Exploring the Types of Messages that Pie Charts Convey in Popular Media

Richard Burns[1], Eric Balawejder[1], Wiktoria Domanowska[1], Stephanie Elzer Schwartz[2], and Sandra Carberry[3]

[1] Dept. of Computer Science, West Chester University, West Chester, PA 19383 USA
rburns@wcupa.edu
[2] Dept. of Computer Science, Millersville University, Millserville, PA 17551 USA
[3] Dept. of Computer Science, University of Delaware, Newark, DE 19716 USA

Abstract. In popular media, information graphics (pie charts, bar charts, line graphs) are frequently used to convey high-level intended messages. This paper focuses on the pie chart graphic type. We have collected a corpus of pie chart information graphics found in popular media, and for each chart, a team of annotators recognized its intended message. In this paper, we report on the types of intended messages that the team of annotators recognized, as well as areas of disagreement. We also briefly survey some of the communicative signals that graphic designers used which helped the annotators recognize these messages. This work is a preliminary step towards developing a system that can automatically hypothesize the intended message of a pie chart.

1 Introduction

Information graphics, such as bar charts and line graphs, are common visual devices frequently incorporated into multimodal document to achieve a set of communicative goals [8] [6]. In popular media (magazines such as *Time* and newspapers such as *USA Today*), information graphics are sometimes included in an article to convey some additional, supplemental high-level message that transcends supporting data, rather than simply providing raw, low-level data points. For example, the grouped bar chart in Figure 1 ostensibly conveys a high-level message that *"Women are more likely than men to delay medical treatment"*.

The idea that information graphics can be considered a form of language follows Clark [3] who noted that language is any "signal" or lack thereof, where a signal is any deliberate action that is intended to convey a message, including gestures and facial expressions. Thus, we view information graphics as a form of language, where the designer of a graphic is able to deliberately use *communicative signals* to help convey an intended message to the viewer of the graphic.

This paper presents preliminary results in our study of designing a methodology that can automatically reason about the most likely intended message of a pie chart, using the present or absent communicative signals in the graphic as evidence.

It is non-trivial to identify the intended message of an information graphic; Carberry [2] found that a graphic's message is often not contained in the graphic's caption or in the article accompanying the graphic. Thus, the use of natural language processing techniques only on the graphic's caption or only on surrounding article text cannot be relied on to provide enough evidence to recognize the graphic's high-level message.

Communicative intent within the domain of information graphics has been studied previously by other researchers: Kerpedjiev et al. [9] proposed a methodology for automatically generating graphics that could realize desired intentions, Fasciano [5] generated graphics from the input of a communicative intention and a data set, and Mittal [10] implemented a process that automatically generated captions to explain data in graphics.

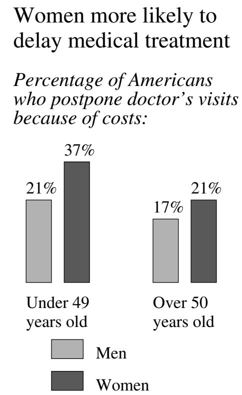

Fig. 1. From *USA Today.*

Previously, our research group has implemented intended message recognition systems for other kinds of information graphics: simple bar charts [4], line graphs [11], and grouped bar charts [1]. These three implemented systems use a Bayesian network to probabilistically capture the relationships between high-level intended messages and communicative signals that help signal the messages. Because each type of information graphic is able to convey a unique set of possible messages compared to the other information graphic types, the end-result for each of the systems has been very different. Simple bar charts, line graphs, and grouped bar charts each have a different set of message categories, and different communicative signals are utilized by graph designers to help convey the high-level intended messages.

This work is the first of our knowledge that studies the problem of recognizing the intended high-level message of a pie chart when it drawn in popular media.

We have collected a set of pie chart information graphics occurring in popular media, and examined these charts to identify (1) the types of high-level messages that graphic designers convey using pie charts, and (2) the kinds of communicative signals present in pie charts that appear likely to assist the recognition of high-level messages. Unsurprisingly, in our preliminary investigation so far, the types of recognized high-level messages and identified communicative signals are different than those in simple bar charts, line graphs, and grouped bar charts.

One application of this research is for sight-impaired individuals who cannot visualize information graphics. In best case, alternative access screen readers can convert the content of a pie chart to text, but only at the level of low-level raw data: (e.g. *"the first pie chart slice is 18.5%, the second pie chart slice is 7.3%, ..."*). Our research aims to generate the high-level message as text for sight-impaired users. A second application for this work is to use the recognized intended message of a pie chart as an indexing feature in an information retrieval system.

Section 2 of the paper introduces the types of messages categories that we identified for the pie charts that we collected. Section 3 describes the current progress of our research, and the additional steps necessary to develop a system that can hypothesize the intended message of a new, previously unseen pie chart. Finally, Section 4 briefly presents some unexpected properties of pie charts in popular media that could be avenues for interesting future work.

