Combining Text and Graphics for Interactive Exploration of Behavior Datasets

Martin Molina, Enrique Parodi
Department of Artificial Intelligence
Technical University of Madrid, Spain
martin.molina@upm.es, eparodi@fi.upm.es

Amanda Stent
AT&T Labs - Research
Florham Park, NJ, USA
stent@research.att.com

***Abstract*—Modern sensor technologies and simulators applied to large and complex dynamic systems (such as road traffic networks, sets of river channels, etc.) produce large amounts of behavior data that are difficult for users to interpret and analyze. Software tools that generate presentations combining text and graphics can help users understand this data. In this paper we describe the results of our research on automatic multimedia presentation generation (including text, graphics, maps, images, etc.) for interactive exploration of behavior datasets. We designed a novel user interface that combines automatically generated text and graphical resources. We describe the general knowledge-based design of our presentation generation tool. We also present applications that we developed to validate the method, and a comparison with related work.**

Keywords – intelligent user interface; multimedia presentation; interactive data exploration

I. INTRODUCTION

There is currently rapid growth in the availability of large data sets about the behavior of dynamic systems (e.g., networks of sensors, telecommunications, the internet, etc.). Examples include SoilWeather, a network of geographically distributed sensors in a river basin that measures water quality, soil conditions and weather year-round [8], and Nericell, a system that models traffic behavior using GPS in smart phones [10]. Analysis of these types of data is useful for many tasks, including trend analysis (e.g. fleet monitoring), planning (e.g. buildout planning), and problem identification (e.g. detection of traffic accidents or flooding).

Multimedia presentations combining text and graphics provide multiple and complementary views which can offer a number of important benefits to help users understand the underlying data [14][16]. Only a few research projects have paid equal attention to text generation and to the integration of presentation elements across different modalities. For example, COMET [9] and WIP [1][15] generate multimedia explanations about how to manipulate devices.

The addition of more complex text presentations about behavior datasets (including summaries, comparisons, and other types of presentation) can provide improvements to existing visualization tools by facilitating more natural and effective analysis. However, there are research challenges to adding more complex text output to visualization tool, such as how to automatically generate coordinated mixed-media presentations that are adapted to the user and context [2][3][4] or how to design effective text that augments graphical presentations [7][12].

In this paper we describe results of our ongoing research on automatic multimedia presentation generation (including text, graphics, maps, images, etc.) for interactive exploration of behavior datasets. We designed a novel user interface that combines automatically generated text and graphical resources. We describe the general knowledge-based design of our software tool together with the representations and reasoning methods we used for the generation of presentations. We also present domains that we used to validate the method, and a comparison with related work.

II. THE USER INTERFACE

Our goal is to produce multimedia presentations (combining text with graphics) in a layout that will be familiar to general users, and require minimum learning effort. For this purpose we have chosen a journalistic presentation style. This includes the following main components:

- *Headline*. The headline summarizes the main idea of the text.
- *Text body*. The text body contains content that supports or elaborates on the headline but summarizes the information in the graphics. According to the journalistic style, the text body organizes a news story as an *inverted pyramid*. This style puts the essential and most interesting elements of the story at the beginning, with supporting information following in order of diminishing importance. This structure enables readers to stop reading at any point and still take the essence of a story.
- *Figures*. Complementary figures are included that can help the user to understand the text. Different types of figures can be presented: a 2D graphic, map, 3D animation, photograph, etc.

Proceedings 22th International Conference on Tools with Artificial Intelligence, ICTAI 2010, pp. 150-155. IEEE Computer Society. Arras, France, October 2010. ISBN: 978-0-7695-4263-8.

Figure 1. Example of user interface for interactive exploration of behavior datasets. The text body summarizes the information, describing the relevant data with an appropriate discourse structure. The user can click on the text to see complementary figures and animations generated by a 3D graphics engine.

Users can interact with generated presentations in several ways. First, the body text is clickable. When the user selects a part of the text, the figure changes to show related details. Second, some figures are interactive. For example, a figure can be a video that can be operated by the user with the corresponding typical control buttons (play, forward, backward, pause and stop). Another type of interactive figure can be a 3D exploration tool with more specific control buttons (e.g. the Google Earth widget or a specific graphical viewer for aircraft models). Figure 1 shows an example of this type of user interface. This example corresponds to a prototype for exploring data about military missions generated by a simulator. This type of presentation can be displayed in a web browser, on a desktop or on a mobile phone. On a small screen, the text can automatically be read to the user, and display of figures can be aligned with the spoken material.

