Processing Information Graphics in Multimodal Documents*

Richard Burns
Dept of Computer Science
Univ. of Delaware
Newark, DE 19716 USA
burns@cis.udel.edu

Sandra Carberry
Dept of Computer Science
Univ. of Delaware
Newark, DE 19716 USA
carberry@cis.udel.edu

Stephanie Elzer
Dept of Computer Science
Millersville Univ.
Millersville, PA 17551 USA
elzer@cs.millersville.edu

Abstract

Information graphics, such as bar charts, grouped bar charts, and line graphs, are an important component of multimodal documents and cannot be ignored. When such graphics appear in popular media, such as magazines and newspapers, they generally have an intended message. We argue that this message represents a brief summary of the graphic's high-level content, and thus can serve as the basis for more robust information extraction from multimodal documents. The paper describes our methodology for automatically recognizing the intended message of an information graphic, with a focus on grouped bar charts.

Introduction

Information graphics are non-pictorial graphics such as bar charts, grouped bar charts (sometimes referred to as clustered bar charts), and line graphs that depict attributes of entities and relations among entities. Although some information graphics are only intended to display data, the majority of information graphics in popular media such as magazines and newspapers are intended to convey a message. For example, the information graphic shown in Figure 1 conveys the message that there was a substantial increase in Delaware bankruptcy personal filings in 2001 compared with the preceding decreasing trend from 1998 to 2000.

This paper addresses the importance of information graphics in extracting information from multimodal documents and provides a methodology for identifying what we view as the primary content of the graphic—namely, its intended message. It argues that information graphics in multimodal documents cannot be ignored, either in summarizing the document or in extracting information from it. We then provide a brief overview of our methodology for identifying the intended message of an information graphic. This methodology has been implemented and tested for simple bar charts (Elzer2005b). In this paper, we focus on how the methodology is being extended to grouped bar charts; we discuss the kinds of messages that can be conveyed by a grouped bar chart, and we describe the communicative signals that appear in grouped bar charts and how they can be extracted from an information graphic. Our methodology employs a Bayesian network which uses the communicative signals present in a graphic as evidence to hypothesize the graphic's intended message. We contend that this message captures the high-level content of the graphic and thus can be used as the basis for taking information graphics into account in extracting information from multimodal documents.

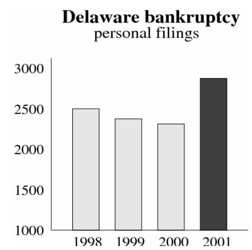

Figure 1: Graphic from a local newspaper.[1]

Information Graphics in Multimodal Documents

Information graphics in popular media, such as magazines and newspapers, generally have a message that they are intended to convey. This message is often not duplicated in the surrounding text (Carberry2006). For example, Figure 2 shows a grouped bar chart from *NewsWeek*. The primary message conveyed by the graphic is ostensibly that '*the percentage of pirated software in China is much higher than in the world as a whole*' and the secondary message is ostensibly that *'the decrease in pirated software in 2002 compared with 1994 was smaller in China than in the world'*. The article itself is about Microsoft's commitment to China and the issues of pirated software. The closest the article comes to mentioning the graphic's message is: "Ninety percent of Microsoft products used in China are pirated." No comparison is ever made between piracy in China and in the world, nor the decline in piracy between 1994 and 2002.

Figure 3 shows a grouped bar chart from *Business Week*. The article itself interviews individuals who are separated from technology either because of geography or limited income. This graphic, part of a set of graphics in the article, supplements the article's text with two high-level

* This material is based upon work supported by the National Science Foundation under Grant No. IIS-0534948.

[1] Taken from *Wilmington News Journal*, 2002.

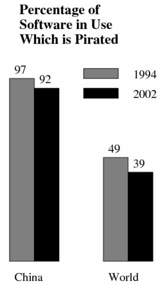

Figure 2: Graphic from *NewsWeek*, "Microsoft's Cultural Revolution", June 28, 2004.