2 Pie Charts

We collected XX pie chart information graphs from popular media.[4]. Of those, we retained 91 of the charts, as the rest appeared to contain *only* data, and did not appear to us to convey any intended message. (Inter-annotator agreement is discussed later.) We then analyzed the corpus to generalize the kinds of high-level intended messages that we recognized into *message categories*. This section describes and presents examples of some of our identified pie chart message categories.

2.1 Message Categories

There are nine pie chart message categories that we defined. Of the 91 pie charts in the corpus, XX have intended messages that fit into these below categories. In this section, we formally define the name of the category, the number of parameters that the category takes, and provide a short description. Because of space constraints, we only present a subset of the message categories.

SingleSlice($< s >$). Single slice messages recognize a high-level message that involves a single, salient, pie chart slice. Generally, the pie charts that fall within this category seem designed so that the graph viewer compares a specific, single slice against the other slices in the pie chart. For example, consider the pie chart in Figure 2. This pie chart ostensibly conveys that *Landfills are a significant source of U.S. methane emissions, the third highest, behind the natural gas and petroleum industry as well as animal digestion.* The parameter $< s >$ in the message category syntax is instantiated with the single pie chart slice that is to be compared against the other slices. That is, this message would be represented as: *SingleSlice*($s = Landfills$).

Versus($< s_1, s_2 >$). Versus messages captures two salient slices, which are compared against each other. In contrast to single slice messages in which a salient pie chart slice is compared against the rest of the slices in the pie chart, the two salient slices in versus messages are compared with each other rather than the other slices. For example, the pie chart in Figure 3 ostensibly conveys the message that *most prisoners were turned over to coalition forces because of bounties, rather than being captured by troops.* The versus message category is

[4] The corpus of pie charts is available at: http://www.cs.wcupa.edu/∼rburns/piecharts

figs/singleslice-eps-converted-to.pdf

Fig. 2. Graphic from *National Geographic*.

figs/verses-eps-converted-to.pdf

Fig. 3. Graphic from *Time Magazine*.

instantiated with two parameters: s_1 and s_2, the slices that should be compared with each other.

BiggestSlice(). Biggest slices messages identify a single slice of the pie chart that is larger than all of the other slices. Because only one slice can be the largest (assuming no ties), the biggest slice message category has no parameters. For example, presumably the intended message in the pie chart in Figure 4 is that *there were a greater number of male deaths than female deaths in which illicit fentanyl was detected*.

figs/biggest-eps-converted-to.pdf

Fig. 4. Graphic from *The Philadelphia Inquirer*.

figs/nomajority-eps-converted-to.pdf

Fig. 5. Graphic from *The Philadelphia Inquirer*.

NoMajority(). Finally, no majority messages capture that none of the slices in the pie chart are larger than 50%. Like the biggest slice message category, the no majority message category also has zero parameters. For example, the pie chart in Figure 5 ostensibly intends to convey the high-level message that *individuals in search of work take a variable range in time in order to find a job*.

2.2 Annotation and Inter-Coder Agreement

The annotation of the corpus was performed with the following process: we first individually recognized the intended message for each pie chart and classified it into its appropriate intended message. Then, we conducted a consensus-based annotation by meeting as a group and discussing each of our annotations, revising any annotations if we were strongly swayed. The final annotation for each pie chart was decided by majority vote.

To date, we have met and deliberated final annotations for 32 of the 91 pie charts in the corpus. Notably, there was some disagreement amongst our annotations. XX of XX were ....

3 Implementing a System for Recognizing the Intended Message of a Pie Chart

Our current research is focused on implementing a complete, automatic system that is able to hypothesize the intended message of a new, previously unseen pie chart. The implementation follows a Bayesian methodology: a hand-constructed Bayesian network learns the probabilistic relationship between pie chart intended messages and the communicative signals that are present or absent in them.

3.1 Communicative Signals

Communicative signals can assist the recognition of pie chart intended messages.

Visual Signals. One visual signal that a graphic designer may use to help communicate some intended message is prominence, by coloring a specific pie chart slice a salient coloring, or boldfacing the label of a pie chart slice. An example of this communicative signal is present in Figure 2, which helps signal that *Landfills* should be compared against the other pie chart slices. Another example of a visual signal found in the pie chart corpus is the use of similar colors across multiple pie chart slices. For example in Figure 3, the slices for *Bounty* and *Troops* are colored similarly (though not exactly identical), helping signal that they should be compared, while still contrasting them against the *Unlabeled 9%* slice.[5]

Linguistic Signals. Although it does not always fully capture a graphic's intended message, the caption text of a pie chart can sometimes serve as a

[5] In the original graphic, *Bounty* is colored yellow, *Troops* is orange, and the unlabeled slice is gray.

linguistic signal that helps convey its message. For example, in the pie chart in Figure 6, the verb *split* helps signal the intended message that *there is no majority slice amongst the slices: "will", "will not", and "unsure"*. We have also observed instances of the article headline of a multimodal article helping to signal the intended message of a pie chart.