III. THE ARCHITECTURE

Figure 2 shows the general knowledge-based architecture of our method. It includes components for the main tasks: discourse planning, text generation, graphics generation and data abstraction. The first task, discourse planning, generates the discourse plan for an output multimedia presentation. This is done in a loop with the abstraction task. At each iteration, the discourse planning task generates abstraction goals and the abstraction task returns the abstractions corresponding to these goals. Discourse planning is directed by a hierarchical search over possible discourse patterns; the discourse patterns are taken from the discourse model.

In the following sections, we describe this architecture in more detail. Because the focus of this paper is the combination of text and graphics, we pay special attention to the components for multimedia presentation (more details about the rest of the components can be found in [11]).

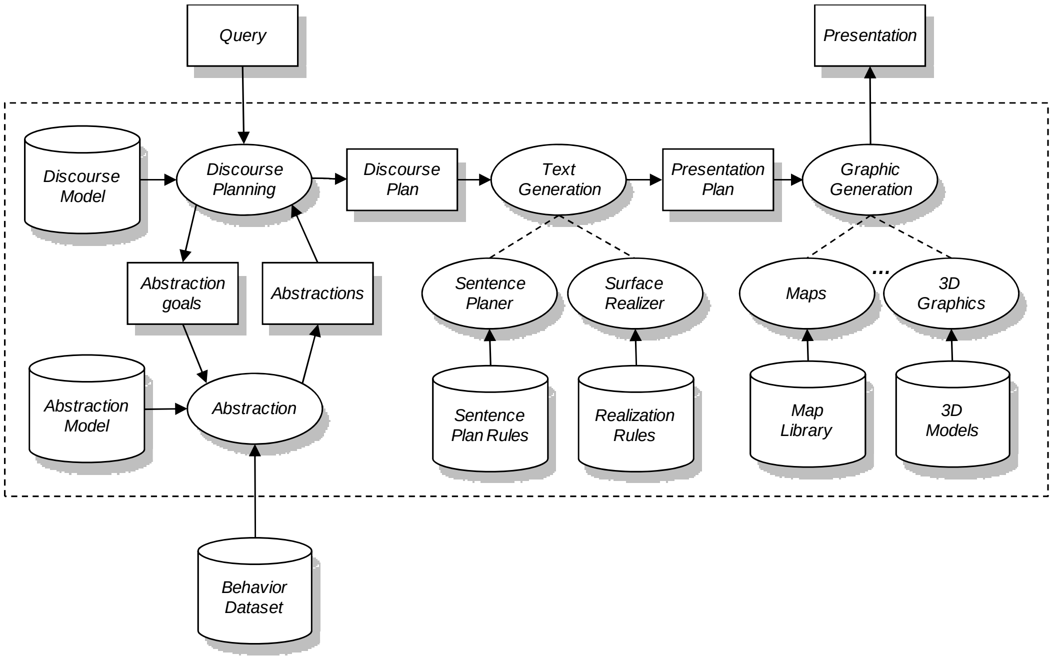

Figure 2. Main components of our method for interactive exploration of behavior datasets. In the figure, rectangles represent input-output data, ellipses represent tasks or subtasks and cylinders represent knowledge bases and data bases.

A. *The Discourse Planner*

In our system, a presentation includes (1) text that summarizes the data (headline and body text) and (2) a set of figures that complement the text. The relation between text and figures is expressed as a *presentation plan*. It is used to interact with the user in such a way that, when the user clicks on certain text segments, the corresponding figures are presented according to the defined relations between figures and text.

```
<presentation>
<head>
  <title>Air Interdiction in Granada</title>
  <subtitle>Scenelab Simulation</subtitle>
  <date>April 28, 2010</date>
</head>
<body initial-fig="general">
 <p> The <a fig="general">Air Interdiction Mission in
     Granada </a> was an attack made by the <a
     fig="alphaf">Alpha Forces</a> against Two Air
     Defense Units (<a fig="adu3">ADU-3</a> and <a
     fig="adu4">ADU-4</a>) of the <a fig="bravof">Bravo
     Forces</a> located near  <a fig="granada">Granada
     City</a>
     ...
 </p>
 ...
</body>
...

<figure label="The ADU-3 Air Defence Unit" ref="adu3">
 <set-time t="0"/>
 <locate actor="ADU-3" label="ADU-3" highlight="true"/>
 <show-attribute actor="ADU-3" attribute="location"/>
 <show-attribute actor="ADU-3" attribute="range"/>
</figure>
...
</presentation>
```

Figure 3. Partial presentation plan.

