Artificial Intelligence 175 (2011) 526–555

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

The automated understanding of simple bar charts

Stephanie Elzer[a,*], Sandra Carberry[b], Ingrid Zukerman[c]

[a] *Department of Computer Science, Millersville University, P.O. Box 1002, Millersville, PA 17551, USA*
[b] *Department of Computer & Information Sciences, University of Delaware, 103 Smith Hall, Newark, DE 19716, USA*
[c] *Faculty of Information Technology, Monash University, Clayton, Victoria 3800, Australia*

ARTICLE INFO

Article history:
Received 2 October 2009
Received in revised form 8 October 2010
Accepted 11 October 2010
Available online 15 October 2010

Keywords:
Natural language processing
Bayesian networks
Bar charts
Communicative signals

ABSTRACT

While identifying the intention of an utterance has played a major role in natural language understanding, this work is the first to extend intention recognition to the domain of information graphics. A tenet of this work is the belief that information graphics are a form of language. This is supported by the observation that the overwhelming majority of information graphics from popular media sources appear to have some underlying goal or intended message. As Clark noted, language is more than just words. It is any "signal" (or lack of signal when one is expected), where a signal is a deliberate action that is intended to convey a message (Clark, 1996 [15]).

As a form of language, information graphics contain communicative signals that can be used in a computational system to identify the message that the graphic conveys. We identify the communicative signals that appear in simple bar charts, and present an implemented Bayesian network methodology for reasoning about these signals and hypothesizing a bar chart's intended message. Once the message conveyed by an information graphic has been inferred, it can then be used to facilitate access to this information resource for a variety of users, including 1) users of digital libraries, 2) visually impaired users, and 3) users of devices where graphics are impractical or inaccessible.

1. Introduction

Information is the key to knowledge and effective decision-making. As more and more information becomes available electronically, the population requiring access to the information has grown and the ways in which we find and obtain information have expanded. It is crucial that we develop techniques for providing effective access to the vast electronic resources so that the information is readily available when needed and so that all individuals can benefit from these resources.

Information graphics such as bar charts, line graphs and pie charts are an important component of many documents. As noted by [39,31], a set of data can be presented in many different ways, and graphs are often used as a communication medium or rhetorical device for presenting a particular analysis of the data and enabling the viewer to better understand this analysis.

The rapidly increasing availability of these important resources electronically poses some interesting challenges in terms of access to the information contained in information graphics. For example, information graphics pose a problem when attempting to search the content of mixed-media publications within digital libraries. The searchable index of such documents should include not only the important content of the text, but also of the information graphics contained in the document. In addition, individuals with impaired eyesight have limited access to graphical displays, thus preventing them from fully utilizing the available electronic resources. In order to provide effective access to information graphics, there is a need for

[*] Corresponding author. Tel.: +1 717 872 3470; fax: +1 717 872 3149.
E-mail addresses: elzer@cs.millersville.edu (S. Elzer), carberry@cis.udel.edu (S. Carberry), Ingrid.Zukerman@monash.edu (I. Zukerman).

0004-3702/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.artint.2010.10.003

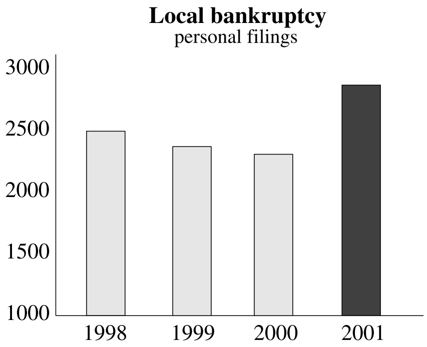


Fig. 1. Graphic from a 2001 local newspaper.

tools that convey their content in other modalities. This article presents an important first step for such a tool — namely, an implemented and evaluated system for recognizing the intended message of simple bar charts.

A tenet of this work is the belief that information graphics are a form of language, and therefore contain communicative signals that can be used in a computational system to identify the message that the graphic conveys. Consider the graphic in Fig. 1 — despite the lack of accompanying text, it conveys the message that there has been a sharp increase in local bankruptcies in 2001 compared with the previously decreasing trend. In this article, we identify the communicative signals that appear in simple bar charts (plus the accompanying caption, if any), and present an implemented and evaluated Bayesian network methodology for reasoning about these signals and hypothesizing a graphic's intended message. Once the message conveyed by an information graphic has been inferred, it can then serve as the basis for an effective summary. This summary could then be used in a variety of ways. For digital libraries, the summary of the information graphic could be used to appropriately index the graphic and to enable intelligent retrieval. If there is accompanying text, the summary of the graphic can be used in conjunction with a summary of the document's text to provide a more complete representation of the document's content. For individuals who are sight-impaired, the core message of the information graphic can be used as the basis for a summary conveyed via speech, thereby providing access to the informational content of the graphic in an alternative modality. Rather than providing alternative access to what the graphic looks like or a listing of all of the data points contained in the graphic, our approach would provide the user with the message and high-level knowledge that one would gain from viewing the graphic.

In addition to producing a methodology for identifying the intended message of a simple bar chart, this paper has the objective of relating information graphics to more standard forms of language by

- identifying the kinds of communicative goals that are achieved by designers of simple bar charts and the kind of communicative signals that help achieve these goals;
- computationalizing the extraction of communicative signals from simple bar charts;
- showing how plan recognition techniques, which have been successfully used to recognize the communicative goals of natural language utterances, can be extended to the recognition of the intended message of a simple bar chart.

In doing so, our research not only provides a system that will be useful in summarizing information graphics, but it suggests that other forms of language, such as advertisements that incorporate both text and illustrations, as well as other kinds of graphics, may also be amenable to such treatment.

2. Information graphics as language

Although some information graphics are only intended to display data, the majority of information graphics that appear in popular media such as newspapers and magazines are intended to convey a message. Newspapers often contain information graphics that are stand-alone and constitute the entire document. Even when an information graphic appears on its own, it can be used to convey a message to the viewer (see Fig. 1, for example). In other cases, information graphics appear as part of a document. When this occurs, the message conveyed by the graphic is often not included in the document's text. The author has some communicative intention or purpose in constructing the document, and makes the choice to convey some of the information supporting that intention in an information graphic. Thus the graphic generally expands on the text and contributes to the overall discourse purpose [33] of the document. For example, Fig. 2 illustrates a graphic from an article in *Newsweek* showing that the income of black women in the United States has risen dramatically over the last decade and has reached the level of the income of white women. Although this information is not conveyed elsewhere in the article, it contributes to the overall communicative intention of this portion of the article — namely, that there has been a "monumental shifting of the sands" with regard to the achievements of black women.

The observation that the overwhelming majority of information graphics from popular media sources appear to have some underlying goal or intended message leads us to consider information graphics as a form of language. As Clark noted,

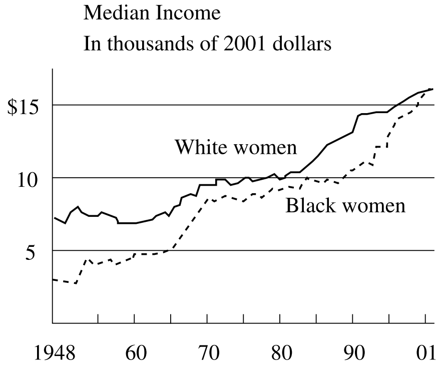


Fig. 2. Graphic from Newsweek magazine.[1]

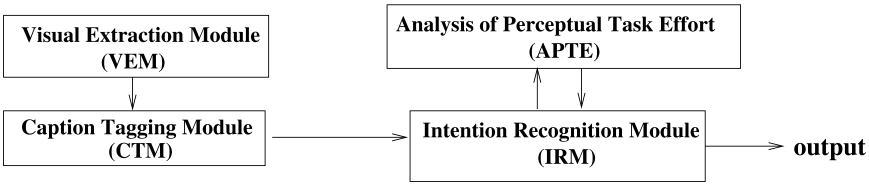

Visual Extraction Module (VEM)
Caption Tagging Module (CTM)
Analysis of Perceptual Task Effort (APTE)
Intention Recognition Module (IRM)
output

Fig. 3. System architecture.

language is more than just words. It is any "signal" (or the lack of a signal when one is expected), where a signal is a deliberate action that is intended to convey a message [15]. Clark's expanded definition of language includes forms of communication such as gesture, facial expression, eye gaze, and so forth. The common factors among these varied forms of expression are the communicative intention underlying them and the presence of deliberate signals to aid in recognizing these intentions.

Research on understanding utterances has posited that a speaker or writer executes a speech act whose intended meaning he expects the listener to be able to deduce, and that the listener identifies the intended meaning by reasoning about the observed signals and the mutual beliefs of author and interpreter [32,15]. Applying Clark's view of language to information graphics, it is reasonable to presume that the author of an information graphic similarly expects the viewer to deduce from the graphic the message that he intended to convey by virtue of communicative signals that are present in the information graphic. Furthermore, a poorly designed graphic may convey a different message from what the graphic designer intended, just as inappropriate words, intonation, etc. may result in misinterpretation of a speaker's utterance.

The design choices made by the designer when constructing the information graphic provide the communicative signals necessary for understanding the graphic. The design choices include selection of graphic type (bar chart, line graph, pie chart, etc.), organization of information in the graphic (for example, the order of bars in a bar chart), and attention-getting devices that highlight certain aspects of a graphic (such as coloring one bar of a bar chart differently from the others, mentioning data elements in the caption, etc.).

3. System architecture

Fig. 3 shows the overall architecture for processing an information graphic. The Visual Extraction Module (VEM) is responsible for analyzing the graphic's image file (currently a .gif) and producing an XML representation containing information about the components of the information graphic including the graphic type (bar chart, pie chart, etc.) and the caption of the graphic. For a bar chart, the representation includes the number of bars in the graph, the labels of the axes, and information for each bar such as the label, the height of the bar, the color of the bar, and so forth [13]. The XML representation is then passed to the Caption Tagging Module (CTM) which extracts information from the caption and passes the augmented XML representation to the Intention Recognition Module (IRM). The IRM is responsible for recognizing the intended message of the information graphic, which we hypothesize can serve as the basis for an effective summary of the graphic. The IRM interacts with the Analysis of Perceptual Task Effort (APTE) Module to obtain estimates of task effort as a source of evidence during the intention recognition process.

The remainder of this paper focuses on intention recognition from simple bar charts. By simple bar charts, we mean bar charts that display the values of a single independent attribute and the corresponding values for a single dependent attribute. Fig. 1 is an example of a simple bar chart. Although simple bar charts constitute only a subset of the full range

[1] Taken from an article entitled The Black Gender Gap, Newsweek, March 3, 2003.

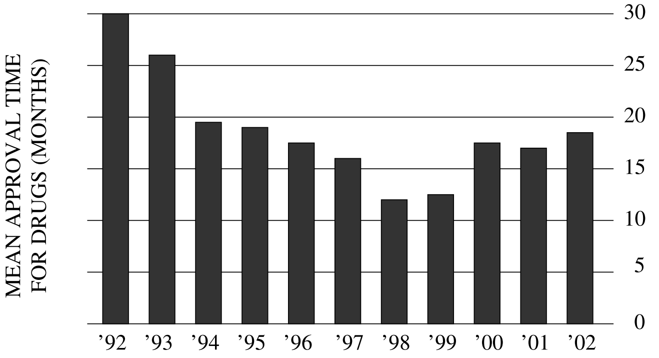


Fig. 4. Bar chart showing the mean approval time for drugs.[2]

of information graphics, we believe that the concepts, mechanisms and framework of our methodology are broadly applicable and extensible to other types of graphics. In fact, Section 12 describes work currently underway to extend this basic methodology to line graphs and grouped bar charts.

Section 4 categorizes the kinds of messages that can be conveyed via simple bar charts. Section 5 discusses the kinds of communicative signals that appear in simple bar charts and our mechanisms for extracting them. Section 6 discusses plan inference in language understanding and its application to information graphics; Sections 7 and 8 then describe the structure of our Bayesian network for recognizing the intended message of a simple bar chart and how the network is built. Section 9 presents two evaluations of our implemented system, and Section 10 presents several variations of a simple bar chart that illustrate how the presence of different communicative signals effect the message recognized by our system. Finally, Section 11 presents related research while Section 12 describes our current and future work.

4. An annotated corpus of simple bar charts

We collected a corpus of 110 simple bar charts from various publications, including business and news-oriented magazines such as Newsweek, BusinessWeek, Time, Fortune and Money, as well as local and national newspapers. We also constructed a list of high-level message schemas that we believed were likely to capture the graphic designer’s intended message for simple bar charts. This list included categories such as conveying trends in the data (increasing, decreasing or stable), presenting the element with the maximum or minimum value in a data set, and so forth. We included a category called “Present-Data” which was meant to capture those cases where there was no apparent message being conveyed by the bar chart (the bar chart seemed to simply present data with no underlying message). Two coders, given a bar chart along with any existing caption, then identified 1) the intended message of each graphic using the provided list of possible message schemas, and 2) the instantiation of the parameters in the schema. For example, if the coder classified the intended message of the graphic shown in Fig. 4 as *Change-Trend*, the coder was also asked to identify where the first trend began, its general slope (increasing, decreasing, or stable), where the change in trend occurred, the end of the second trend, and the slope of the second trend. The coders were also given the option of creating a new message schema if they felt it was warranted.

Each coder independently classified the intended message and determined the instantiation of the parameters for all of the bar charts in the corpus. If there was disagreement between the coders on either the intention or the instantiation of the parameters, we utilized consensus-based annotation [4], in which the coders discussed the graphic to try to come to an agreement as to its intention and the appropriate parameters. As observed by Ang et al. [4], this allowed us to include the “harder” or less obvious graphics in our study, since the coders were likely to immediately agree upon the bar charts where the underlying message was clear, but might only agree on the bar charts where the message is more subtle after some discussion. Including these less obvious graphics in our corpus may lower our system performance when it comes to evaluation, but we have a richer and more complete set of data included in those evaluations.

Table 1 presents the twelve categories of high-level messages that were assigned as the intentions of the bar charts in our corpus, as well as their distribution within our corpus (as identified by the coders). Ten of these message categories appeared in the list of possible messages that was provided to the coders at the beginning of the tagging process, while two of the categories, *Rank-All* and *Contrast-Pt-Trend*, were added by the coders.

5. Communicative signals in simple bar charts

As discussed in Section 2, we view information graphics in popular media as a form of language with communicative signals that help convey the graphic’s intended message. These signals can appear within the graphic itself as well as in the

[2] This graphic is based on a bar chart from Money magazine’s April 2003 issue.

Table 1
Categories of high-level intentions.