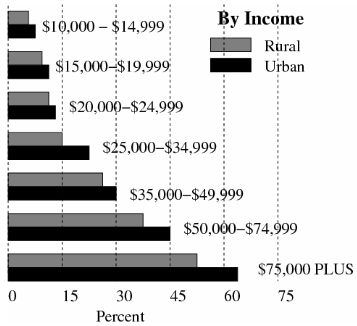

Figure 3: Graphic from *Business Week*, "A Small Town Reveals America's Digital Divide", October 4, 1999.

messages: that '*the percentage of households with internet access increases with income level*' and secondarily that *'a smaller percentage of households have internet access in rural areas than urban areas at every income level'*.

These two examples illustrate the important role of information graphics in multimodal documents. In terms of the discourse theory espoused by Grosz and Sidner (Grosz1986), the message conveyed by an information graphic constitutes the discourse goal of the graphic and, together with the discourse goals of the textual discourse segments, contributes to achieving the discourse purpose of the overall article. Thus, it is impossible to fully comprehend a multimodal document without understanding its information graphics.

We contend that the messages (both primary and secondary messages) conveyed by an information graphic should be integrated with a summary of the article's text to provide a more complete summary of the document in a digital library. Similarly, we contend that a graphic's intended message captures the high-level content of the graphic and should be used along with the article's text when extracting information from a multimodal document. For example, consider the graphic in Figure 3. The primary and secondary messages of the graphic capture the high-level information that the graphic is intended to convey. Thus they can be used as the basis for taking the graphic into account when extracting information from the document. Furthermore, using the intended message to index a graphic allows it to be efficiently retrieved to extract more detailed information in a question-answering system. For example, the secondary message of the graphic in Figure 3 can be logically represented as:

Entity-Relationship-Same(Household-internet-access,Percent,Income-levels,{Rural,Urban})

The message category *Entity-Relationship-Same* captures message types in which the bars in each group are in the same relationship (less-than, equal, greater-than), as is the case for the *Rural* and *Urban* bars in the groups in Figure 3. Based on the message category of *Entity-Relationship-Same* as well as the instantiated parameters in the logical representation of the message, a system might decide to retrieve this graphic to answer a detailed question such as *"What is the difference between rural and urban internet access at each income level?"* or *"Does the difference between rural and urban internet access increase as income changes?"*

In summary, we contend that information graphics in multimodal documents play an important role in achieving the discourse purpose of the document and cannot be ignored. Furthermore, the intended message of the graphic captures its high-level content, represents a graphic's core summary, and can be used for taking information graphics into account in extracting information from a multimodal document.

Methodology for Processing Information Graphics

Our methodology for identifying the primary intended message of an information graphic contains several components. A Visual Extraction Module first processes an electronic information graphic image and produces an XML representation of the graphic which captures all features of the graphic, such as the height and value of each bar, the color of bars, any annotations, the caption, etc., (Chester2005). A Caption Processing Module then performs a shallow processing of the graphic's caption to identify helpful verbs and nouns in the caption (discussed later) and enters them into an augmented XML representation of the graphic. A Feature Extraction Module analyzes the augmented XML representation to identify communicative signals present in the graphic and enters them as evidence in a Bayesian network. The Bayesian network then hypothesizes the message that the graphic is intended to convey.

This architecture has been implemented and tested for simple bar charts such as the graphic in Figure 1 (Elzer2005b). The system is presented with an electronic image of a graphic and outputs a logical representation of the graphic's intended message. We are now extending our methodology to recognize the primary intended message of a more complex kind of information graphic: grouped bar charts.

Grouped Bar Charts

Categories of Messages

We have collected a corpus of approximately 100 grouped bar charts from popular media. After analyzing this corpus,

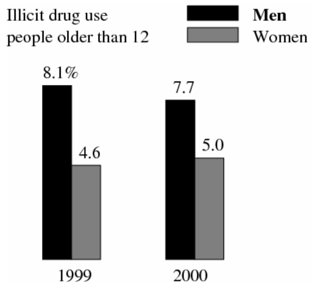

Figure 4: Graphic from *NewsWeek*, "Stop Pretending Nothing's Wrong", June 16, 2003.

we identified 14 categories of possible messages. Below is a sampling:

- **Rising-Trend-All** – Trend of values is increasing for every series of bars.
- **Trend-Contrast** – Trend in one series of bars differs from the trends in all other series
- **Rank** – Rank of a group according to one or more entities.
- **Gap-Largest** – Gap between entities in a group is the largest for some group.
- **Gap-Decreasing** – Gap between entities in a group decreases across groups.
- **Gap-Comparison** – A comparison of the gap between entities between two groups.
- **Entity-Relationship-Same** – Relation of entities in a group is the same for all groups.
- **Entity-Relationship-Contrast** – Relation of entities in one group is different from the relation of entities in all other groups.