In order to represent a presentation plan we have designed an XML-based representation (Figure 3). The representation is divided into the following sections: (1) header, (2) body and (3) a list of figures. In the header we define attributes such as title, date, sources and meta-information related to the presentation. The body contains the labeled text with references to the corresponding figures. The figures are defined separately using a primitive set (see more details below).

OPERATOR: *operator-15*
GOAL: *elaborate_with_quantity* (*o*, *q*)
CONDITIONS: (*o*: *river-section*) ∧ (*q* = *flow*) ∧ *value*(*q*, *o*, *time*, *v*)
∧ *attribute*(*o*, *geometry*, *g*} ∧ *attribute*(*o*, *name*, *n*})
ACTION: {*generate-text*({"there is a flow of ", *v* }, *t1*),
generate-text({"at ", *o*}, *t2*),
generate-figure(2DBars,{*plot*(*o*, *q*, *time*, 3*h*), *title*(*n*)}, *f1*),
generate-figure(2DMap,{*locate*(*g*, *n*), *balloon*(*g*, *f1*)}, *f2*),
link-text-figure(*t2*, *f2*)}

Figure 4. Example of presentation operator.

To dynamically construct the presentation, we use a discourse planner. The discourse planner uses a knowledge base with presentation operators. Figure 4 shows an example of a presentation operator. Each operator is associated with a communication goal. In this example, the goal is *elaborate_with_quantity*, i.e., to give details showing the quantitative value of a parameter that characterizes the state of a certain object of the dynamic system. In general, there are several candidate operators for a given communication goal. Each operator includes a set of conditions that must be matched to the communicative goal.

In this example, the conditions establish that the operator is applicable for a particular type of object (river section).

In addition, the operator includes an ordered set of actions to generate a part of the presentation. There are three types of actions: (1) actions to generate text, (2) actions to generate figures and (3) actions to link text segments to figures. In this example, the actions to generate text are based on templates. However, in order to be more flexible, they could use techniques based on natural language generation. In order to generate figures, we assume that there is a prefixed set of presentation primitives (see below).

PATTERN: *pattern-32*
GOAL: *elaborate_state*(*x*)
CONDITIONS: (*state*(*y*, *s*) ∈ *x*) ∧ *single_component*(*y*) ∧ *quantity*(*y*, *q*)
DISCOURSE: {*elaborate_with_quantity*(*y*, *q*),
contrast_to_previous (*y*, *q*),
contrast_to_average(*y*, *q*)}

Figure 5. Example of discourse pattern.

Our discourse planner is a knowledge-based hierarchical planner that constructs the output presentation plan through a search directed by partial presentation strategies that are progressively refined until the final presentation plan is generated. The planner itself is a modified version of HTN (Hierarchical Task Network) planning [5]. The HTN planner includes control knowledge represented by planning-tasks and planning-methods that helps select operators. In our approach, each planning-task corresponds to a communication goal (e.g., to inform about a relevant fact) and each planning-method corresponds to a discourse pattern. Figure 5 shows an example planning. Besides the goal and conditions (which are interpreted as are the goal and conditions in a presentation operator), the planning-method includes a *discourse*: an ordered set of communication goals.

B. *Text and Graphics Generation*

Our architecture can use two alternative methods for text generation: (1) template-based generation and (2) natural language generation. The template-based method uses templates of text with variables that are instanced with particular values. However, to provide richer and more flexible texts, our method can also use natural language generation. In this case, the planning operators of the presentation planner use a modified representation (see details about this in [11]).

We generate each figure of the presentation according to a sub-plan produced by the presentation planner. The sub-plan is an ordered set of presentation actions $\{a_1, a_2, \ldots, a_n\}$. Each action invokes a presentation primitive that can be interpreted by graphics generation tools. Figure 6 shows a list of presentation primitives that we use in our method. They correspond to general primitives for generating graphics related to geospatial areas and temporal references using maps, 3D animations on virtual terrains, 3D objects and time line.