Intention	Description	Distribution
Get-Rank	Viewer to believe that $\langle param_1 \rangle$ is ranked $\langle rank \rangle$ among the elements in the graphic	3.6%
Rank-All	Viewer to believe that the elements in the graph have an ordering $\langle element_1 \ldots element_n \rangle$	9.1%
Increasing-Trend	Viewer to believe that there is an increasing trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$	23.6%
Decreasing-Trend	Viewer to believe that there is a decreasing trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$	12.7%
Stable-Trend	Viewer to believe that there is a stable trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$	0%
Change-Trend	Viewer to believe that there is a $\langle slope_1 \rangle$ trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$ and a significantly different $\langle slope_2 \rangle$ trend from $\langle param_2 \rangle$ to $\langle param_3 \rangle$	6.4%
Contrast-Pt-Trend	Viewer to believe that there is a $\langle slope_1 \rangle$ trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$ and that the value of subsequent element $\langle param_3 \rangle$ contrasts with this trend	10.9%
Relative-Difference	Viewer to believe that the value of element $\langle param_1 \rangle$ is $\langle comparison \rangle$ the value of element $\langle param_2 \rangle$, where $\langle comparison \rangle$ is greater-than, less-than, or equal-to	0%
Relative-Difference-Degree	Viewer to believe that the value of element $\langle param_1 \rangle$ is $\langle comparison \rangle$ the value of element $\langle param_2 \rangle$, where $\langle comparison \rangle$ is greater-than, less-than, or equal-to, and the $\langle degree \rangle$ of that difference is large, medium, or small	5.5%
Maximum	Viewer to believe that $\langle param_1 \rangle$ has the largest value among the entities in the graphic	22.7%
Minimum	Viewer to believe that $\langle param_1 \rangle$ has the smallest value among the entities in the graphic	3.6%
Present-Data	Graphic simply presents data with no underlying message	1.8%

graphic's caption. Section 5.1 discusses the types of communicative signals present in the graphic itself (such as highlighting, annotations, and perceptual task effort), while Section 5.2 discusses the communicative signals present in captions.

5.1. Communicative signals in the graphic itself

5.1.1. Salience

Many researchers have examined ways in which graphic designers might make particular elements of a graphic salient to the viewer; much of this work has been done for the purpose of improving the design of graphics (for example, [16] and [45], among others), while [51] considered some of the design decisions that make elements of the graph salient for the purpose of biasing viewer inferences. Our contention is that if the graphic designer goes to the effort of employing attention-getting devices to make certain elements of the graphic particularly salient, then the salient elements serve as communicative signals — i.e., the designer probably intends for them to play a prominent role in the intended message of the graphic. We have identified several of the most common design techniques that graphic designers employ to increase the salience of an element or elements in simple bar charts.

In order to draw attention to a particular element (or elements) of a bar chart, the graphic designer may choose to *highlight* it. Graphic designers typically highlight an element or elements of a bar chart by drawing the viewer's attention to the bar itself or to an attribute of the bar, such as its label or its annotated value. For example, the designer could highlight a bar in the graphic by making it a different color, shade or texture than the other bars. This is a communicative signal, conveying to the viewer of the graphic that the bar (and thus the data element that it represents) is of significant import to the graphic's message. Consider the graphic in Fig. 1 which appeared in a local newspaper. The graphic appeared in shades of gray, as it is depicted here, with the bar representing 2001 in a darker shade than the other bars. The design choice to highlight this bar by making it a darker shade of gray than the other bars seems to signal the importance of the bankruptcy rate in 2001 to the message that the designer is attempting to convey with the graphic — ostensibly, that there has been a sharp increase in local bankruptcies in 2001 compared with the previously decreasing trend (a *Contrast-Pt-Trend* message).

A graphic designer can also convey the significance of an element in an information graphic by annotating the salient element in some way. The most common form of annotation in our corpus of information graphics is the annotation of an element with its exact value. Annotating individual elements with their exact values can signal salience if the annotations are *not* a general design feature of the graphic. If all of the elements are displayed with their exact values, then we consider this to be a general design feature of the graphic since the annotations do not draw attention to a specific subset of elements. However, if only a subset of the elements are annotated with their values, the annotations signal the salience of those elements. This is the case in Fig. 5 where only the first and last elements are annotated with their values. Annotations of bars in a bar chart are not limited to the exact value represented by the bar — they can also include content such as dates or other additional notes.

We have also identified several factors that increase the salience of an element in a graphic without the application of any particular design techniques. Although no specific action is required on the part of the graphic designer to make these elements salient, we posit that it is mutually believed by both designer and viewer that such elements will be salient to the viewer. These elements include any element that is significantly taller than all of the other elements in the graphic and the most recent date on a time-line, since the viewer will certainly notice the height of a bar that is much taller than all of the others, and will naturally be interested in what has occurred most recently.

In our architecture, the Visual Extraction Module (VEM) is responsible for producing an XML representation of a bar chart that includes information about each bar such as bar height, annotations, color, and so forth. By analyzing this XML rep-

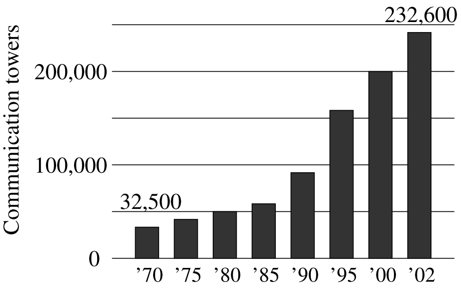


Fig. 5. Bar chart showing the number of communication towers.[3]

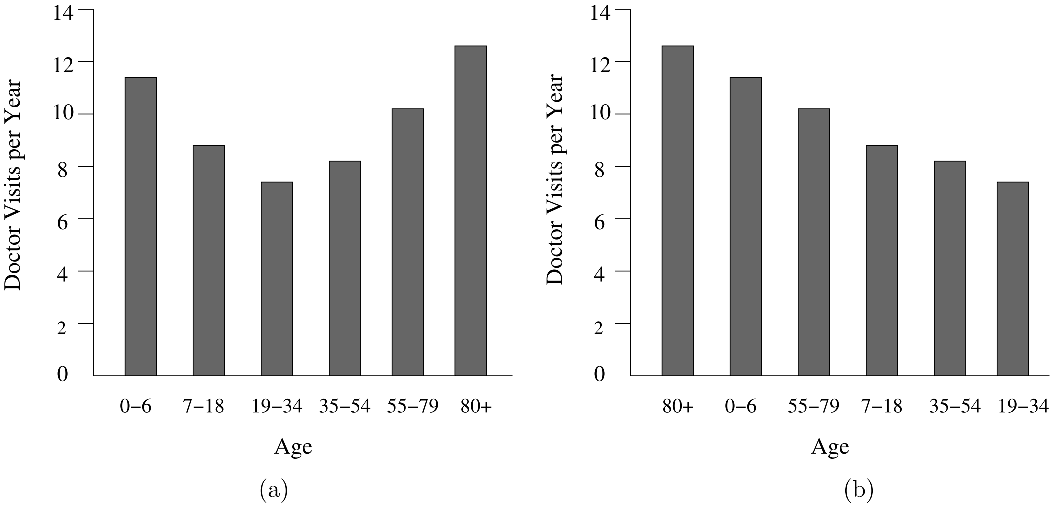


Fig. 6. Two alternative graphs from the same data on doctor visits by age.[4]

resentation, our system is able to identify any particularly salient elements of the graphic according to the aforementioned criteria for salience.

5.1.2. Perceptual task effort

Given a set of data, the graphic designer has many alternative ways of designing a graphic. In their work on automated graph generation, Fasciano and LaPalme [23] showed that the writer's intentions affect the appropriate design of a good graph. Larkin and Simon note that information graphics that are *informationally* equivalent (all of the information in one graphic can also be inferred from the other) are not necessarily *computationally* equivalent (enabling the same inferences to be drawn quickly and easily) [47].

Consider, for example, the two alternative graphs shown in Fig. 6, which have been constructed from the same set of data. The leftmost graphic orders the bars by consecutive age group. This organization facilitates a comparison of the number of doctor visits per year by consecutive age classes and thus conveys a changing trend in the frequency of doctor visits over the average person's lifetime (see Fig. 6a). The same data could also be presented in a bar chart with the bars ordered by height as shown in Fig. 6b. This organization does not facilitate the same comparisons as the graphic in Fig. 6a and thus does not convey the same Change-Trend message.[5]

Peebles and Cheng [56] further observe that even in graphics that are informationally equivalent, seemingly small changes in the design of the graphic can affect viewers' performance on graph reading tasks. Much of this can be attributed to the fact that design choices made while constructing an information graphic will facilitate some perceptual tasks more than others. The AutoBrief project considered the generation of graphics to achieve communicative goals, and posited

[3] This is based on a bar chart from the newspaper USA Today.

[4] The data displayed in this graphic is for illustration purposes only and is not based on factual findings.

[5] Although it might seem odd to order bars by height instead of by an ordinal label such as age, such graphics do occur. For example, a graphic from the U.S. Department of Labor, posted on the CNN web site and entitled "*Worst Annual Job Losses*", ordered the bars by number of jobs lost (bar height) rather than by their labels (year of occurrence). Similarly, a bar chart in USA Today ordered Steven Spielberg movies by revenue (bar height) rather than by their label (year of release).

```
Rule-B1: Estimate effort for task
         PerceiveValue(⟨viewer⟩, ⟨g⟩, ⟨att⟩, ⟨t⟩, ⟨v⟩)

Graphic-type: bar-chart

Gloss: Compute effort for finding the exact value ⟨v⟩ for attribute ⟨att⟩
         represented by top ⟨t⟩ of a bar ⟨b⟩ in graph ⟨g⟩

Conditions:
    B1-1: IF the top ⟨t⟩ of bar ⟨b⟩ is annotated with a value,
            THEN effort = 150 + 300
    B1-2: IF the top ⟨t⟩ of bar ⟨b⟩ aligns with a labelled tick mark on the dependent axis,
            THEN effort = 230 + (scan + 150 + 300) × 2
```

Fig. 7. APTE rule for estimating effort for the perceptual task *PerceiveValue*.

that the graph designer chooses a design that best facilitates the tasks that are most important to conveying his intended message, subject to the constraints imposed by competing tasks [43,31]. Thus we contend that the relative difficulty of different perceptual tasks serves as a communicative signal about which tasks the viewer was intended to perform in recognizing the graphic's intended message — i.e., the greater the effort required for a task, the less likely the task is to be part of the perceptual tasks needed to identify the graphic's message.

Computationalizing perceptual task effort. In order to reason about which tasks are easier to perform within a given information graphic, we needed to be able to estimate the effort of each perceptual task so that the relative effort could be directly compared against the estimated effort of other tasks. Therefore, we constructed a set of rules, encapsulated as a module called APTE (Analysis of Perceptual Task Effort), that estimate the relative effort required for different perceptual tasks within a given information graphic. Each rule consists of a set of condition–computation pairs; the conditions are examined sequentially, and the computation associated with the first satisfied condition in the graphic is used to estimate the effort for that perceptual task. To develop these rules, we applied the results of research from cognitive psychology.

We adopted a GOMS-like approach [11] to estimate the relative effort involved in performing a task, decomposing each task into a set of component tasks. Following other cognitive psychology research (e.g. [48]), we take the principal measure of the effort involved in performing a task to be the amount of time that it takes to perform the task, and our effort estimates are based on time estimates for the component tasks.[6] Wherever possible, we utilize existing time estimates (primarily those applied in Lohse's UCIE system [48]) for the component tasks; otherwise we conducted eye tracking experiments to acquire the necessary estimates [21,22].

For example, the rule shown in Fig. 7 estimates the effort required to determine the exact value represented by the top of a bar in a bar chart, given that the viewer is already focused on the top of the bar.[7] In the case of condition–computation pair B1-1 (finding the exact value for a bar where the bar is annotated with the value), the effort is estimated as 150 units for discriminating the label (based on work by Lohse [48]) and 300 units for recognizing a 6-letter word [40]. In the case of B1-2 (finding the exact value for a bar where the top of the bar is aligned with a tick mark on the dependent axis), the effort estimate includes scanning over to the dependent axis (measured in terms of distance in order to estimate the degrees of visual arc scanned [44]) in addition to the effort of discriminating and recognizing the label. Our eye tracking experiments showed that when the top of the bar is aligned with a tick mark, participants frequently repeat the task of scanning to the axis and reading the label (presumably to ensure accuracy), so our effort estimate also includes 230 units [63] to perform a saccade back to the top of the bar before repeating the task. Our set of APTE rules for estimating the effort of tasks in bar charts and our eye tracking experiments that validated those rules are described in [21,22].

The evidence provided by perceptual task effort can sometimes be subsumed (probabilistically) by other communicative signals, such as a bar being highlighted or the presence of helpful words in a caption. However, perceptual task effort is a very important communicative signal in graph understanding since, in the absence of all other evidence, we always have communicative signals provided by perceptual task effort. For example, an information graphic might not have any salient elements and may be lacking a helpful caption, but it is still possible to reason about the relative ease or difficulty with which tasks can be performed on the graphic and to thereby draw useful inferences about the intended message of the graphic designer.

5.2. Communicative signals in captions

One might suggest relying on a graphic's caption to identify its primary message. However, Corio and LaPalme [18] conducted a corpus study whose objective was to categorize the kinds of information contained in captions in order to form rules for generating captions to accompany graphics; they noted that captions are often missing or very general.

[6] The units of effort estimated by our rules roughly equate to milliseconds.

[7] *Rule B1* does not estimate the effort required to get the value represented by the top of a bar in the case where the viewer must scan to the axis and interpolate an estimated value. This task is represented by a separate rule in our system.

Table 2
Analysis of 100 captions on bar charts.

Category	#
Category-1: Captures intention (mostly)	34
Category-2: Captures intention (somewhat)	15
Category-3: Hints at intention	7
Category-4: No contribution to intention	44

We conducted our own corpus study 1) to identify how well the intended message of a bar chart appearing in popular media was captured by the graphic's caption, and 2) to determine how easily a general-purpose natural language system could understand such captions [20]. We analyzed the first 100 graphics from our corpus of bar charts, each of which had previously been annotated with its intended message. We examined the caption of each graphic and classified it into one of four categories based on the extent to which the caption captured the graphic's intended message. Table 2 shows the results. Slightly more than half of the captions (51%) were judged to convey none of the graphic's intended message or to only hint at it. An example was the caption "*The Sound of Sales*" that appeared on a bar chart conveying a changing trend in record album sales. Furthermore, for the 49 captions that conveyed at least some of the graphic's message, we found that almost half were fragments (for example, "*A Growing Biotech Market*") or involved some other form of ill-formedness (such as the caption "*Running tops in sneaker wear in 2002*"), and that 16% would require extensive domain knowledge or analogical reasoning to understand (such as the caption "*Bad Moon Rising*" where *bad moon* was a reference to something undesirable, in this case *delinquent debts*).

