or illegal fishing, where stop locations are major information. The visualization (a) on the figure is an example for the “Stop map” that was identified in previous part, after the reasoning process. The part (b) of the figure shows the equivalent in the ontology of the map view. The circles correspond to stop locations, where size and color (yellow to red) stand for the duration of this stop. The color of the line stands for the speed of the ship on the part of its trajectory. The graph in the upper right part of the figure is a speed graph, showing times where the ship stopped, and changes in the speed. While passing the mouse over this graph, a marker is displayed on the map for linking the graph and the map.

V. Conclusion

We have introduced a system based on ontologies, for the formalization of knowledge in geovisual analytics. This system supports in the use of adequate visualization of spatial, temporal and semantic information. Previous research works have proposed intelligent systems for data-driven visualization; the limits of these work was to adapt these visualizations to the context of use. To extend this research, we proposed a reasoning system that takes into account the context of use, in order to propose useful and adequate visualization methods for geospatial information. This context of use is characterized by the profile of user, the analysis to perform, and the data to use. Moreover, we used ontologies for modeling knowledge about the field of Maritime Domain Awareness for a specific application. Data and tasks concerning known threats or anomalies are used to identify adequate visualizations. From this conceptual model, we developed rules to analyze the context of use in MDA, and to compare visualization possibilities to the data, the tasks and the user profile.

This work can be applied to any other research field dealing with analysis of movement data, based on geo-information. For instance, air traffic control, car traffic or pedestrian movements. The goals and data can differ from specific maritime requirement, but the reasoning process for studying the context of use and comparing it to the visualization possibilities still remains the same.

However, this research shows its limits on the visualizations that can be used. We have chosen visualization methods that have already been published, to use their results of usability as rules. As our reasoning system is based on the knowledge of capabilities and limits (data amount, visualized time, etc.), it cannot be used with visualizations, whose usability has not been studied. Thus, this knowledge-based system can only advise on existing and implemented geovisual analytics methods and environments, and cannot create “on the fly” visualizations as Show Me or AutoVis did.

Further research of great interest in this field could concern an intelligent display of the various visualization methods that are selected by the knowledge-based system. Indeed, using adequate visualizations with a poor display is another brake in the usability of such geovisual analytics environment. This problematic raises major issues on the ergonomics of visual analytics.

REFERENCES