The generation of the figures by using the primitives is done by different software tools specialized in graphics manipulation. For example, a figure can be defined by several actions that invoke the primitive *Locate*(*x*, *y*, *z*); this primitive may be realized in a web-based viewer (supported by tools such as GoogleMaps) as a map that shows the locations of a set of objects together with either the name of each object or a number for each object and a legend associating numbers and names (when there are too many objects).

Locate(actor, label, highlight)
Show-Relative-Location(a, b, label1, label2)
Show-Route(actor, start-time, end-time, color, show-points)
Show-Route(waypoints, color, show-points)
Show-Attribute(actor, attribute-name)
Set-Focus(actor, focus-mode, …)
Set-Time(time)
Set-Time-Lapse(start, end, looping-enabled)
Highlight(actor, mode)
Show-Image(source, label, actor)
Show-Video(source, label, actor)

Figure 6. Presentation primitives for generating figures.

Figures can be also generated and manipulated with the help of a 3D viewer using animated 3D visualizations of different objects (e.g., aircrafts, missiles, etc.). We developed a graphical viewer using OpenGL-based libraries to generate animations by interpreting the general primitives. For example, the primitives *Set-Focus*(*x*, *y*, *z*) and *Set-Time-Lapse*(*x*, *y*, *z*) can be used to generate an animation that allows the user to move forward and backward through a time interval. This type of animation can be done to illustrate, for example, a missile launch with the camera fixed on the missile while keeping the target in sight.

IV. Applications

We validated our general design using three different domains: hydrology, travel and aircraft simulation. Figure 7 shows an example of an application called VSAIH that we developed in the domain of hydrology following our general design. For this purpose we used hydrologic information from a national system in Spain. Using this system, data is received periodically at control centers about rainfall at certain locations, water levels, and flow discharge in reservoirs and flows in certain river channels.

We validated the presentation models by continuous operation of this application for more than one year with the help of experts in hydrology who refined details of the summaries. The application includes the three models (flood risk, water management and sensor validation) and generates summaries every hour with approximately 48,000 measures (time series containing hourly readings from about 2000 sensors). The discourse model for flood risk includes 103 operators and 142 discourse patterns. This version uses templates to generate text.

We also applied the general design to the domain of travel, with GPS data as the sensor data. Details about this application can be found in [11].

Figure 7. Example of user interface for interactive exploration of datasets in the domain of hydrology. This corresponds to an application developed in Spain to explain data recorded by sensors about the behavior of rivers and basins. The headline says that there is an important increase of the flow in two rivers of Spain (Ebro basin).

In addition, we are applying our general design to the military domain to generate summaries about aircraft missions. Here, the datasets are created by a simulator that generates the detailed behavior of a set of actors (aircrafts, missiles, radars, etc.) in a battle scenario (see Figure 1). A preliminary evaluation with different scenarios showed that our method is also able to capture presentation needs in this domain.

V. Related Work

The work that we describe in this paper is related to multimedia presentation systems. For example, our method integrates and extends some general ideas about presentation planning as it was done by prototype systems such WIP [1][15] and COMET [9]. The goal of these systems was to generate text explanations accompanied by static illustrations to help users to manipulate a device (e.g., a radio receiver-transmitter or a coffee machine). Compared to these systems, our method is oriented for a different task: generating text explanations combined with interactive illustrations to help users understand the behavior of a complex dynamic system (e.g., a network of river basins or a battle simulation). In addition, our method uses a different type of user interface and richer graphics.

Our method for data summarization has some similar general components to other data-to-text systems (for weather forecasting [13], engineering [17], or medicine [6]. However, our method uses a different abstraction approach, directed by the planning process, and our presentations are multimodal.

We designed a representation for presentation plans that defines the relationships between text and figures in the output presentation. This representation can be compared to the language SMIL (Synchronized Multimedia Integration Language) which also integrates multimedia features. As a main difference, our language establishes a particular coordination between text and language during the communication which cannot be supported by the SMIL language.