Our corpus study indicated that even helpful captions could be difficult to process and fully understand; moreover, once the caption was understood, we would still need to relate it to the information extracted from the graphic itself, which appears to be a difficult problem. Thus we decided to focus on communicative signals that could be extracted via shallow processing of captions. Our analysis yielded the following observations:

- Verbs in a caption often suggest the general category of message being conveyed by the graphic. An example from our corpus is "*American Express total billings still lag*"; the verb *lag* suggests that the graphic conveys that some entity (in this case *American Express*) falls behind some others.
- Adjectives in a caption can also suggest the general category of message being conveyed by the graphic. An example from our corpus is "*Soaring Demand for Servers*" which is the caption on a graphic that conveys the rapid increase in demand for servers. Here the adjective *soaring* is derived from the verb *soar*, and suggests that the graphic is conveying a strong increase.
- Words that usually appear as verbs, but are used in the caption as a noun, may function similarly to verbs. An example is "*Cable On The Rise*"; in this caption, *rise* is used as a noun, but suggests that the graphic is conveying an increase.
- Nouns in a caption that reference a label on the independent axis of the graphic serve to highlight the associated bar and make it salient. An example from our corpus is "*Germans miss their marks*" where the graphic displays a bar chart in which Germans correlates with a label in the graphic and the graphic is intended to convey that Germans are the least happy with the Euro.

Based on these observations, we compiled a set of *helpful* verbs and adjectives (identified through our corpus study, WordNet [68] and a thesaurus [50]) and manually divided them into similarity classes[8]; for example, the verbs *rise* and *soar* were placed in the same class, whereas the verbs *lag* and *trail* were placed in a different class. Adjectives derived from verbs, such as *soaring*, are treated as verbs. We then used a part-of-speech tagger and a stemmer to implement a type of shallow processing of captions to identify 1) the presence of one of our verb or adjective classes (adjectives and nouns derived from verbs, such as "soaring", are reduced to their root form and treated as verbs), and 2) nouns which match the label of a data element in the bar chart. The Caption Tagging Module (CTM) is responsible for entering this caption evidence into an augmented XML representation of a graphic.

6. Plan inference

Plan inference is the recognition of an agent's plan from his observed actions. Pollock [57] distinguishes among scenarios in which 1) the agent is actively uncooperative (trying to prevent his plan from being recognized, as in adversarial situations [27,26,60]), 2) passive (unconcerned about whether his plan is recognized), and 3) actively cooperative (attempting to facilitate the recognition of his plan). Cohen, Perrault, and Allen [17] refer to the latter two situations respectively as *keyhole*

[8] Although there are several resources that could be used for this task, we chose to use WordNet since it provides sets of synonyms (synsets) with relations between them; other resources provide more elaborate lexical structures. For example, FrameNet [62] provides frames which fully describe situations, objects, or events, and which have associated lexical units. Our verb classes do not necessarily correspond with FrameNet lexical classes. For example, the lexical units associated with the frame for Change_position_on_a_scale include *soar*, *decline*, and *fluctuate*; *soar* and *decline* are members of different verb classes in our system, and *fluctuate* is not a member of any of our verb classes since it does not suggest a kind of message that typically occurs in a simple bar chart.

plan recognition and *intended recognition*. Examples of keyhole plan recognition include identifying an agent's quest and next action in an adventure game [1], recognizing the state of individual team members via their routine communications to one another in order to monitor the team's progress on a task [42], and identifying the goals of an elderly or impaired individual in order to provide helpful advice [58]. Activity recognition is similar to keyhole plan recognition, in that the agent is unaware that a system is attempting to infer the activity in which he is engaged. However, as pointed out by Geib and Goldman [28], the objective of activity recognition is to identify a single activity as opposed to a plan consisting of multiple actions. An example of activity recognition is the use of information provided by multimodal sensors to identify the physical activity (such as walking) in which an individual is engaged [14] in order to provide advice and/or support for improving fitness.

Plan recognition for language understanding falls into the category of intended recognition, since the speaker (or writer) intends for the listener (or reader) to deduce the communicative goal of an utterance (or piece of text). Beginning with the seminal work of Allen [2] who developed a system for deducing the intended meaning of an indirect speech act, researchers have applied plan inference techniques to a variety of problems associated with understanding utterances, particularly utterances that are part of a dialogue. Examples of recent work in intended plan recognition include a dialogue system for obtaining weather information [49] and speech and typed command understanding in a role playing game [29,30]. In addition, research on assistive devices (such as wheelchairs) uses plan recognition to analyze the user's actions and identify his plan in order to make the device achieve the user's goals [38].

Our research takes the ideas and concepts developed for plan inference in understanding natural language and extends them to recognize communicative intention in the domain of information graphics. Treating information graphics as another form of language is appropriate when one considers 1) Clark's [15] view of language as any "signal" (or lack of signal when one is expected), where a signal is a deliberate action that is intended to convey a message, and 2) the observation that the overwhelming majority of the graphics that we have examined (taken from newspapers and magazines) appear to have some underlying communicative goal. But extending existing plan inference techniques to the recognition of intentions from bar charts is not a straightforward task and requires that a number of issues be addressed.

First, the communicative signals present in a bar chart are necessarily different from the communicative signals used to reason about the intention of an utterance in a dialogue. For example, Allen [3] utilized signals such as the mood of the utterance and expectations about likely goals of the speaker. Carberry [9] tracked and utilized the focus of attention in an ongoing dialogue. Lambert [46] used the surface form of the utterance, along with stereotypical beliefs, as communicative signals in recognizing complex discourse acts such as expressions of doubt. Gorniak and Roy [29,30] took into account the physical context in identifying the intended meaning of spoken language. If, as we contend, information graphics are an alternate form of language, it stands to reason that there will be an analogous (but different) set of communicative signals present in a bar chart that we can utilize in order to reason about the graphic designer's intended message. The previous section discussed the set of communicative signals that we have identified, including perceptual task effort, salience techniques, and nouns, verbs, and adjectives extracted from captions.

Other issues that must be addressed in extending plan inference to information graphics include modeling the strength of the evidence provided by the various communicative signals, and determining how to combine evidence and resolve conflicting evidence. The next section addresses these issues, and presents our probabilistic framework for inferring the communicative message of an information graphic. We discuss our use of a Bayesian belief network [55], the plan operators that we developed in order to capture knowledge about how the graphic designer's goal of conveying a message can be achieved via the viewer performing certain perceptual and cognitive tasks, and the way these plan operators are captured in our network structure.

7. A Bayesian network for identifying a bar chart's message

Plan inference involves searching through a space of possible plans to identify the one most likely to represent the agent's intentions. Charniak and Goldman [12] were the first to use a Bayesian belief network to address the uncertainty inherent in plan recognition. A Bayesian belief network is a probabilistic framework based on Bayes' rule, and it is also sometimes referred to as a belief network or a Bayesian network. In a Bayesian belief network, the nodes represent propositions and the arcs between the nodes represent causal dependencies (in contrast with generic probabilistic dependencies). These dependencies are captured using conditional probability distributions that represent the probability of a proposition given the various values of its parent node(s). Evidence propagation is used to compute the posterior probability of each proposition. In Bayesian networks for plan inference, the root nodes typically represent various high-level hypotheses about an agent's plan. In recent years, researchers have often used some form of Bayesian network in modeling plan recognition [1,59,60,38,28].

In all plan inference systems, there is some explicit plan structure which defines the relationships among goals, subgoals and primitive actions. In early plan inference systems, the plan structure (or domain knowledge) was represented in the form of operators that decomposed goals into a sequence of subgoals and eventually into primitive actions. The plan inference mechanism then utilized these plan operators, along with evidence in the form of an observed action (such as an utterance in Allen's seminal work [3]), to chain backwards on the plan operators to deduce one or more high-level goals that might have led the agent to perform the observed action as part of an overall plan for achieving his goal(s). In Bayesian networks, the plan structure is captured by the network itself, rather than by plan operators. Each goal, subgoal, and primitive action is represented as a piece of a network. If a goal can be decomposed into a particular set of subgoals or primitive actions, an arc from the goal to each subgoal (or primitive action) is used to represent this causal relationship. In this way, plan operators can be "mapped" to a Bayesian network structure — although, of course, the process can become more complicated than this since the plan structure might include explicit sequencing of actions, conditionalization, iteration and context. Huber et al. [36] provide an example of automatically mapping a general and potentially complex plan structure to a Bayesian network.

One of our claims is that information graphics in popular media are a form of language with a communicative goal and that plan recognition can be extended to identifying the intended message of an information graphic. Thus we chose to use plan operators to describe communicative goals and their decomposition into lower-level task goals. In our work, the plan that is developed by expanding goals into their constituent subgoals is the plan that the graph designer intends the graph viewer to execute in order to recognize the intended message of the information graphic. Thus we are getting at the intended message of the graphic by recognizing not the actions of the graph designer, but rather the actions that the graph designer intends for the viewer to perform (such as perceiving the rank of a bar in a bar chart) and which the graph designer has facilitated via the chosen design for the graphic. The observations are the features of the graphic that were included by the graph designer in order to facilitate the viewer's performing these actions and thereby result in the viewer's recognition of the graph designer's intended message. We chose to use a Bayesian network since it nicely facilitates the comparison of different hypotheses based on the evidence present in the graphic and provides a means of subsequently evaluating the impact of different evidence sources [10]. Thus we mapped our plan operators to a Bayesian network structure. Section 7.1 describes our plan operators and Section 7.2 describes the resulting network structure, including how our identified communicative signals are used as evidence in the Bayesian network. Section 7.3 then discusses the construction of the conditional probability tables. In our implementation, we utilize Netica [54]; this software supports the construction of Bayesian networks and, given the probabilities for the conditional probability tables of a network as well as any evidence or "findings", calculates the probabilities of each node in the network.

7.1. Plan operators

Our plan operators capture knowledge about how the graphic designer's goal of conveying a message can be achieved via the viewer performing certain perceptual and cognitive tasks, as well as knowledge about how perceptual and cognitive tasks decompose into sets of simpler tasks. Fully expanding the subgoals of a high-level operator would produce a complete plan for a viewer to achieve the goal of the high-level operator. Therefore, the operators capture what the graphic designer wants to ensure can be accomplished via a given information graphic. Each plan operator consists of the following components:

- **Goal:** the goal that the operator achieves.[9]
- **Data requirements:** requirements which the data must satisfy in order for the operator to be applicable in a graphic planning paradigm.
- **Display constraints:** features that constrain how the information graphic is eventually constructed if this operator is part of the final plan.
- **Body:** lower-level subgoals or primitive actions that must be accomplished in order to achieve the overall goal of the operator.

Consider the plan operator shown in Fig. 8 which represents the goal of having the viewer believe that the value corresponding to a particular element displayed in the graphic has a specific rank among the values of all elements in the graphic. This plan operator represents the high-level message that our system recognizes for the graphic shown in Fig. 9 — that is, that American Express ranks fourth among the credit card companies shown in the graphic in terms of number of U.S. credit cards that it has in circulation.

When the operators are used for plan inference, the display constraints eliminate operators from consideration — if the graphic does not satisfy the display constraints, then the operator could not have been part of a plan that produced the graphic. The data requirements in our operators are used to instantiate parameters in the operator. That is, the data must have had certain characteristics for the operator to have been included in the designer's plan, and these limit how the operator's arguments can be instantiated. The data requirements for the RankFromBar operator (Fig. 8) guide the instantiation of the primary key and dependent attribute of the graphic, specify that the values of the dependent attribute must have a natural ordering along a quantitative scale, and ensure that parameters representing the element (the particular bar), the rank, and the dataset are properly instantiated.

The body of our plan operators specifies a set of subgoals or primitive actions that can be carried out in order to accomplish the goal of the operator. For example, the goal of determining the rank of a particular bar in a bar chart (Fig. 8) can be accomplished by perceiving whether or not the bars are sorted according to height within the graphic, perceiving

[9] Note that we have shortened the names of our goals and subgoals, so that they will fit into our network diagrams; for example, the goal *RankFromBar* in Fig. 8 refers to the goal of *recognizing* the rank of a particular bar in a bar chart.

Goal: RankFromBar($\langle \mathit{viewer}\rangle$, $\langle \mathit{ds}\rangle$, $\langle g\rangle$, $\langle b_x\rangle$, $\langle v_x\rangle$, $\langle \mathit{rank}\rangle$)
Gloss: Viewer to believe from graphic $\langle g\rangle$ that the element of dataset $\langle \mathit{ds}\rangle$ depicted as bar $\langle b_x\rangle$ with the value $\langle v_x\rangle$ for primary key $\langle \mathit{att}_1\rangle$ has a value for $\langle \mathit{att}_2\rangle$ that is $\langle \mathit{rank}\rangle$ among the values of $\langle \mathit{att}_2\rangle$ for elements of dataset $\langle \mathit{ds}\rangle$
Data requirements:
1. $\langle \mathit{att}_1\rangle$ is the primary key attribute for dataset $\langle \mathit{ds}\rangle$
2. The values of $\langle \mathit{att}_2\rangle$ have a natural ordering along a quantitative scale
3. The value of $\langle \mathit{att}_2\rangle$ for $\langle b_x\rangle$ has rank $\langle \mathit{rank}\rangle$ among the elements of $\langle \mathit{ds}\rangle$

Display constraints:
1. Graph $\langle g\rangle$ is of type bar-chart
2. For each value of $\langle \mathit{att}_1\rangle$ in dataset $\langle \mathit{ds}\rangle$, the value of $\langle \mathit{att}_1\rangle$ and associated value of $\langle \mathit{att}_2\rangle$ are displayed via a bar on graph $\langle g\rangle$

Body:
1. PerceiveIfSorted: Viewer performs the perceptual task of determining whether the values of $\langle \mathit{att}_2\rangle$ for successive values of $\langle \mathit{att}_1\rangle$ occur along the independent axis in $\langle \mathit{sorted}\rangle$ (ascending or descending) order
2. PerceiveRank: Viewer performs the perceptual task of finding the $\langle \mathit{rank}\rangle$ relative to $\langle \mathit{att}_2\rangle$ of bar $\langle b_x\rangle$ in graph $\langle g\rangle$
3. GetLabel: Viewer performs the perceptual task of finding the value $\langle v_x\rangle$ for $\langle \mathit{att}_1\rangle$ where $\langle v_x\rangle$ corresponds with the bar $\langle b_x\rangle$ on graph $\langle g\rangle$

Fig. 8. Operator for finding the rank of an element.