3.2 Towards a Bayesian Implementation

We are currently constructing a Bayesian network, which has a top-level node with states that enumerate all possible pie chart messages. This top-level node is linked to children leaf nodes that represent the possible communicative evidence in a graphic. Given our corpus of pie chart graphics, we will train the network to learn the probabilistic relationships between pie chart high-level intended messages and the communicative evidence that is present or absent in the charts.

Visual extraction of communicative signals in the graphic is possible with a system similar to Huang et al. [7], which identifies an information graphic within a noisy pdf document, performs OCR on the text within the graphic, and represents the semantic structure of information graphics (heights of the bars, sizes of the pie chart slices, etc.). Caption text and article headlines can be parsed and post-processed with natural language processing techniques to identify the presence of any signal verbs (such as the verb *"split"* in Figure 6) as well as analyze the clausal structure if multiple pie chart slice entities are present in the caption/headline.

After the communicative evidence of a new pie chart is automatically extracted and entered into the network, the posterior beliefs of the network can then be used to hypothesize the most likely intended message of the graphic.

4 Conclusion

Future Work. There are a couple of avenues of interesting future work that we are exploring. First, we have observed numerous instances of *multiple pie charts* drawn adjacent to one another, where the single intended message of the graphic seems to involve *both* pie charts, rather than two individual and separate intended messages. For example in the multiple pie charts shown in Figure 7, the high-level message conveyed is that *the percentage of births to unmarried U.S. women 35 and older increased from 1990 to 2008*. This avenue of future work explores the unique types of messages and communicative signals that can be found when multiple pie charts are purposely drawn adjacent to each other.

Another area of future work is predicting when there is a high-level message in a pie chart that is intended to be conveyed, rather than when a pie chart is only displaying data. Although our investigation is limited to the domain of popular media rather than scientific text, pie charts appear more likely to sometimes only contain data, compared to bar charts and line graphs.

Summary. In this paper, we have presented novel research that introduces (1) a corpus of pie charts that we have collected from popular media, (2) a figs/split-eps-converted-to.pdf

Fig. 6. Graphic from *USA Today.*

figs/mothering-eps-converted-to.pdf

Fig. 7. Graphic from *???*.

sampling of the types of messages that pie charts are able to convey, and (3) examples of communicative signals that help communicate these messages. These identified messages and communicative signals are unique compared to other types of information graphics—notably bar charts and line graphs—and can be used in a system that automatically hypothesizes the intended message of a pie chart.

References

1. Burns, R., Carberry, S., Elzer, S., Chester, D.: Automatically recognizing intended messages in grouped bar charts. In: Proceedings of the Seventh International Conference on the Theory and Application of Diagrams. pp. 8–22. Springer-Verlag, Berlin, Heidelberg (2012)
2. Carberry, S., Elzer, S., Demir, S.: Information graphics: An untapped resource of digital libraries. In: Proceedings of 9th International ACM SigIR Conference on Research and Development on Information Retrieval. pp. 581–588. ACM, New York, NY (2006)
3. Clark, H.: Using Language. Cambridge University Press (1996)
4. Elzer, S., Carberry, S., Zukerman, I.: The automated understanding of simple bar charts. Artificial Intelligence 175(2), 526–555 (February 2011)
5. Fasciano, M., Lapalme, G.: Intentions in the coordinated generation of graphics and text from tabular data. Knowledge and Information Systems 2(3), Springer London (August 2000)
6. Green, N.L., Carenini, G., Kerpedjiev, S., Mattis, J., Moore, J.D., Roth, S.F.: Autobrief: an experimental system for the automatic generation of briefings in integrated text and information graphics. International Journal of Human-Computer Studies 61(1), 32–70 (2004)
7. Huang, W., Tan, C.L.: A system for understanding imaged infographics and its applications. In: Proceedings of the 2007 ACM symposium on Document engineering. pp. 9–18. DocEng '07, ACM, New York, NY, USA (2007)

8. Iverson, G., Gergen, M.: Statistics: The Conceptual Approach. Springer-Verlag, New York (1997)
9. Kerpedjiev, S., Green, N., Moore, J., Roth, S.: Saying it in graphics: from intentions to visualizations. In: Proceedings of the Symposium on Information Visualization. pp. 97–101. InfoVis '98, IEEE, Research Triangle Park, NC, USA (1998)
10. Mittal, V.O., Carenini, G., Moore, J.D., Roth, S.: Describing complex charts in natural language: A caption generation system. Computational Linguistics 24(3), 431–467 (September 1998)
11. Wu, P., Carberry, S., Elzer, S., Chester, D.: Recognizing the intended message of line graphs. In: Proceedings of the 6th International Conference on Diagrammatic Representation and Inference. pp. 220–234. Diagrams' 10, Springer-Verlag (2010)