Consider, for example, the graphic in Figure 2. The graphic's primary message (that *the percentage of pirated software in China is much higher than in the world as a whole*) falls into the **Rank** category, and the secondary message (that *the decrease in pirated software in 2002 compared with 1994 was smaller in China than in the world*) falls into the **Gap-Comparison** category. Similarly, the primary message of the graphic in Figure 3 (that *the percentage of households with internet access increases with income level*) falls into the **Rising-Trend-All** category and the secondary message (that *a smaller percentage of households have internet access in rural areas than urban areas at every income level*) falls into the **Entity-Relationship-Same** category. Although the message categories are useful for classifying the kinds of messages that can by conveyed by grouped bar charts and form the predicate for a logical representation of the conveyed message, our system must identify the full message conveyed by the graphic, not just the message category. For instance, a message involving a gap-comparison category must include which specific entities are being compared as a gap, over which specific groups.

Communicative Signals

When designing an information graphic, a designer will incorporate communicative signals to help convey an intended message. This is similar to natural language discourse, where signals such as discourse markers and intonation help the listener identify the intended meaning of an utterance. Thus a system which automatically recognizes the message conveyed by an information graphic must extract and reason about the communicative signals that appear in the graphic.

Previously, our project successfully extracted communicative signals from simple bar charts (bar charts without any concept of groups) (Elzer2005b). A graphic's caption is one source of communicative signals. Unfortunately, captions in graphics are often very general and fail to convey the graphic's intended message. Moreover, even when a caption might be helpful, it is often ill-formed or requires significant domain knowledge to process and understand. However, shallow processing of the caption can extract communicative signals about the graphic's message. As with simple bar charts, verbs in a caption can suggest the general category of message conveyed by a grouped bar chart. For example, the presence of the verb *shifts* in the subcaption of the graphic in Figure 5 suggests a message in the **Entity-Relationship-Contrast** category. Our caption processing module uses a part-of-speech tagger and a stemmer to identify the presence of one of our identified helpful verbs in a caption. (Elzer2005a)

Communicative signals in an information graphic can make certain entities in the graphic salient, and thereby suggest that they play a prominent role in the message that the graphic is intended to convey. For example, nouns in captions can assist in making a set of bars salient. Consider the graphic in Figure 4 which was part of a set of graphics whose overall caption was *"Boys Don't Cry: Men and Depression"*. The noun *men* matches a reference in the graphic's legend, thereby making the bar referring to men salient in both groups; this suggests that the intended message might be emphasizing men, perhaps in a comparison.

Sets of bars can also become salient via design choices made by the graphic designer. The position of a group or position of a bar within each group can lead to a set of bars becoming salient. For example, in Figure 5, the group "Life Sciences" is made salient by its position (first group in the graphic) as well as by its large bar values and its mention in the caption.

Coloring can also create salience, as illustrated by the graphic in Figure 6 from *Time Magazine*. Here the '04 bar in the first group is colored differently from the '04 bars in the other groups, thereby drawing attention to the increased instruction on reading in contrast with the decrease in time instructed for other subjects.

In the simple bar chart implementation, the relative effort required to perform different perceptual tasks was the communicative signal that had the greatest impact on successful recognition of an intended message

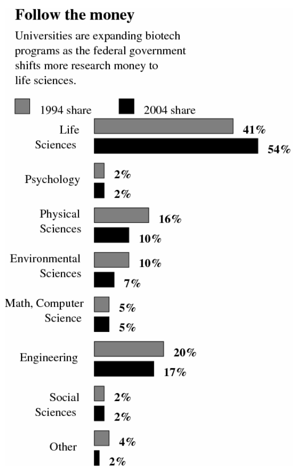

Figure 5: Graphic from *USA Today*, "Universities grid for battle for bioscience supremacy", June 24, 2005.