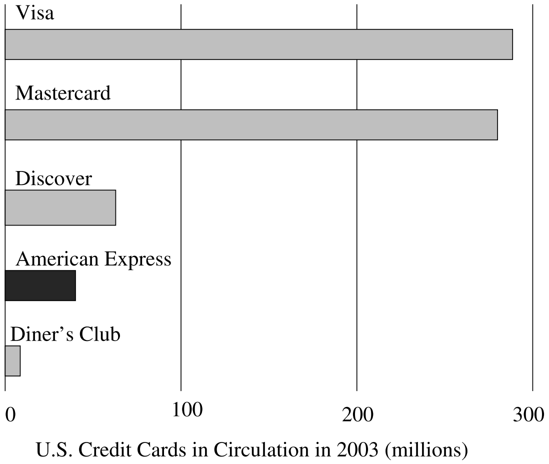


Fig. 9. Bar chart showing credit card circulation data.[10]

the rank of the particular bar within the graphic (the estimated effort for this task is affected by whether the bars were in sorted order), and determining the label of that bar. This would be the most natural set of tasks to carry out if the bar was salient, as is the case in Fig. 9. Alternatively, the plan operator in Fig. 10 shows a different task decomposition for the same goal — the differences between Figs. 8 and 10 can be seen in the bodies of the two operators. The body of the RankFromLabel operator (Fig. 10) states that the goal of determining the rank of a particular bar in a bar chart can be accomplished by finding the bar in the graphic that corresponds to a particular label, perceiving whether or not the bars are sorted according to height within the graphic, and determining the rank of the bar. This would be a more natural set of tasks to carry out if the label of the bar was mentioned in the caption. Notice that although the subtasks in the body of our operators are numbered, the ordering of the steps does not impact our plan inference process, and the steps do not always have to be performed in the order of their appearance.

Our library of plan operators includes *Data Presentation* operators, *Supporting Goal* operators, and *Perceptual* and *Cognitive Primitives*. *Data Presentation* operators describe methods for conveying via bar charts the kinds of messages described in Section 4; Figs. 8 and 10 are examples of *Data Presentation* operators. *Supporting Goal* operators decompose subgoals of the *Data Presentation* operators; for example, Figs. 11 and 12 show two alternative *Supporting Goal* operators that can be used to achieve the goal of the viewer finding the value of the dependent attribute represented by the top of a bar. The body of the *ExactValuePercep* operator in Fig. 11 consists of the primitive perceptual task *PerceiveValue*, in which the viewer simply perceives the value (the bar could be annotated with the exact value or the top of the bar could align with a labelled tick mark). On the other hand, the *ExactValueInterp* operator in Fig. 12 specifies how the same goal can be achieved, admittedly with more effort, using a combination of perceptual and cognitive tasks. The first subgoal, *PerceiveInfoToInterpolate*, is a

[10] This is based on a bar chart from an article in the September 13, 2004 issue of BusinessWeek.

Goal: RankFromLabel($\langle viewer\rangle$, $\langle ds\rangle$, $\langle g\rangle$, $\langle b_x\rangle$, $\langle v_x\rangle$, $\langle rank\rangle$)
Gloss: Viewer to believe from graphic $\langle g\rangle$ that the element of dataset $\langle ds\rangle$ depicted as bar $\langle b_x\rangle$ with the value $\langle v_x\rangle$ for primary key $\langle att_1\rangle$ has a value for $\langle att_2\rangle$ that is $\langle rank\rangle$ among the values of $\langle att_2\rangle$ for elements of dataset $\langle ds\rangle$
Data requirements:
1. $\langle att_1\rangle$ is the primary key attribute for dataset $\langle ds\rangle$
2. The values of $\langle att_2\rangle$ have a natural ordering along a quantitative scale
3. The value of $\langle att_2\rangle$ for $\langle b_x\rangle$ has rank $\langle rank\rangle$ among the elements of $\langle ds\rangle$

Display constraints:
1. Graph $\langle g\rangle$ is of type bar-chart
2. For each value of $\langle att_1\rangle$ in dataset $\langle ds\rangle$, the value of $\langle att_1\rangle$ and associated value of $\langle att_2\rangle$ is displayed via a bar on graph $\langle g\rangle$

Body:
1. PerceiveBar: Viewer performs the perceptual task of finding the bar $\langle b_x\rangle$ on graph $\langle g\rangle$ that corresponds to the element whose value for $\langle att_1\rangle$ is $\langle v_x\rangle$
2. PerceiveIfSorted: Viewer performs the perceptual task of determining whether the values of $\langle att_2\rangle$ for successive values of $\langle att_1\rangle$ occur along the independent axis in $\langle sorted\rangle$ (ascending or descending) order
3. PerceiveRank: Viewer performs the perceptual task of finding the $\langle rank\rangle$ relative to $\langle att_2\rangle$ of bar $\langle b_x\rangle$ in graph $\langle g\rangle$

Fig. 10. Alternative operator for finding the rank of an element.

Goal: ExactValuePercep($\langle viewer\rangle$, $\langle ds\rangle$, $\langle g\rangle$, $\langle b\rangle$, $\langle v\rangle$)
Gloss: Given a bar $\langle b\rangle$ depicted in graph $\langle g\rangle$, viewer performs the task of finding the exact value $\langle v\rangle$ of attribute $\langle att\rangle$ in dataset $\langle ds\rangle$ that corresponds to $\langle b\rangle$
Data requirements:
1. $\langle att\rangle$ is not the primary key attribute for dataset $\langle ds\rangle$
2. The values of $\langle att\rangle$ have a natural ordering along a quantitative scale

Body:
1. PerceiveValue: Viewer performs the perceptual task of finding the exact value $\langle v\rangle$ of attribute $\langle att\rangle$ that corresponds to the top of bar $\langle b\rangle$ depicted in graph $\langle g\rangle$

Fig. 11. Supporting goal operator for finding the exact value perceptually.

Goal: ExactValueInterp($\langle viewer\rangle$, $\langle ds\rangle$, $\langle g\rangle$, $\langle att\rangle$, $\langle b\rangle$, $\langle v\rangle$)
Gloss: Given a bar $\langle b\rangle$ depicted in graph $\langle g\rangle$, viewer performs the task of finding the exact value $\langle v\rangle$ of attribute $\langle att\rangle$ in dataset $\langle ds\rangle$ that corresponds to $\langle b\rangle$
Data requirements:
1. The values for $\langle att\rangle$ have a natural ordering along a quantitative or chronological scale
2. $\langle att\rangle$ is not the primary key attribute for dataset $\langle ds\rangle$

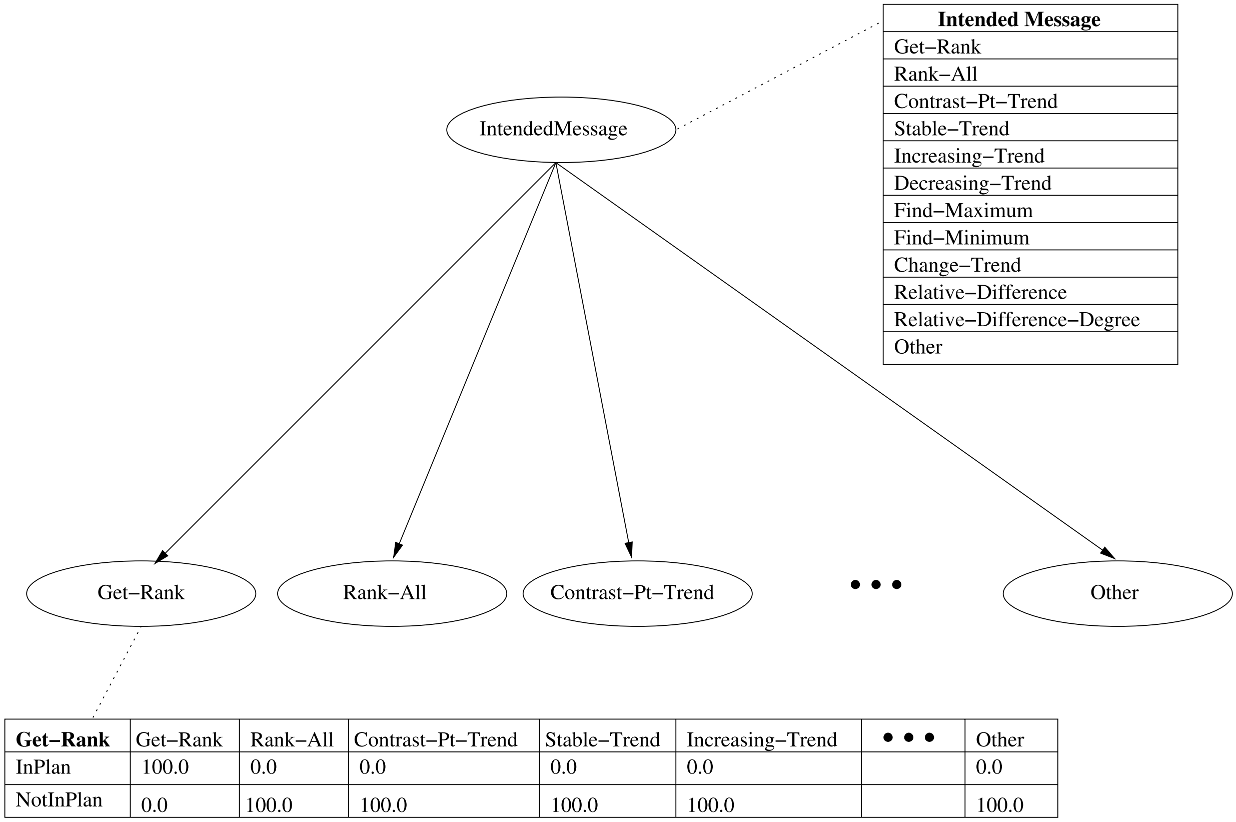
Body:
1. PerceiveInfoToInterpolate: Viewer performs the perceptual task of finding the location $\langle loc\rangle$, the values $\langle v_i\rangle$ and $\langle v_j\rangle$ on the axis $\langle a\rangle$ that displays $\langle att\rangle$, and the fraction $\langle f\rangle$ such that $\langle loc\rangle$ corresponds to the top of the bar $\langle b\rangle$ depicted in graph $\langle g\rangle$, the values $\langle v_i\rangle$ and $\langle v_j\rangle$ surround $\langle loc\rangle$ on axis $\langle a\rangle$, and $\langle f\rangle$ represents the distance between $\langle loc\rangle$ and $\langle v_i\rangle$ relative to the distance between $\langle v_i\rangle$ and $\langle v_j\rangle$
2. Interpolate: Viewer performs the cognitive task of interpolating between $\langle v_i\rangle$ and $\langle v_j\rangle$ using fraction $\langle f\rangle$ to find the value $\langle v\rangle$ of attribute $\langle att\rangle$ that corresponds to $\langle loc\rangle$ on the axis $\langle a\rangle$

Fig. 12. Supporting goal operator for finding the exact value cognitively.

primitive perceptual task in which the viewer perceives the values $\langle v_i\rangle$ and $\langle v_j\rangle$ immediately below and above the location on axis $\langle a\rangle$ corresponding to the top of the bar $\langle b\rangle$ of graph $\langle g\rangle$ and the fraction $\langle f\rangle$ of the distance that this location lies between $\langle v_i\rangle$ and $\langle v_j\rangle$. The second subgoal, *Interpolate*, is a primitive cognitive task in which the viewer computes (via interpolation) the value $\langle v\rangle$ of attribute $\langle att\rangle$ for bar $\langle b\rangle$ based on $\langle v_i\rangle$, $\langle v_j\rangle$ and $\langle f\rangle$. Each Perceptual Primitive is associated with a perceptual task rule in APTE (Section 5.1.2).

7.2. Network structure

In our domain, we utilize a Bayesian network which causally links the graphic designer's possible communicative intentions to perceptual and cognitive goals which the graphic designer wants the viewer to achieve, rather than reasoning



Intended Message
Get−Rank
Rank−All
Contrast−Pt−Trend
Stable−Trend
Increasing−Trend
Decreasing−Trend
Find−Maximum
Find−Minimum
Change−Trend
Relative−Difference
Relative−Difference−Degree
Other

Get−Rank	Get−Rank	Rank−All	Contrast−Pt−Trend	Stable−Trend	Increasing−Trend	• • •	Other
InPlan	100.0	0.0	0.0	0.0	0.0		0.0
NotInPlan	0.0	100.0	100.0	100.0	100.0		100.0

Fig. 13. Top levels of the Bayesian network.

about utterances or observed actions [12,36,1,35]. Our network structure captures the plan operators described in the previous section. The top-level node (or root node) in the network captures the probability of all of the possible categories of high-level messages underlying a graphic. The possible values for this node are the twelve categories of high-level messages detailed in Section 4. Each node in a Bayesian network has a conditional probability table. The probabilities in the root node's table (IntendedMessage in our network) represent the prior probabilities of each message category — the probability that a message category will occur without taking any additional evidence into account. The prior probabilities in the table of the IntendedMessage node are based on the distribution of the categories of high-level messages in our corpus. After completing the inference process, the entry with the highest probability in this node represents the category most likely to represent the graphic designer's primary intention or communicative goal for the graph. However, our Bayesian network must recognize not only the high-level category of a graphic's message but also the instantiation of the message's parameters.

Each individual category of high-level message is represented (again) as a child of the top-level node, as is shown in Fig. 13; Fig. 13 also shows the conditional probability table associated with the Get-Rank child node. This additional level does not affect the inference process or the computation of the probabilities in the network. Its presence merely simplifies the conditional probability tables in the network by limiting the size of the tables that would typically occur just below the top-level node.

7.2.1. Alternative instantiations

We refer to the process of replacing one or more of the parameters of a goal or perceptual task with specific elements or entities of the graphic as *instantiating* that goal or task. The network nodes in the top two levels are not instantiated — they represent the general categories of high-level messages in bar charts. Specific instantiations appear in the network as children of the nodes representing the high-level intentions. For example, the Get-Rank node is shown with several children in Fig. 14, illustrating several of its possible instantiations — namely, finding the rank of the first, second, or third bar in the graphic. (For readability, only the instantiation of the bar parameter $\langle b_x \rangle$ in each Get-Rank node is shown in Fig. 14.) The children of Get-Rank represent alternative instantiations. It is our contention that simple bar charts have a single primary message that they are intended to convey. For example, if two bars are highlighted in a bar chart where the bars are ordered by height, the primary message seems to be a comparison of the bars' values (either our Relative-Difference or our Relative-Difference-Degree category of message), not two Get-Rank messages. Thus inhibitory links [36] are used to capture

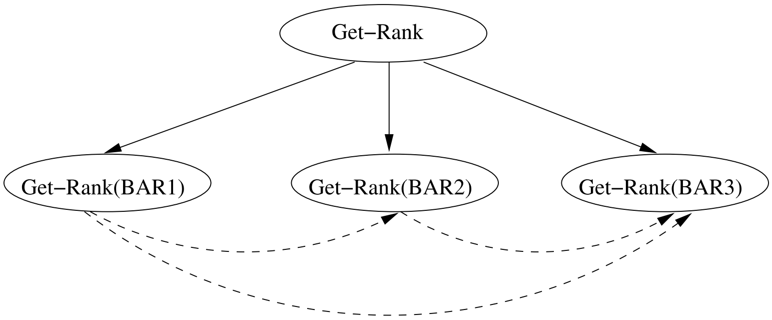


Fig. 14. Possible instantiations of Get-Rank.

