User Task Adaptation in Multimedia Presentations

Giuseppe Carenini, Cristina Conati, Enamul Hoque, Ben Steichen

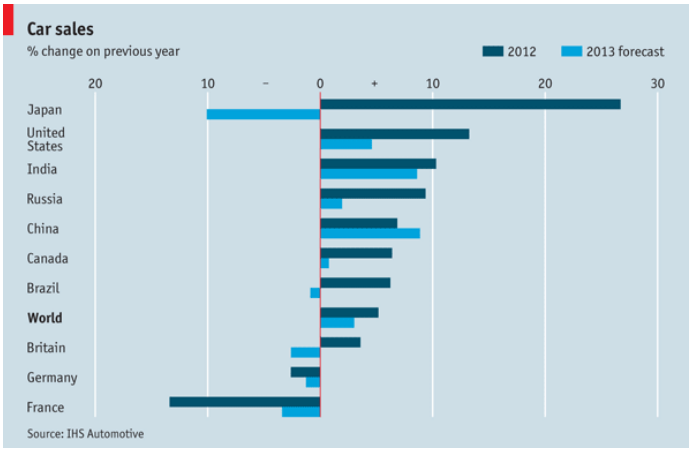
Department of Computer Science, University of British Columbia, Vancouver, Canada
{carenini, conati, enamul, steichen}@cs.ubc.ca

1 Multimedia Presentations Combining Visualizations and Text

It is quite common that documents ranging from newspaper articles to scientific papers convey complex information by combining visualizations with textual material. Presenting information in different modalities not only makes the presentation more engaging, but could also better suit users with different cognitive skills (visual vs. verbal). In these multimedia presentations graphics and text play complementary roles. While graphics can convey large amounts of data compactly and support discovery of trends and relationships, text is much more effective at pointing out and explaining key points about the data, in particular by focusing on specific temporal, causal and evaluative aspects [1]. For illustration, Figure 1 shows an example of a multimedia presentation from The Economist magazine. Notice, for instance, how the sentence "*The end of subsidies to car buyers will lead to a slump in Japan, just as its carmakers' output recovers from the 2011 tsunami."* provides a causal explanation for the noticeably extreme data about current (year 2012) and forecasted (year 2013) car sales in Japan. Generally speaking, the textual part of a multimedia presentation can be seen as suggesting to the reader a set of visual tasks that can be performed by inspecting the visualization. For example, when reading the two sentences *"India and China will have further strong rises—though not at the double-digit rates seen until 2010. Brazil and Britain will suffer reverses."* the reader is prompted to verify in the visualization (the deviation chart) that all the bars for India and China are on the right side of the chart (i.e., sales are increasing) and less than 10%, while the bars for Brazil and Britain are on the right for 2012, but on the left side (i.e., sales are decreasing) for the 2013 forecast.

2 Task Adaptation in Multimedia Presentations

As we have illustrated, the textual component of a complex multimedia presentation typically specifies several visual tasks for the reader. Although the reader should be able to perform all these tasks in the visualization, the visualization components cannot be designed to favour the performance of any particular one of those tasks. Arguably, designing the visualization to support a specific task will likely hinder the performance of the other ones. The main idea of this short paper is that if a system could track what part of the text the reader is currently reading, and from that text it could infer the corresponding visualization task, such a system could **dynamically adapt the visualization** to support the reader in more effectively performing the inferred current task. For instance, if the user read the sentence *"India and China will*



The global car industry: Wheels of mixed fortune *[Source: The Economist- Dec 22nd 2012]* America will enjoy a fourth consecutive year of growth in car sales in 2013, predicts IHS, a research firm. India and China will have further strong rises—though not at the double-digit rates seen until 2010. Brazil and Britain will suffer reverses. The end of subsidies to car buyers will lead to a slump in Japan, just as its carmakers' output recovers from the 2011 tsunami. In the European Union, car sales will fall for the sixth year in a row: they are now back at early 1990s levels. Although some European car factories face closure, elsewhere assembly lines are being built at a rapid clip. So once again worldwide car production, at 82.8m, will exceed sales, at 81.9m. As the metal stacks up on dealers' forecourts, motorists can look forward to some great deals on wheels in 2013.