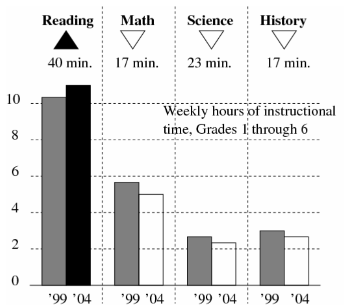

Figure 6: Graphic from *Time*, "How to Fix No Child Left Behind", June 4, 2007.

(Carberry2007). Intuitively, this makes sense, and is also supported in the literature, which posits that a graphic designer will facilitate as much as possible the tasks that a viewer must perform in order to understand a graphic (Green2004).

Effort also impacts the message conveyed by a grouped bar chart. Consider Figure 7; the two grouped bar charts were constructed from exactly the same data.[2] The message conveyed by the graphic on the left is that salaries are lower for women than men in all disciplines shown. Note that this graphic makes it easy to compare female and male salaries in each discipline. However, extracting that same message from the graphic on the right would require much more effort. In fact, a different message would most likely be inferred from the right graphic.

Most communicative signals can be extracted from an analysis of the augmented XML representation of the graphic. For example, whether an entity or group is salient by virtue of color can be determined by examining the XML representation of the graphic to determine whether they are colored differently from other entities or groups in the graphic. However, estimating relative effort is more difficult, particularly for grouped bar charts which are much more complex than simple bar charts.

In a motivational experiment with human subjects and an eye tracker, we observed how overall time, the number of fixations, and fixation durations were all affected when the number of groups, size of a graphic, density of the bars, and the complexity of a graphic was varied for graphics having an *inter*-trend message (a trend across groups, such as the graphic in Figure 3). Surprisingly, as the number of groups increased, the effort required to recognize an *inter*-trend decreased. When the number of groups was kept constant, graphics of increased size (and therefore lower density) required greater cognition for *inter*-trend recognition. Additional complexity in the graphic, such as visual clutter, or an exception contradictory to a previously witnessed pattern also added complexity. Using ACT-R (Anderson1993) and EMMA (Salvucci2001), we developed a model that takes these factors into account in estimating the relative effort required for recognizing an *inter*-trend.

Evaluation of Effort Model

We performed an eyetracking experiment to validate our model of relative effort. The experiment compared the relative effort as estimated by our model with the time taken by subjects to process the graphics. Statistical analysis produced a Spearman Rank Order Correlation Coefficient (ρ) of 0.7874 (df=17), which is statistically significant at α=0.0001.[3] Thus we conclude that our model successfully determines the relative effort for recognizing an *inter*-trend on different graphics. We are now extending our effort model to other recognition tasks in grouped bar charts.

Evaluation of Bayesian Methodology

Our methodology involves automatically extracting the communicative signals from an information graphic into a Bayesian network that then hypothesizes the intended message of the graphic. Our methodology has been implemented for simple bar charts. Using leave-one-out cross-validation, the system was shown to have an accuracy of 79.1% in recognizing the messages conveyed by a corpus of 110 simple bar charts whose messages had been previously identified by human encoders (Elzer2005b). The system was also positively evaluated by human subjects judging the correctness of posited messages, which varied between matching and differing from our system's actual inferred message (Elzer2008). We anticipate that the Bayesian network will have similar success in using the communicative signals present in grouped bar charts as evidence to hypothesize the primary message that the graphic conveys.

[2] The grouped bar chart on the left appeared in the 2000 Report of the NSF Committee on Equal Opportunities in Science and Engineering. We know its intended message since a colleague was on the panel that constructed the graphic and the report.

[3] The Spearman Rank Order Correlation Coefficient measures the correlation between two ordinal sets of data---in this case, the actual effort required by human subjects compared to the estimated effort by the model.

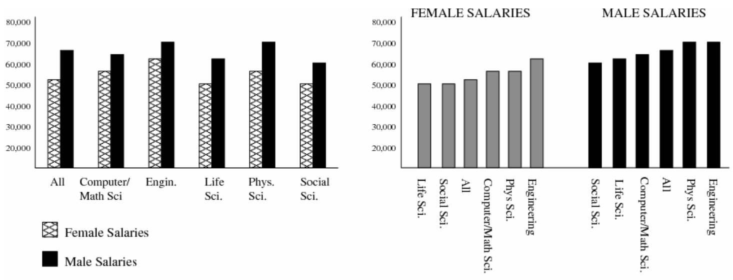

Figure 7: Two graphics which are informationally but not computationally equivalent.