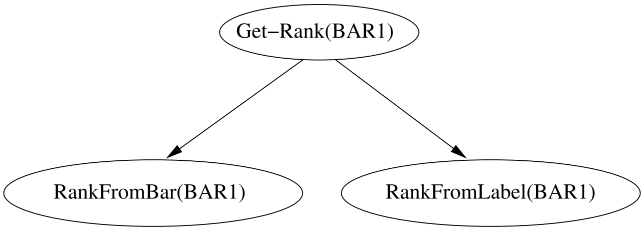


Fig. 15. Alternative ways to achieve the Get-Rank message.

the mutual exclusivity of the instantiated messages. The inhibitory links between the various instantiations of Get-Rank in Fig. 14 are shown as dashed lines. The effect of the inhibitory links is that the positive belief that one of the instantiations is part of the intended message (InPlan) will inhibit belief in the other possible instantiations being part of the intended message. Thus the child node of Get-Rank with the highest probability represents the instantiation of that goal most likely to be part of the graphic designer's intended message. Therefore, if after the inference process, the entry for Get-Rank has the highest probability in the IntendedMessage root node, our system selects the instantiated child node of Get-Rank with the highest probability, and produces the instantiated version of Get-Rank as its hypothesis about the intended message of the graphic.

7.2.2. Alternative subgoal decompositions

If there are multiple ways for a goal to be achieved, these are captured as children of the instantiated goal node. It is at this level that our plan operators are mapped into the structure of the Bayesian network. For example, the RankFromBar and RankFromLabel operators shown in Figs. 8 and 10 represent alternative ways of achieving the high-level Get-Rank message. These alternatives are captured as children of the instantiated Get-Rank node, as shown in Fig. 15.

Notice that there are no inhibitory links — no dashed lines — between the nodes representing the alternative ways of achieving the same goal. This is because the graphic designer may have enabled or facilitated multiple tasks that would allow the viewer to recognize the intended message, and the support for those individual tasks should combine to strengthen the belief in the high-level message rather than inhibiting each other. For example, the graph designer might have both colored the bar differently from other bars in the graphic and mentioned its label as part of the caption.

Each node representing an instantiated Data Presentation operator will have child nodes that represent their instantiated subgoals or perceptual or cognitive primitives. Any subgoal which corresponds to a non-primitive operator will in turn have children representing its instantiated subgoals. Continuing this expansion until the lowest node in each branch of the network corresponds to a perceptual or cognitive primitive produces a full Bayesian network with nodes representing all possible plans for achieving the intended message of the graphic. The subnetwork for Get-Rank(BAR1), including the alternative decompositions of RankFromBar(BAR1) and RankFromLabel(BAR1), is shown in Fig. 16. Notice that a node corresponding to a specific instantiation of a plan operator or primitive may be part of the subgoal decomposition of several plan operators, and so nodes may have multiple parents. This is the case for PerceiveIfSorted and PerceiveRank in Fig. 16 since they are part of the instantiated body of both RankFromBar(BAR1) and RankFromLabel(BAR1). Each subgoal or primitive task has a conditional probability table that captures the probability that the subgoal or primitive task is part of the plan for recognizing the graphic's message given that any one (or more) of its parent nodes is believed to be part of that plan.

7.2.3. Evidence nodes

In addition to encoding the plan structure, our Bayesian network also needs to explicitly represent evidence that should influence the credibility of different hypotheses. The evidence represented in our network corresponds to the communicative

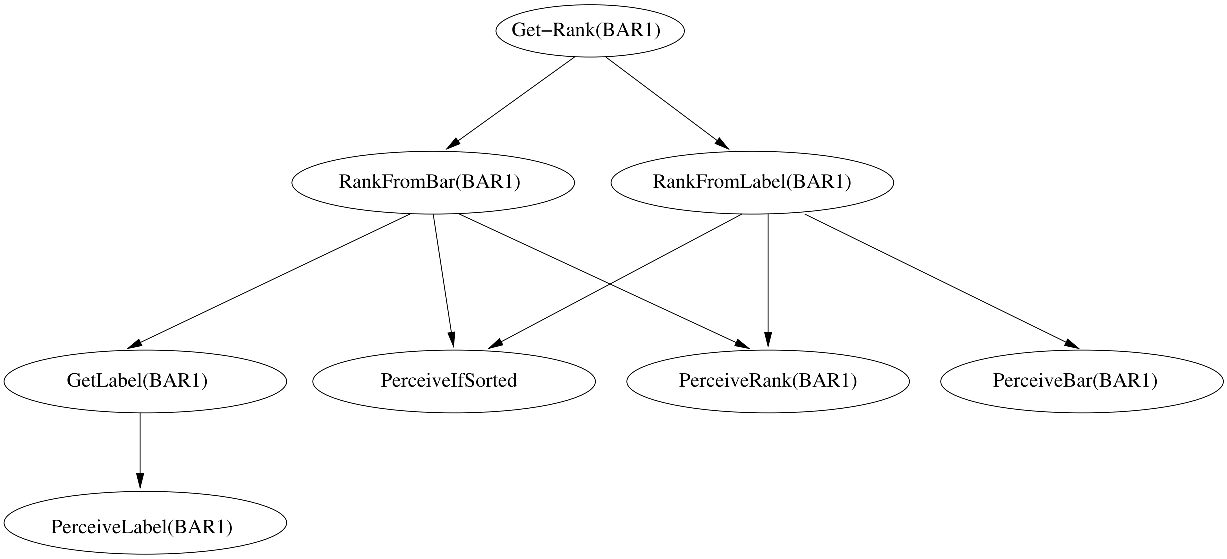


Fig. 16. Get-Rank(BAR1) subnetwork.[11]

signals that we have identified for information graphics, namely perceptual task effort, salience evidence, and evidence extracted from the caption.

Most of the signals, such as perceptual task effort or whether a particular bar in a bar chart is highlighted, provide evidence regarding the probability that a particular perceptual task (or primitive) is part of the plan for achieving the overall intended message of a graphic. Therefore, evidence nodes representing these communicative signals are added to the network as children of the nodes representing perceptual tasks. There are six different types of evidence nodes that are attached to perceptual tasks:

- *Effort*: captures the relative effort required for the perceptual task that is its parent.
- *Highlighting*: captures whether a parameter in the perceptual task is instantiated with a highlighted element in the graphic.
- *Annotation*: captures whether a parameter in the perceptual task is instantiated with an element that has a special annotation in the graphic (a special annotation occurs when only a subset of the bars in a bar chart is annotated).
- *MostRecentDate*: captures whether a parameter in the perceptual task is instantiated with an element that is associated with the most recent date in the graphic.
- *SalientHeight*: captures whether a parameter in the perceptual task is instantiated with an element whose height makes it salient in the graphic.[12]
- *NounInCaption*: captures whether a parameter in the perceptual task is instantiated with an element whose label matches a noun in the caption.

Each instantiated perceptual task in the network will have a unique set of evidence nodes as its children. An example of a perceptual task, PerceiveRank(BAR1), with its associated evidence nodes is shown in Fig. 17. In our diagrams, nodes shown with dashed outlines represent evidence nodes.

With the exception of the Effort evidence node where the possible values are *easy*, *medium*, *hard*, or *impossible*, the possible values of the evidence nodes vary depending on how many bars are involved in the particular task. For example, PerceiveRank involves finding the rank of just one bar. The possible values for the Annotation node attached to Perceive-Rank(BAR1) capture the following conditions: 1) only the first bar is annotated, 2) the first bar and some other bars are annotated, 3) only other bars are annotated, or 4) no bars are annotated. Note that the possible values for the evidence nodes capture not just positive evidence (such as the bar involved in the task being annotated), but also negative evidence (such as other bars being annotated or no bars being annotated). In this way, the *lack* of an explicit signal can also convey information about the likelihood of a task being part of the intended plan. Table 3 displays the conditional probability table for the Annotation evidence node attached to the PerceiveRank(BAR1) task node. It shows that if PerceiveRank(BAR1) is part of the plan for the viewer to recognize the graphic's message, then the probability that just BAR1 is annotated

[11] The graph is not intended to imply an order to the subtasks.

[12] Currently, only bars which are significantly taller than the other bars are considered to be visually salient. Additional testing would need to be conducted to determine if significantly shorter bars are also visually salient.

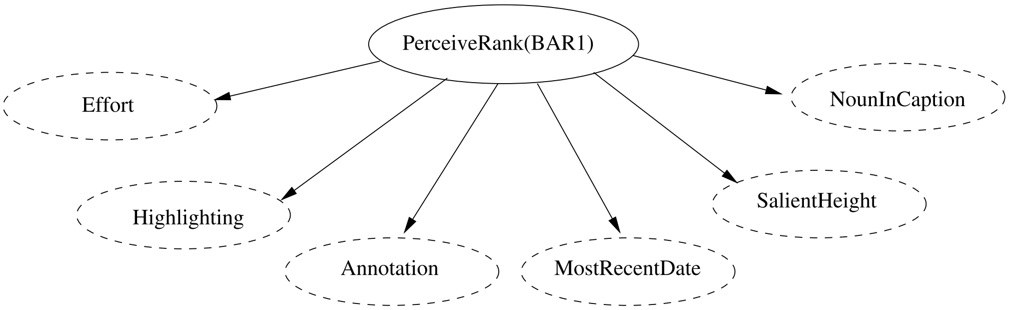


Fig. 17. Perceptual task node with evidence nodes.

Table 3
A sample conditional probability table for an evidence node. *InPlan* and *NotInPlan* represent the possible values of the parent node, while the rows represent the combinations regarding the presence or absence of annotations involving a bar instantiated in the parent node.

Annotation	InPlan	NotInPlan
Only ⟨*bar*⟩ annotated	24.99	2.3
⟨*bar*⟩ and others annotated	0.01	0.9
Only others annotated	0.01	19.5
No bars annotated	74.99	77.3

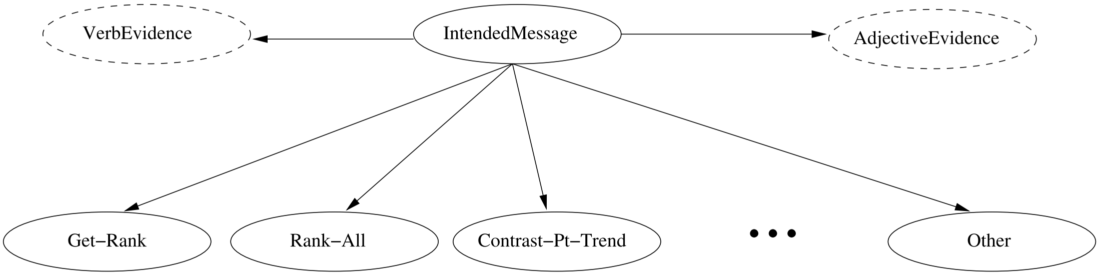


Fig. 18. Verb and adjective evidence nodes.

is 24.99%, whereas the probability that only other bars are annotated is .01%. On the other hand, if PerceiveRank(BAR1) is not part of the plan, then the probability that just BAR1 is annotated is only 2.3% whereas the probability that only other bars are annotated is 19.5%. The acquisition of the probabilities for the conditional probability tables is discussed in Section 7.3.

As discussed in Section 5.2, certain verbs and adjectives in the caption are communicative signals that provide evidence about the general category of message conveyed by the graphic.[13] For example, in the caption "More Communication Towers", the adjective "more" suggests that the viewer is intended to recognize an increasing trend in the number of communication towers, and in the caption "South Africa Tops in Gold Production", "tops" suggests that the viewer is intended to recognize the element with the maximum gold production in the graphic. In order to reflect the fact that verbs and adjectives provide evidence regarding the high-level message category, the evidence nodes representing verb and adjective evidence are children of the top-level node as shown in Fig. 18.

7.3. *Constructing the conditional probability tables*

A critical component in building a Bayesian network is obtaining the probabilities that are required for the conditional probability tables in the network. As Druzdzel and van der Gaag [19] note, "Where do the numbers come from?" is a commonly asked question. And, unfortunately, it is sometimes unclear that the necessary numbers can be reasonably or accurately obtained. In contrast to many other probabilistic plan recognition models where it is difficult to empirically determine the probabilities, the probabilities used in our belief network have been obtained through a corpus analysis.

[13] Note that nouns and adjectives derived from verbs are treated as the root form of the verb.

In order to calculate the necessary probabilities, each graphic in the corpus was analyzed to determine for each possible instantiation of each perceptual task:

- the relative effort (categorized as easy, medium, hard, or impossible) as estimated by our effort estimation rules implemented in the APTE component.
- which parameters referred to elements that were salient in the graphic and the kind of salience (highlighted bars, annotated bars, bars with labels that appear as nouns in the caption, etc.).
- whether elements in the graphic other than those referred to by parameters in the task were salient and the kind of salience (highlighted bars, annotated bars, bars with labels that appear as nouns in the caption, etc.).

In addition, we noted the occurrence in the caption of one of our identified verb or adjective classes.

As described in Section 4, two coders identified the intention of each graphic in the corpus. We applied our plan operators to manually construct a plan (constrained by what appeared in the graphic) for achieving the posited intention of each graphic. The fully expanded plans, including the goal–subgoal relationships and the low-level perceptual and cognitive tasks, were recorded. We then had all of the information necessary to calculate the probability of a particular Data Presentation operator being part of the plan given the high-level intention of the graphic. For example, given that the high-level intention of the graphic was Get-Rank, we could now calculate how probable it was that RankFromBar or RankFromLabel was part of the plan.

This data also provided the information necessary to calculate the conditional probability tables for the evidence nodes. This was fairly straightforward for the verb and adjective evidence nodes. We simply needed to calculate the likelihood of the presence in the caption of a verb or adjective from each of the various classes given the high-level intention of the graphic. For the evidence nodes that appear as children of the perceptual tasks (such as highlighting, annotations, and perceptual task effort), the calculations were a bit more complex. However, formulas were constructed to compute all of the required conditional probability tables. Examples of the needed conditional probability tables include: 1) the conditional probability of a particular perceptual task being easy, medium, hard or impossible given that the perceptual task is (or is not) part of the plan, 2) the conditional probability of a bar's height being salient given that recognizing the intended message entails (or does not entail) performing a particular perceptual task involving that bar, and 3) the conditional probability of a bar being annotated given that recognizing the intended message entails (or does not entail) performing a particular perceptual task involving that bar. Table 3 shows an example of this latter conditional probability table for the perceptual task PerceiveRank. As an example of calculating the probabilities required for this table, the probability of a particular bar (and only that bar) being annotated given that the intended message for the graph *does* entail perceiving the rank of that bar is

$$\frac{\mathrm{Count}(\mathrm{InPlan}(\mathrm{PerceiveRank}(\langle \mathit{bar}\rangle)) \wedge \text{only } \langle \mathit{bar}\rangle \text{ annotated})}{\mathrm{Count}(\mathrm{InPlan}(\mathrm{PerceiveRank}))}$$

Since the structure of the conditional probability tables includes values for the likelihood of tasks being both *in plan* and *not in plan*, data from each graph in the corpus is utilized in each conditional probability table. Consider, for example, a graphic in our corpus whose message has been coded as Get-Rank(BAR4) where BAR4 is the only bar in the graphic annotated with its value. The plan for achieving the Get-Rank message will include the perceptual task PerceiveRank(BAR4). This graphic will contribute to the calculation of the probabilities shown in Table 3. On the other hand, a graphic whose coded message category is Rising-Trend and in which no bars are annotated also contributes to the calculation of the probabilities shown in Table 3 even though PerceiveRank is not part of the plan for achieving the message of this particular graphic.