Figure 1: Sample Multimedia Presentation Combining a Chart with Text

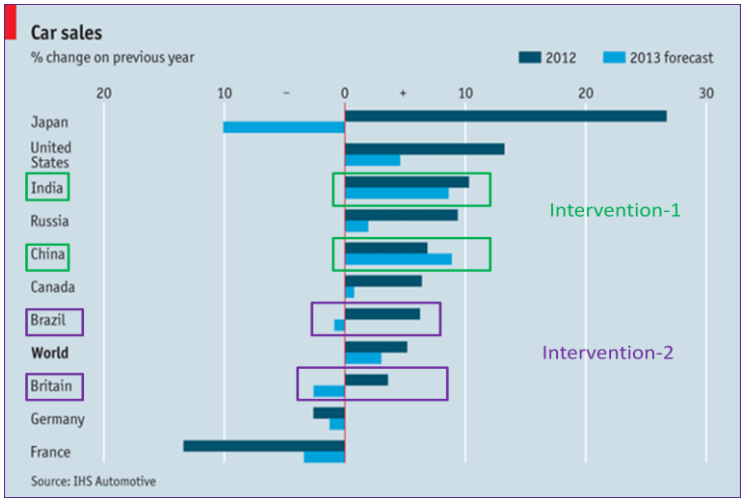


Figure 2: Sample Interventions Adapted to the User Task

have further strong rises—though not at the double-digit rates seen until 2010.", and then turned to the visualization, the system could highlight the information relevant to the corresponding task in the visualization, as shown in Figure 2 by intervention-1. Similarly, if the user turned to the visualization after reading the sentence "*Brazil and Britain will suffer reverses.*", the system could generate intervention-2. Let us now discuss more in detail the modules of a system that can perform task adaptation in multimedia presentations and how each module could be implemented

1) Identify what part of the text the user is currently reading. This module could rely on eye-tracking technology. Although eye tracking devices are still too costly for mass use, more affordable solutions (e.g., based on standard webcams) will be available in the near future [2,3]. **2) Verify if the text the user is currently reading is relevant to the visualization.** Preliminary ideas on how this could be done are presented in [4]. **3) Infer visual tasks for text the user is currently reading.** This is a challenging Natural Language Processing (NLP) problem. Some progress on a related problem was made by Elzer et al. [5] (Sec. 5.2) and [6], while working on using the caption of a visualization as one of the sources to infer the message the visualization was intended to convey. They process the caption to identify and combine verbs (e.g., *lag*), nouns (e.g., *growth),* nouns referring to labels in the visualization, and adjectives (e.g., *soaring*) that are typically used in captions to highlight key points about the displayed data . To be applicable in the system we envision, this work will need to be expanded to deal with more complex and more sophisticated NLP techniques.**4) Provide adaptive interventions.** This module will need to solve at least the following two sub-problems, which also may make use of eye-tracking. First, it needs to decide when an intervention should be triggered and when it should be faded away. A simple approach could be to trigger an intervention every time the user, after reading a chunk of text corresponding to one or more visual tasks, switches her gaze from the text to the visualization. The intervention could then be removed, only if the user has looked at it and possibly returned to the text. The second key problem is selecting the most appropriate intervention for a given task on the specific visualization [7].

3 Further Issues and Questions for Discussion at the Workshop

So far we have described interventions on the visualization to support visual task(s) specified by the text that the user is currently reading. We can also envision interventions on the text, triggered by the user inspection of the visualization. For instance, if eye-tracking data show that the user is inspecting the bottom part of the chart in Figure 1, the sentence about the EU could be highlighted. Such interventions generate a number of research questions: How could the performance of specific visual tasks be detected? (See [8] for preliminary results). Would these graphic to text interventions be useful? What are the implications of allowing both types of interventions?

With the rapid progress in Intelligent User Interfaces, it will become more and more common for multimedia presentations to be generated automatically by computer systems [9, 10]. What are the implications of this for adaptive interventions? Is it the case that for these presentations it may be easier to perform user task adaptation? Next steps in our research include: developing a prototype, run user studies, and also explore adaptation in the context of text-to-speech interaction, which could be more feasible in the short term, as it does not require eye-tracking in the loop.

References

[1] E. R. Tufte. Visual Explanations: Images and Quantities, Evidence and Narrative. 1997.

[2] Sesma, L., Villanueva, A., & Cabeza, R. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12), (2012), 217-220.

[3] Samsung Galaxy S4 eye-tracking smartphone unveiled, http://www.bbc.co.uk/news/technology-21791023

[4] Peng Wu and Sandra Carberry. Toward Extractive Summarization of Multimodal Documents. Proceedings of the Canadian AI Workshop on Text Summarization, pages 53-64, 2011

[5]Elzer, S., Carberry, S., & Zukerman, I. The automated understanding of simple bar charts. Artificial Intelligence Journal 175(2), (2011), 526-555.

[6] S. Elzer, S. Carberry, D. Chester, S. Demir, N. Green, I. Zukerman, K. Trnka, Exploring and exploiting the limited utility of captions in recognizing intention in information graphics, in: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, 2005, pp. 223–230.

[7] Mittal, V.O. Visual Prompts and Graphical Design: A Framework for Exploring the Design Space of 2-D Charts and Graphs. Proc. AAAI/IAAI, 57-63, 1997.

[8] Steichen B., Carenini G., Conati C, User-Adaptive Information Visualization - Using Eye Gaze Data to Infer Visualization Tasks and User Cognitive Abilities. Proceedings of the International Conference on Intelligent User Interfaces, Santa Monica, CA, USA March 19-22, 2013

[9] N. Green., G. Carenini., S. Kerpedjiev, J. Mattis, J. Moore, S. Roth, AutoBrief: an Experimental System for the Automatic Generation of Briefings in Integrated Text and Information Graphics. International Journal of Human-Computer Studies Vol. 61, Issue 1, pag, 32-70, July 2004

[10] Carenini G., Ng R., Pauls A. Interactive Multimedia Summaries of Evaluative Text, *Proceedings of the 10th International Conference on Intelligent User Interfaces* Sydney, Australia, Gen29-Feb1, 2006.