VI. CONCLUSIONS

In this paper we have described a method for interactive exploration of behavior datasets that combine text summaries and graphics (maps, 3D animations, etc.). Our method shares some components with other presentation systems, such as text generation and presentation planning. However, it includes several novel aspects: (1) it defines a novel user interface following the idea of text browsing complemented by interactive figures to explore behavior datasets, (2) we designed a specific representation for presentation plans together with a knowledge-based method to generate the plans, and (3) the method has been validated in three different domains (hydrology, travel and aircraft simulation).

The method presented in this paper is an experimental prototype that corresponds to a preliminary result of our ongoing research on multimedia presentation generation. We are currently working on a deeper and more extensive evaluation of this method using additional domains. Our future plans related to this work include to identify and represent computationally reusable presentation strategies and to experiment with them in general domains (e.g., geographic, temporal, social, etc.).

ACKNOWLEDGMENT

This work was partially supported by the Ministry of Science and Innovation of Spain within the VIOMATICA project (TIN2008-05837/TIN).

REFERENCES

[1] E. André and T. Rist, "The design of illustrated documents as a planning task", in Intelligent Multimedia Interfaces, M. Maybury, ed., AAAI Press, 1993, pp. 94–116.

[2] G. Carenini and J. Moore, "A strategy for generating evaluative arguments", Proc. First Int'l Conf. on Natural Language Generation, 2000.

[3] V. Demberg and J. Moore, "Information presentation in spoken dialogue systems", Proc. 11th Conf. European Chapter of the Association for Computational Linguistics, 2006.

[4] T. Gaasterland and P. Godfrey and J. Minker, "An overview of cooperative answering", Journal of Intelligent Information Systems, vol. 1, no. 2, pp. 387–416, 1992.

[5] M. Ghallab D. Nau, and P. Traverso, Automated planning: Theory and practice, Morgan Kaufmann, 2004.

[6] J. Hunter, A. Gatt, F. Portet, E. Reiter, and S. Sripada, "Using natural language generation technology to improve information flows in intensive care units", Proc. 5th Conf. on Prestigious Applications of Intelligent Systems, 2008.

[7] S. Kopp, P. Tepper, and J. Cassell, "Towards integrated microplanning of language and iconic gesture for multimodal output", Proc. 6th International Conference on Multimodal Interfaces, 2004.

[8] N. Kotomäki, S. Thessler, J. Koskiaho, A. O. Hannukkala, H. Huitu, T. Huttula, J. Havento, M. Järvenpää, "Wireless in-situ sensor network for agriculture and water monitoring on a river basin scale in Southern Finland: Evaluation from a data user's perspective", Sensors, vol. 9, pp. 2862-2883, 2009.

[9] K.R. McKeown and S.K. Feiner, "Interactive multimedia explanation for equipment maintenance and repair", Proc. DARPA Speech and Language Workshop, 1990.

[10] P. Mohan, V. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones", Proc. 6th ACM Conference on Embedded Network Sensor Systems, 2008.

[11] M. Molina and A. Stent, "A knowledge-based method for generationg summaries of spatial movement in geographic areas". International Journal on Artificial Intelligence Tools, Vol. 19, No. 4, 2010.

[12] M. O'Donnell, "Intermixing multiple discourse strategies for automatic text composition", Revista Canaria de Estudios Ingleses (RCEI), Special Issue on Intercultural and Textual Approaches to Systemic-Functional Linguistics, no. 40, 2000.

[13] E. Reiter, S. Sripada, J. Hunter, J. Yu, and I. Davy, "Choosing words in computer-generated weather forecasts", Artificial Intelligence, vol. 67, no. 1-2, pp. 137-169, 2005.

[14] R.J. Sclabassi, "Computational support for the development of medical multimedia applications", Center for Clinical Neurophysiology, University of Pittsburgh Medical Center, 1993.

[15] W. Wahlster, E. André, W. Finkler, H.J. Profitlich, T. Rist, "Plan-based integration of natural language and graphics generation", Artificial Intelligence, vol. 63, pp. 387-427, 1993.

[16] M.Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for using multiple views in information visualization," Proc. Working Conf. on Advanced Visual Interfaces, 2000.

[17] J. Yu, E. Reiter, J. Hunter, and C. Mellish, "Choosing the content of textual summaries of large time-series data sets", Natural Language Engineering, vol. 13, no. 1, pp. 25-49, 2007.