It is time-consuming to collect a corpus of bar charts, manually construct a plan for each bar chart that achieves its high-level goal and is consistent with the features of the graphic, and then extract the needed probabilities from the generated plans and the evidence in the graphics. Blaylock and Allen investigated the automatic generation of plan corpora [6]. They modified an AI planner so that it stochastically generates plans in the domain, with every plan (not just optimal ones) being a possible result. They note that the planner would need to be given the a priori probability of each top-level goal schema, along with the a priori probabilities of each parameter value. In addition, it is assumed that disjunctive subgoals are equally probable, although a weighted distribution could be used if the probability of each were known.[14] Recall that in our work, the plan that is recognized (and thus the plan that must be generated) is the plan that the graph designer intends that the viewer will execute to recognize the graphic's intended message. To automate the generation of such plans, one might apply Blaylock's methodology and modify a planner so that given a graphic and its intended message, the planner would produce a plan for the viewer that achieves recognizing the message and is consistent with the graphic; we could then extract our probabilities from the set of such automatically generated plans. While this is a promising avenue for future research, there are a number of issues that would need to be addressed, such as ensuring that the resultant plans took advantage of all the cues in the graphic (since this is what the graph designer intends for the viewer to do in recognizing the graphic's message).

[14] Personal correspondence.

8. The inference process

8.1. A starting point for plan inference

One of the characteristics that make intention recognition in information graphics unique is that the communication between the graphic designer and the viewer of the graphic is not incremental. In dialogue, written communication, and in domains where the user is completing a task through interaction with the system, there is an incremental nature to the interaction between the user and the system. For example, as each utterance is made in a dialogue, the plan inference system can assume that it is relevant and attempt to fit the utterance into the overall plan. However, when recognizing intentions in information graphics, the system is presented with a complete information graphic. Upon first consideration, this complete knowledge about the information graphic might seem to simplify the inference process for information graphics. Given the graphic and the network structure, the system could potentially build a network containing every possible instantiation of every possible task and record the evidence for all of the candidate tasks. However, automatically adding to the network all of the possible instantiations of all possible perceptual tasks rapidly becomes infeasible due to the overwhelming size of the resultant network and practical constraints on memory. There are two basic factors that impact the size of the network, and the subsequent memory burden: 1) the number of nodes in the network, and 2) the size of the conditional probability tables. In practice, it is the size of the conditional probability tables that results in the primary drain on memory resources. This is not surprising when you consider the size of some of the conditional probability tables that would result from a network containing every possible instantiation of every possible task. If, for example, a bar chart contains 10 bars, there will be 45 different instantiations of the Relative-Difference-Degree goal (45 different combinations of bars that could be compared). Because of the inhibitory links that connect the different instantiations, the largest conditional probability table for one of the instantiated Relative-Difference-Degree nodes would need to contain entries for all of the possible combinations of values for the other 44 nodes as well as the possible values of its uninstantiated high-level parent node. Since each node has two possible values (InPlan or NotInPlan), this results in 2^{45} (or 35,184,372,088,832) rows in this single conditional probability table.

In order to address this problem, when the plan inference process is begun, we limit the size of the network by only adding nodes representing tasks that the communicative signals in the graphic suggest might be part of the inferred plan. These communicative signals include salience techniques, naturally salient items, nouns in the caption which correspond to labels of bars, and perceptual task effort. The explosive expansion of the number of nodes in the network is a problem that was also encountered by Charniak and Goldman [12], who used marker passing to limit the size of the network under consideration. The next sections discuss the process of identifying tasks for inclusion in the network by reasoning about the communicative signals in the graphic.

8.1.1. Perceptual task effort

As discussed in Section 5.1.2, the design choices made by the graphic designer facilitate some perceptual tasks more than others. Thus the relative effort required for a perceptual task serves as a communicative signal about what tasks the designer expects the viewer to perform. Following this reasoning, we use a set of ten low effort perceptual tasks that can be performed on a given bar chart as a starting point for the plan inference process and add nodes representing these tasks to the Bayesian network.

To identify the set of low effort perceptual tasks, we begin by estimating the perceptual effort required to perform each instantiated perceptual task on the given graphic, using our APTE rules (Section 5.1.2) to generate the estimates. In order to ensure that the set of ten tasks represents a fairly broad range of the possible perceptual tasks that can be performed on a graphic, we limit the set to containing a single instantiation of each particular task. We would not, for example, want the set to contain ten different instantiations of the PerceiveLabel task. If several instantiations of the same task require the same estimated perceptual effort, we choose just the first (or leftmost) instantiation — for example, if all of the instantiations of PerceiveLabel have an estimated cost of 680, PerceiveLabel(BAR1) will be included in the set. It is important to note that, as described in Section 8.1.2, if there is any evidence that another bar is salient (and therefore likely to be important to the intended message), nodes representing the instantiations of perceptual tasks involving this bar will be added to the network in a separate step.

Consider the graphic in Fig. 19 which shows the number of communication towers from 1970 to 2002. The set of ten low effort perceptual tasks for which nodes will be added to the Bayesian network, along with the estimated effort for each task, is shown in Fig. 20. Note that the effort estimates were generated for the original version of the information graphic as it appears in our corpus, and so some of the estimates will not be accurate for the graphic as it appears here because it has been enlarged.

8.1.2. Salience

The communicative signals that increase the salience of a particular element in the graphic include the highlighting of a bar, the special annotation of a bar, the salient height of a bar, the salient recency of a bar, and the use of the bar's label as a noun in the caption. Our contention is that if the graphic designer goes to the effort of employing attention-getting devices to make certain elements of the graphic particularly salient, or if the graphic designer believes that certain elements will naturally be salient to the viewer, then the salient elements serve as communicative signals — i.e., the designer probably

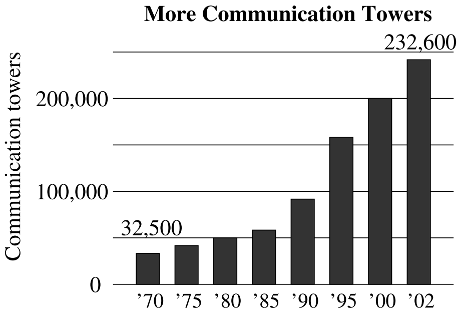


Fig. 19. Bar chart showing the number of communication towers.[15]

PerceiveIfSorted()	12
PerceiveMaximum(BAR8)	162
PerceiveIsMaximum(BAR8)	242
PerceiveIsMinimum(BAR1)	242
PerceiveRank(BAR1)	381
PerceiveMinimum(BAR1)	392
PerceiveValue(BAR1)	450
PerceiveRelativeDifference(BAR1, BAR2)	472
PerceiveLabel(BAR1)	680
PerceiveBar(BAR1)	681

Fig. 20. Set of ten lowest effort perceptual tasks from the bar chart in Fig. 19.

intends for them to be part of the intended message of the graphic. We do not, however, know anything about the particular task that the designer intends the viewer to perform with the salient element. Therefore, for each salient element, we add nodes to the Bayesian network that represent the possible perceptual tasks in which it could play a role.

Consider, again, the graphic in Fig. 19 which shows the number of communication towers from 1970 to 2002. The first and last bars are specially annotated — that is, those bars have annotations that are not a general design feature of the graphic. It is also the case that the last bar in the graphic (corresponding to the year 2002) is the most recent data in the graphic. Therefore, the first and last bars (the first and eighth bars of the graphic) are considered likely to be salient to the message of the graphic and nodes representing instantiations of perceptual tasks involving these bars will be added to the Bayesian network.

The list of perceptual tasks for which nodes will be added to the network due to the salience of the first and eighth bars is shown in Fig. 21. For most of the perceptual tasks, the process of instantiating the tasks with salient bars is very straightforward. For example, PerceiveLabel is instantiated once with each salient bar and nodes representing these instantiations are added to the network. For tasks involving two bars, such as PerceiveRelativeDifference, if two or more bars are salient, any possible instantiations involving two of the salient bars are added — hence, PerceiveRelativeDifference(BAR1, BAR8) is added to the network. The salient bars are then considered individually. For PerceiveRelativeDifference, we initially considered adding each possible instantiation involving a salient bar. For example, since the first bar is salient in Fig. 19, we would add a node for PerceiveRelativeDifference(BAR1, BAR2), PerceiveRelativeDifference(BAR1, BAR3), and so forth. However, we were again limited by the practical memory constraints of the system. We hypothesize that it is extremely unlikely (without additional signals) that the intended message of a bar chart containing a large number of bars will be Relative-Difference or Relative-Difference-Degree because there are many possible combinations of two bars which could be compared, and with so many choices, it is unlikely that the intention is for the viewer to pick out one of them.[16] Therefore, if a bar chart has four or fewer bars, all of the possible instantiations of PerceiveRelativeDifference involving a salient bar are added as nodes in the network. If a bar chart contains more than four bars, only the lowest effort instantiation is added. This accounts for the inclusion of PerceiveRelativeDifference(BAR1, BAR2) and PerceiveRelativeDifference(BAR7, BAR8) in the list in Fig. 21.

The other tasks in Fig. 21 that warrant some special attention are the three instantiated PerceiveIncreasingTrend tasks. The following section discusses the detection of possible trends in our system.

8.1.3. Detecting trends

In order to detect and reason about the trends that may be present in a bar chart, we have devised an algorithm for identifying the various ways that a bar chart might be divided into segments or "chunks" that convey trends. We define

[15] This is based on a bar chart from the newspaper USA Today.

[16] This observation is similar to the forking heuristic used by Allen [2] where candidate plans had their ratings downgraded if they were produced by mutually exclusive inferences.

PerceiveLabel(BAR1)
PerceiveLabel(BAR8)
PerceiveInfoToInterpolate(BAR1)
PerceiveInfoToInterpolate(BAR8)
PerceiveRelativeDifference(BAR1, BAR2)
PerceiveRelativeDifference(BAR7, BAR8)
PerceiveRelativeDifference(BAR1, BAR8)
PerceiveIncreasingTrend(BAR1, BAR8)
PerceiveIncreasingTrend(BAR1, BAR4)
PerceiveIncreasingTrend(BAR4, BAR8)
PerceiveMaximum(BAR8)
PerceiveBar(BAR1)
PerceiveBar(BAR8)
PerceiveRank(BAR1)
PerceiveRank(BAR8)
PerceiveIsMaximum(BAR8)
PerceiveIsMinimum(BAR1)
PerceiveValue(BAR1)
PerceiveValue(BAR8)
PerceiveMinimum(BAR1)

Fig. 21. Perceptual tasks involving salient bars from the bar chart in Fig. 19.

a *division* as consisting of one or more *segments* of bars such that these segments encompass the entire bar chart, from the first bar to the last bar. For example, a division of the bar chart in Fig. 19 might consist of two segments — one encompassing the first bar to the fourth bar and the second encompassing the fourth bar to the eighth bar. The algorithm utilizes the calculated slope of different possible segments, the weighted error of the bars in terms of variance from the slope, and a preference for the fewest possible number of segments. The plausible candidate divisions are identified before we construct the Bayesian network, and the XML representation of the graphic taken as input to our plan inference process is augmented to include the candidate divisions.[17]

Candidate divisions are only calculated for bar charts in which a trend is possible — that is, the primary key attribute must be either a time-line or an ordinal attribute. The basic algorithm for identifying the divisions to be included in the augmented XML is the following:

- Calculate the maximum number of allowed segments (or chunks) in the divisions as the truncation (or floor) of the number of bars in the bar chart divided by two, with an upper limit of four segments per division.[18]
- For $n = 1$ to the maximum number of segments, consider all possible ways of dividing the graphic into n segments:
  - For each segment of each possible division, calculate the weighted error per bar as

$$\sum_{k=i}^{j} \frac{\text{distance}(b_k, \text{line}(b_i, b_j))}{j - i + 1} \tag{1}$$

    where b_i and b_j are the first and last bars in the segment, line(b_i, b_j) is an imaginary line connecting the tops of the bars b_i and b_j, distance(b_k, line(b_i, b_j)) is the distance from the top of bar b_k to the imaginary line in centimeters, and $j - i + 1$ is the number of bars in the segment.
  - If the weighted error per bar of every segment in the division is <0.07, then the division is deemed *acceptable* and is included in the augmented XML, otherwise it is excluded. The threshold of 0.07 was chosen based on an examination of a sampling of graphics in our corpus.

After all of the candidate divisions have been identified, we attempt to choose what would be the most visually apparent division if a viewer was asked to identify a trend (or trends) in the graphic. Our identification of this "best" division is important because this is the only division that we consider when generating the set of the ten lowest effort perceptual tasks for addition to the network. Our guiding principle in identifying the best division is loosely based on Ockham's Razor; we want to identify the simplest division that is consistent with the data. Therefore, we prefer divisions with fewer segments. The process for identifying the best candidate division is the following:

- Sort the acceptable line divisions by number of segments (primary sort) and weighted error per bar (secondary sort).
- If the single segment division of the graphic is acceptable, this will be selected as the best division, *unless* the weighted error per bar of the two-segment division which includes a two-bar segment at the end of the division is lower than the weighted error per bar of the single segment division. This allows for the natural tendency of a viewer to detect changes in the most recent data compared to a previous trend.
- If the single segment division of the graphic was initially deemed unacceptable, but the slope difference between the segments of the best two-segment division is <.25, we select the single segment division as the best division because we believe that the small change in slope would not be visually apparent.
- If the previous steps do not result in a selected division, consider the multiple segment divisions. For $n = 2$ to the maximum number of segments *or* until a division is selected:

[17] The implementation of this process was done by Seniz Demir.

[18] For example, a bar chart with five bars will be allowed to have a maximum of two segments per division. We impose this restriction on the upper limit on the number of segments because we are only attempting to identify visually apparent trends, and also to ensure the feasibility of calculating all possible divisions.

- If there is an acceptable division with *n* segments which does not have a two-bar segment, this division is selected as the best division. Preference is given to those divisions which do not include two-bar segments, since segments consisting of only two data points do not actually represent a trend.
- If there is more than one such division, the one with the lowest weighted error per bar will be selected.
- Otherwise, any acceptable divisions containing a two-bar segment will be considered for selection (again, preference goes to the divisions with the lowest weighted error per bar).