Related Work

Bradshaw (Bradshaw2000) notes that work on image retrieval reasons about the semantics of the images being processed. However, this research is concerned with what is physically represented in the image and the relationships between objects whereas we are concerned with recognizing the message that an information graphic is intended to convey. Futrelle and Nikolakis (Futrelle1995) developed a constraint grammar for parsing vector-based visual displays and producing structured representations of the elements comprising the display. The goal of Futrelle's current work (Futrelle1999) is to produce a graphic that is a summary of one or several more complex graphics. Note that the end result is again a graphic, whereas our goal is to recognize a graphic's intended message so that it can be used in multimodal document summarization and information extraction.

Conclusion

This paper has shown that information graphics are an important component of a multimodal document and cannot be ignored. We contend that the message conveyed by an information graphic can serve as a brief summary of the graphic's high-level content and thus contribute to effective information extraction from a multimodal document. The paper has presented a brief overview of our methodology for recognizing the primary message of an information graphic. This methodology has been successfully implemented for simple bar charts and is now being extended to a more complex kind of information graphic: grouped bar charts. We have identified the categories of messages that are conveyed by grouped bar charts in popular media, discussed the communicative signals present in such graphics, and introduced a model of relative effort for recognizing trend messages in grouped bar charts. To our knowledge, our research is the first to address the problem of recognizing the message conveyed by an information graphic in popular media, and thus serves to advance robust summarization of and information extraction from multimodal documents.

References

Anderson, J. R. 1993. Problemsolving and learning. *American Psychologist* 48:35–44.

Bradshaw, B. 2000. Semantic based image retrieval: a probabilistic approach. In *Proceedings of the 8th ACM International Conference on Multimedia*. 167-176.

Carberry, S.; Elzer, S.; and Demir, S. 2006. Information graphics: An untapped resource of digital libraries. In *Proceedings of 9th Intl. ACM SigIR Conference*. 581-588.

Carberry, S., and Elzer, S. 2007. Exploiting evidence analysis in plan recognition. In *International Conference on User Modeling*. 7-16.

Chester, D., and Elzer, S. 2005. Getting computers to see information graphics so users do not have to. In *Intl. Sym. on Methodologies for Intelligent Systems*. 660-668.

Elzer, S.; Carberry, S.; Chester, D.; Demir, S.; Green, N.; Zukerman, I.; and Trnka, K. 2005a. Exploring and exploiting the limited utility of captions in recognizing intention in information graphics. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 223-230.

Elzer, S.; Carberry, S.; Zukerman, I.; Chester, D.; Green, N.; and Demir, S. 2005b. A probabilistic framework for recognizing intention in information graphics. *Proc. of the Intl. Joint Conf. on Artificial Intelligence*. 1042-1047.

Elzer, S.; Schwartz, E.; Carberry, S.; Chester, D.; Demir, S.; and Wu, P. 2008. Accessible Bar Charts for Visually Impaired Users. In *Fourth Annual IASTED Intl. Conf. on Telehealth and Assistive Technologies.* 55-60.

Futrelle, R. P., and Nikolakis, N. 1995. Efficient analysis of complex diagrams using constraint-based parsing. In *Proceedings of the Third International Conference on Document Analysis and Recognition*. 782-790.

Futrelle, R. P. 1999. Summarization of diagrams in documents. In Mani, I., and Maybury, M., eds., *Advances in Automated Text Summarization*. MIT Press. 403–421.

Green, N. L.; Carenini, G.; Kerpedjiev, S.; Mattis, J.; Moore, J. D.; and Roth, S. F. 2004. Autobrief: an experimental system for the automatic generation of briefings in integrated text and information graphics. *International Journal of Human-Computer Studies* 1:32–70.

Grosz, B., and Sidner, C. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12:175 – 204.

Salvucci, D. D. 2001. An integrated model of eye movements and visual encoding. *Cognitive Systems Research* 1:201–220.