The selected "best" division will be recorded as the first division in the augmented XML representation of the bar chart. Next in the XML representation will be any salient divisions — i.e., divisions where an endpoint of a segment or segments in the division is salient in the graphic. The motivation for this is that one of the roles that the salient elements might play in the intended message is as endpoints of a trend.

This algorithm captures rough heuristics for identifying visually apparent trends in a bar chart. As discussed in Section 12, we are currently working on using machine learning techniques to produce a learned model for dividing a line graph into visually distinguishable trends [69].

Our algorithm for detecting trends results in the inclusion of the three PerceiveIncreasingTrend instantiations in the list of salient tasks in Fig. 21. The best, or most preferred, division for the graphic in Fig. 19 is one with two segments. The first segment, from the first bar to the fourth bar, has a slope of .1 and a weighted error per bar of .00273. The second segment, from the fourth bar to the eighth bar, has a slope of 0.475 and a weighted error per bar of .00675. This two segment division causes the addition of two perceptual task nodes into the network, one for each segment: PerceiveIncreasingTrend(BAR1, BAR4) and PerceiveIncreasingTrend(BAR4, BAR8). In addition, because we have two salient bars, the single segment division in which the two endpoints are annotated (and therefore salient) causes a node representing the perceptual task PerceiveIncreasingTrend(BAR1, BAR8) to be added to the network.

8.2. Building the network

The nodes in the top two levels of the network, representing the high-level intentions (Fig. 13), exist in every network and are added at the beginning of the network construction process. As discussed in Section 8.1, we use perceptual task effort and salience to identify the low-level perceptual tasks that are initially inserted into the network. Once these perceptual tasks are added to the network, we perform upward chaining via the plan operators to add higher level goal nodes until a link is established to one of the top-level goal nodes. Whenever a node representing a goal is added to the network, links are added from that node to nodes representing all of its subgoals — if a node does not already exist to represent the subgoal, the node is added to the network. Thus the resulting network contains all nodes that are in some way related to the perceptual task nodes that were initially added to the network.

For example, when we add the node PerceiveIncreasingTrend(BAR1, BAR8) to the network, the higher-level goal node IncreasingTrend(BAR1, BAR8) will also be added because PerceiveIncreasingTrend(BAR1, BAR8) fulfills a subgoal of IncreasingTrend. IncreasingTrend(BAR1, BAR8) also has the subgoals GetLabel(BAR1) and GetLabel(BAR8), and so nodes representing these tasks are added to the network. GetLabel has PerceiveLabel as a subgoal and so two instantiations of PerceiveLabel, (PerceiveLabel(BAR1) and PerceiveLabel(BAR8)), are also added to the network. IncreasingTrend(BAR1, BAR8) will be linked to the high-level IncreasingTrend node since it is a specific instantiation of this high-level goal. An instantiation of IncreasingTrend may also fulfill a subgoal of ChangeTrend, and so this possibility is investigated. However, there are no possible instantiations of ChangeTrend with IncreasingTrend(BAR1, BAR8) as a subgoal (which makes sense, since IncreasingTrend(BAR1, BAR8) encompasses the entire graph), and so no instantiation of ChangeTrend is added to the network at this point. The entire subnetwork resulting from the addition of a node representing PerceiveIncreasingTrend(BAR1, BAR8) can be seen in Fig. 22.

8.3. Recording the evidence

After all of the task nodes have been added to the network, the final phase in the plan inference process is the recording of the evidence. First, the evidence nodes are added to the network. The adjective and verb evidence nodes are linked to the top-level IntendedMessage node, and the specific low-level salience and perceptual effort nodes are added for each perceptual task, as discussed in Section 7. The appropriate findings are then recorded for each evidence node.

For example, consider the graph in Fig. 19 and the addition of evidence nodes to its network of goal nodes (which is partially shown in Fig. 22). The value of the *annotation* evidence node that will be attached to PerceiveIncreasingTrend(BAR1, BAR8) will reflect the fact that both of the bars that are parameters of this task, and no other bars in the graphic, have salient annotations. In contrast, the value of the *annotation* evidence node that will be attached to PerceiveIncreasingTrend(BAR1, BAR4)[19] will reflect the fact that one of the bars in the task as well as other bars in the graphic have salient annotations. The adjective evidence node will record the presence of the adjective class containing "more", since the caption of the graphic is "More Communication Towers". Fig. 23 shows the subnetwork from Fig. 22 along with the evidence nodes that will be attached to the nodes in the subnetwork.

[19] For readability, this node is not shown in Fig. 22, but it is included in the complete network.

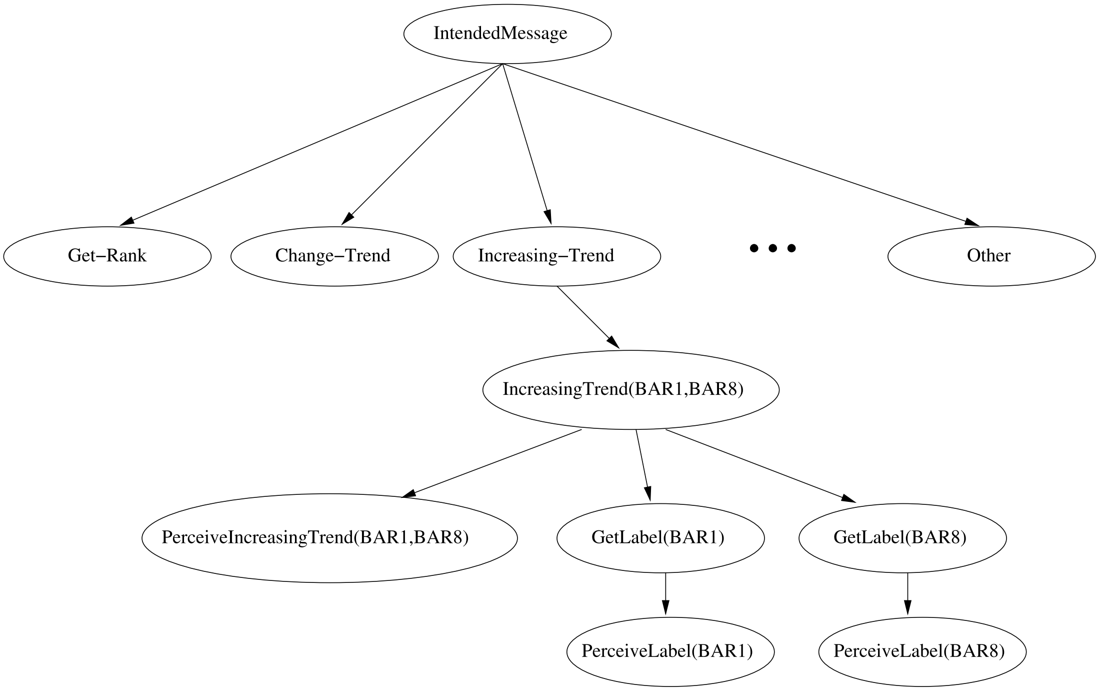


Fig. 22. Result of upward chaining from PerceiveIncreasingTrend(BAR1, BAR8) to IncreasingTrend.

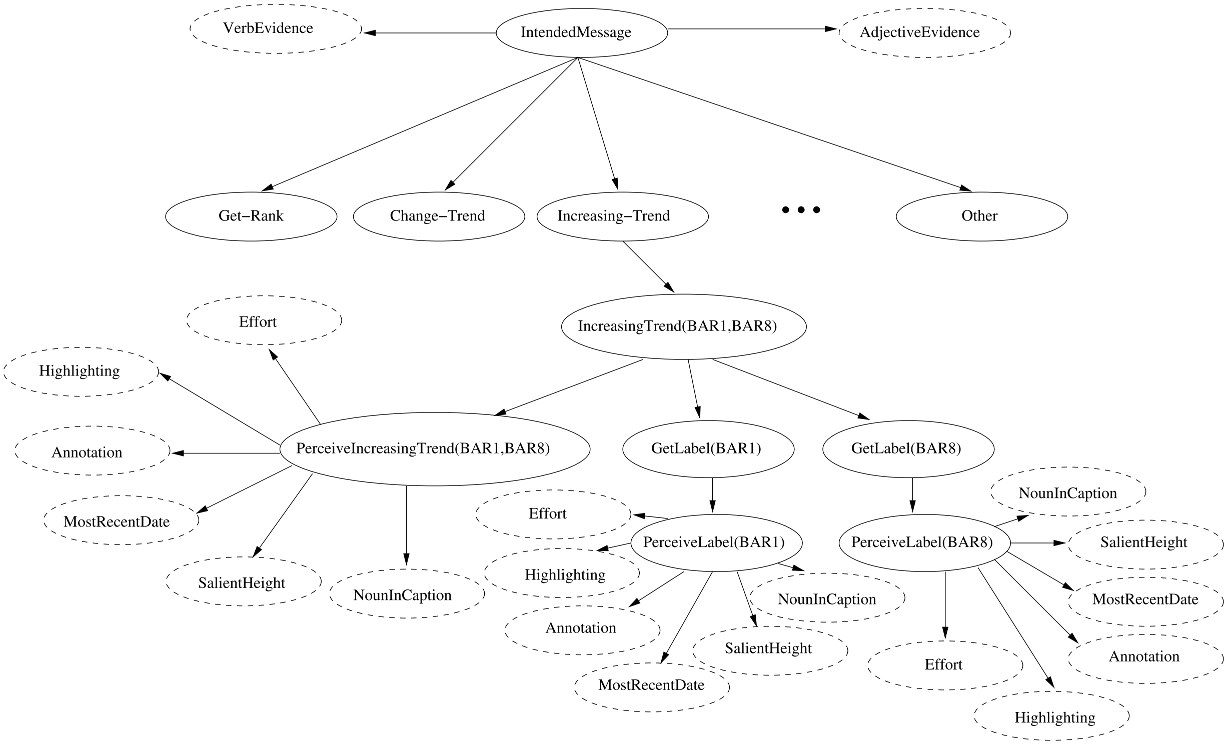


Fig. 23. Subnetwork with evidence nodes.

The *effort* evidence nodes will record the estimated perceptual task effort as *easy* if the estimated effort is less than or equal to 750, *medium* if the estimated effort is less than or equal to 1350, *hard* if the estimated effort is greater than 1350 or *impossible* if the perceptual task cannot be performed on the given bar chart.[20] For example, PerceiveLabel(BAR1) has an estimated effort of 680, so the effort node attached to this task will have a value of *easy*.

Once the evidence is added to the network, the evidence is propagated through the network by recomputing the probabilities at each node. After all of the evidence is recorded, the high-level message inferred by the graphic is the value of the IntendedMessage node with the highest probability. In the case of the network constructed for the graphic shown in Fig. 19, the intended message is inferred to be IncreasingTrend. The inferred instantiation of the message is the child node of the IncreasingTrend node with the highest probability — in this case, it is IncreasingTrend(BAR1, BAR8), with a probability of 99.6%.

9. Evaluation

In order to evaluate the performance of our intention recognition module, we performed two different types of evaluations. The goal of the first evaluation was to measure the system's ability to match the results of the coders who tagged our corpus of bar charts with the intended message of each bar chart. The goal of the second evaluation was to gauge the level of acceptance that users of the system might have for the responses produced by the system. The next section discusses the methods employed in these evaluations and the results of the evaluations.

9.1. Leave-one-out cross validation

The first evaluation of our methodology for intention recognition for bar charts involved measuring the system's output against that of human coders. We used leave-one-out cross validation to evaluate our system. We performed a series of experiments in which each graphic in our corpus was selected once as the test graphic. The data pertaining to the test graphic was removed from the calculation of the probabilities used in the conditional probability tables. The XML representation of the test graphic was augmented by the Caption Tagging Module and then presented to our Intention Recognition Module, and the Bayesian network was constructed with probability tables acquired from the remaining graphics. The system was judged to fail if either its top-rated hypothesis did not match the intended message that was assigned to the graphic by the coders or the probability rating of the system's top-rated hypothesis did not exceed 50%. This is a more stringent requirement than simply choosing the highest rated intention, but because our planned applications include summarizing graphics for users who could not access the content themselves (for example, blind users), we only wanted to utilize hypotheses in which the system had a high degree of confidence.

We computed the system's overall performance as the percentage of experiments in which the system's top-rated hypothesis matched that of the coders. Our overall accuracy was 79.1% — that is, the system's hypothesis matched the coders for 87 out of 110 bar charts. As a baseline, the most commonly occurring message category (at 23.6%) was IncreasingTrend. However, note that the baseline of 23.6% indicates the message category only — in the evaluation, the system's hypothesis had to match not only the high-level message category tagged by the coders but also the instantiation of all of the parameters of the message (for example, if the coders tagged a Rising-Trend from BAR1 to BAR8, and the system recognized a Rising-Trend from BAR1 to BAR7, this was counted as an error). We believe that our success rate provides strong evidence of the success of our methodology.

Some of the errors in our system's performance arise from what we view as special cases or "design quirks", and may be fairly hard to correct. Consider, for example, the bar chart shown in Fig. 24 which shows net income data for Levi Strauss. The eighth bar (representing 1999) is annotated with its value. Our system considers this to be evidence of its salience. However, it appears that the designer annotated the bar (which cannot be seen in the graphic) with its value because its extremely low value made it appear as if there was no value recorded for that year. The system attempts to come up with hypotheses involving this bar as a salient parameter, rather than inferring the *ChangeTrend* message that the coders recognized.

However, the majority of the errors made by our system are caused by sparseness of the data used for training the system; for example, if we only have a single graphic using a verb from the class containing verbs like "recover" to indicate a changing trend, then once the leave-one-out validation excludes the data pertaining to this graphic, we will not have any evidence linking the verb class to the change trend category of intention, and we may get an incorrect result when evaluating the graphic.

One issue that has been largely overlooked in plan recognition is the evaluation of the impact of the different evidence sources. In developing our system, we first identified every feature that we thought might influence the recognized message. Then we manually examined a wide variety of bar charts and eliminated those features that did not appear to have a significant impact; for example, we considered *number of bars in the bar chart* as a possible feature but chose not to include it as evidence in our Bayesian network. Once the recognition system was implemented and tested, we evaluated how each kind of evidence impacted system performance by 1) examining system performance with only one kind of evidence, and 2) examining the degradation in system performance when a particular kind of evidence was disabled [10]. Perceptual task effort and adjective evidence were the evidence sources that respectively had the greatest and least impact on system

[20] The thresholds for effort were assigned based on naturally occurring breaks within the collected eyetracking data.

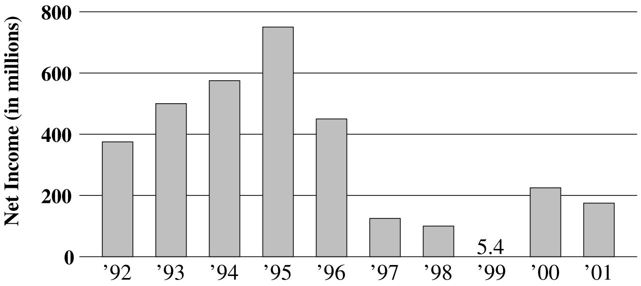


Fig. 24. Bar chart showing net income data.[21]

performance. Evaluations such as this can both confirm the impact of some features, and suggest the removal of features that appear to have little impact.

9.2. Evaluation by human subjects

In our second evaluation, we performed an experiment in which human subjects were asked to examine a set of bar charts and rate a posited primary intention for each bar chart. This type of evaluation is less demanding for the subjects than asking them to code the intentions of many bar charts, and allows us to broaden the participation in our evaluation. It also poses a slightly different question than the previous evaluation — rather than asking the subjects to identify the intention, the participants were asked whether they were satisfied with a provided intention. Thus, the goal of this experiment was to determine whether the intentions being inferred by our system would meet the approval of users. Seventeen undergraduate students took part in the experiment, which contained twenty-seven bar charts. For each bar chart, the participants were asked to answer a set of questions, including their level of agreement with the stated primary intended message of the graphic (strongly agree, agree, not sure, disagree, strongly disagree), and follow-up questions for cases where they did not agree with the stated message.

To produce the statements describing the proposed intended messages of the bar charts, we wrote a template for each category of high-level message and manually inserted information from the bar chart into the template. For twenty of the twenty-seven bar charts, the statement matched the intention hypothesized by our system. For the remaining seven bar charts, we proposed messages that did not match the hypotheses of our system. These "non-matching" cases were interspersed throughout the experiment. They were included to ensure that the participants did not become accustomed to automatically agreeing with the proposed messages, and to demonstrate that the participants were actively engaged in evaluating the system's hypothesized message.

When calculating the results of the experiment, we assigned numeric values to the scale in the first question. An answer of "strongly agree" was counted as a four, "agree" a three, "not sure" a two, "disagree" a one, and "strongly disagree" a zero. The answers of each of the seventeen participants were then averaged to come up with an overall assessment of the participants' level of agreement with the proposed message for a particular bar chart. For the twenty graphs where the proposed message matched the output of our system, we expected the majority of participants to agree with the proposed message. This was, indeed, the case, given that the average agreement score for the twenty graphs was 3.33 (a value between "agree" and "strongly agree" on our scale) with a standard deviation of 1.02 and a 95% confidence interval of .108. The agreement score for the individual bar charts ranged from 2.59 to 3.94.

For the seven bar charts where the proposed message did not match the output of our system, we expected the majority of participants to disagree with the proposed message. And, for the most part, this was the case. The average agreement score for the seven graphs was 1.19 (a value between "not sure" and "disagree" on our scale) with a standard deviation of 1.46 and a 95% confidence interval of .261. The agreement score for the individual bar charts ranged from 0.18 to 1.53. While the numbers certainly show that the majority of the participants disagreed with the erroneous messages, the strength of their disagreement is not quite as strong as expected. Given the nature of some of the responses, we feel that a small number of the participants may have either hurried through the experiment or agreed with any statements that were accurate given the data being displayed.

The results of our experiment, particularly the average level of agreement with the intended messages recognized by our system, demonstrate the success of our system in inferring the intended message of a bar chart. Utilizing these recognized intentions as the basis of a summary of a graphic should, therefore, produce summaries which would be satisfactory to a majority of users.

[21] This is based on a bar chart from the Lancaster Intelligencer Journal on April 9, 2002.

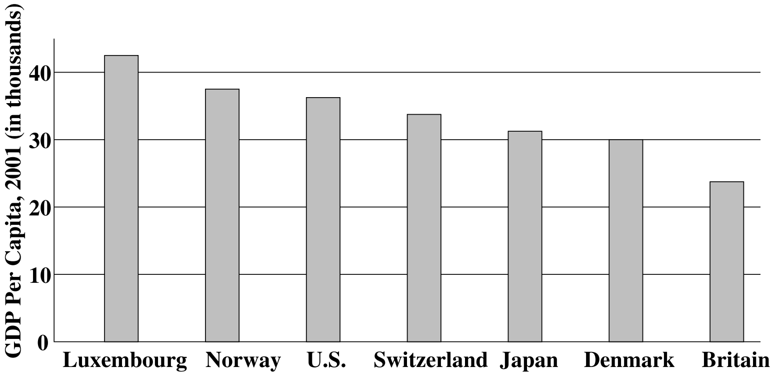


Fig. 25. Information graphic example.

10. Examples

This section presents several examples which illustrate how different kinds of evidence impact our system's hypothesis as to the intended message of a bar chart. In order to clearly show the impact of the choices made by a graphic designer when constructing a bar chart, we will consider multiple variations of bar charts displaying the same underlying data.[22] In each case, the Visual Extraction Module (VEM), shown in the system architecture diagram in Fig. 3, was given the gif representation of the graphic and produced an XML representation that was passed to the Caption Tagging Module which augmented it with communicative signals extracted from the caption (if any). The Intention Recognition Module then took this augmented XML as input and produced its hypothesis about the graphic's intended message after interacting with the Analysis for Perceptual Task Effort Module.

10.1. A bar chart without salient elements

As our first example, consider the bar chart shown in Fig. 25, which displays the gross domestic product per capita for a number of countries. Note that there are no communicative signals within the graphic other than perceptual task effort, and so the Bayesian network is constructed beginning with the addition of nodes representing the ten low effort perceptual tasks.

Given the bar chart as it appears in Fig. 25, our system hypothesizes that the graphic is intended to convey the relative rank in GDP of the various countries displayed in the bar chart and assigns this intention a probability of 87%. Other possibilities also have some probability assigned to them. For example, the intention of conveying that Luxembourg has the highest GDP is assigned a probability of 12.4% because the bars are in sorted order according to height, thus making it relatively easy for the viewer to recognize the maximum, and because finding the entity in the graphic with the maximum value is a fairly common intention (occurring approximately 22.7% of the time in our corpus). However, there is no other evidence suggesting that the bar representing the maximum value is salient (such as that bar being highlighted, or "Luxembourg" being mentioned in the caption). Note that trends are not considered as possibilities for this graph since the values on the independent axis are not ordinal, and so a trend cannot be detected.

10.2. A bar chart with a single salient element

The bar chart shown in Fig. 25 had no salient elements, and so the starting point for the plan inference process was solely based on perceptual task effort. Suppose, however, that the bar representing the U.S. was darker than the other bars (as shown in Fig. 26), thus making this element salient and thereby providing strong evidence that it plays a role in the intended message of the graphic. The set of tasks used to begin the inference process for this bar chart will include not only the set of ten low effort perceptual tasks but also any perceptual tasks involving the salient element (the bar representing the U.S.).

The resultant Bayesian network hypothesizes that the intended message of the graphic shown in Fig. 26 is to convey that the U.S. ranks third among the countries shown in the bar chart with respect to GDP per capita. This hypothesis is believed to be extremely probable, given its calculated probability of 99.5%.

10.3. A bar chart with two salient elements

Elements of the graphic could also be made salient in other ways, such as through annotations. Suppose that the bar representing the U.S. was still darker than the other bars, but that the bars representing the U.S. and Japan (and only those

[22] The data displayed in the example graphics is based on a bar chart that appeared in the June 28, 2004 issue of U.S. News & World Report magazine.

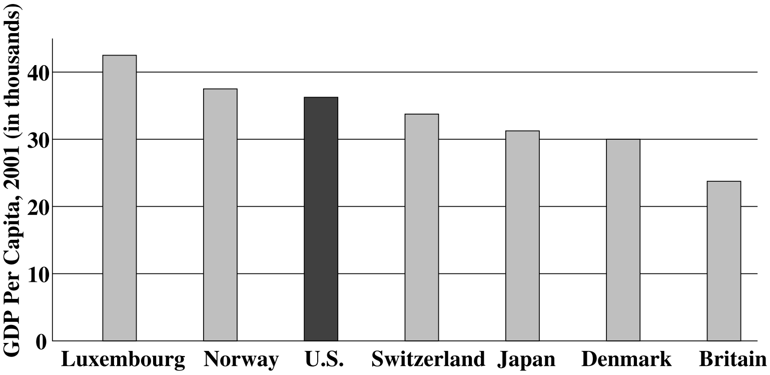


Fig. 26. Information graphic example with a single salient element.

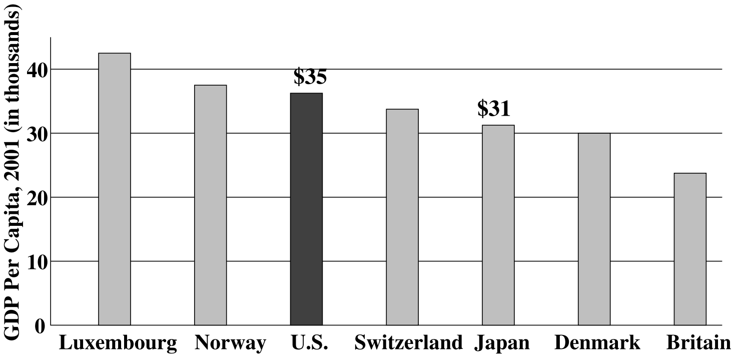


Fig. 27. Information graphic example with two salient elements.

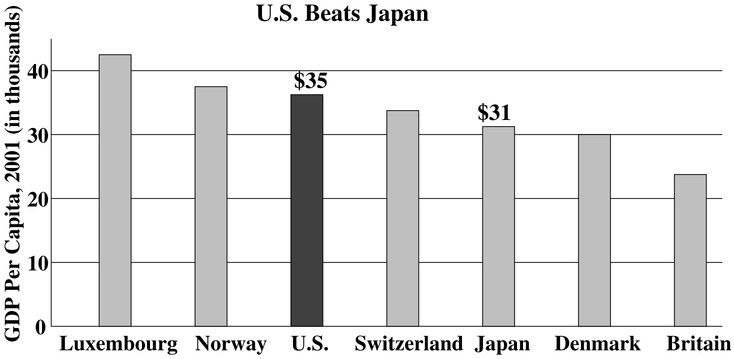


Fig. 28. Information graphic example with a helpful caption.

bars) were annotated with their exact values, as shown in Fig. 27. Here the evidence still suggests the salience of the U.S., as in the previous example, but also suggests that Japan is salient.

The fact that two bars are now salient provides evidence against intentions involving only the U.S. and will favor hypotheses involving both bars. Thus it is not surprising that the system hypothesizes the graphic's intended message to be the relative difference (and the degree of that difference) between the GDP of the U.S. and Japan and assigns it a probability of 87.3%.

10.4. A bar chart with a helpful caption

So far, none of our examples have included a caption. Suppose, however, that the caption of our previous example was "U.S. Beats Japan" (as shown in Fig. 28). In this case, the XML produced by the Visual Extraction Module is augmented by the Caption Tagging Module to include evidence extracted from the caption. The verb "beat" (the root form of "beats") is a member of one of our identified verb classes, and therefore the presence of this verb class is included in the augmented

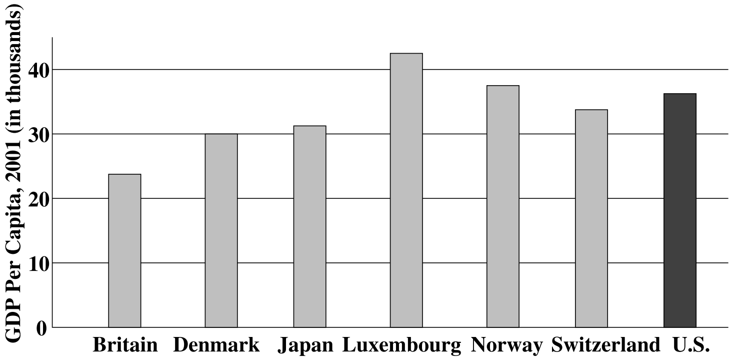


Fig. 29. Information graphic example with different perceptual task effort.

XML representation of the graphic. The nouns "U.S." and "Japan" match the labels of bars in the bar chart, so they are also included in the augmented XML.

Because the bars representing the U.S. and Japan were already salient in the previous example, the appearance of their labels in the caption does not cause any additional tasks to be included in the set of tasks used to begin the plan inference process. However, the inclusion of "U.S." and "Japan" in the caption *does* provide additional evidence of the importance of these elements to the intended message of the graphic. The additional communicative signal provided by the verb "beats", as well as having both labels appear in the caption, strengthens the system's belief in its hypothesis of the relative difference and degree message, and it now assigns it a probability of 99.6% (as opposed to 87.3% without the caption).

10.5. The impact of perceptual task effort

Now consider a significant variation of the graphic design. Suppose, again, that the bar representing the U.S. was darker than the other bars. But now suppose that the bars were sorted by the alphabetical order of their labels, rather than by descending order of their height. This variation is shown in Fig. 29. The perceptual task of determining the rank of the U.S. is now estimated as being *hard* (or difficult) to perform. This new evidence results in the system assigning a probability of only 9.1% to the Get-Rank message. The system hypothesizes the most likely message to be a comparison of relative difference and degree involving the bar representing the U.S.; however, this top-rated instantiation is only assigned a probability of 47.7%, and we only accept hypotheses that are assigned a probability greater than 50%. Therefore, we conclude that the system has been unable to come up with a hypothesis regarding the graphic designer's intended message for this bar chart. This is probably reasonable given the conflicting communicative signals regarding salience and perceptual task effort present in this graphic — that is, 1) the salience of the bar for U.S. suggests that U.S. plays a prominent role in the graphic's message, 2) but the difficulty of the perceptual task of determining the rank of the U.S. makes it unlikely that the graphic conveys a Get-Rank message, 3) and no other bars being salient (besides the bar for U.S.) makes it unlikely that the graphic is conveying a comparison message.

11. Related work

The communicative intentionality of graphics is demonstrated by work on multimodal document generation and by work on multimodal dialogue systems. Research has addressed the generation of multimodal documents that contain images or graphics as well as text [67,23,65,5,31]. The early work described in [67] treated the generation of a multimodal document as an incremental planning process in which the text and graphics are planned to work together to achieve a communicative goal. Other work, such as the AutoBrief project [31] has followed this paradigm and influenced our research. AutoBrief proposed extending speech act theory to the generation of multimodal presentations. The system used high-level planning to transform a high-level goal into communicative acts, which then were assigned to a particular medium (text or graphics) using media allocation rules. If a communicative goal was to be expressed as an information graphic, then it was translated into perceptual and cognitive tasks which the graph design must enable for the viewer. Our research inverts this process. Given an information graphic, we use plan inference to recognize the graph designer's plan for the viewer and thereby recognize the graphic's high-level communicative goal. In the case of multimodal dialogue systems, presentation planning divides the system's response into a coordinated set of goals that are achieved via speech, graphics, or facial expressions [66,53,41,8]. This work has generally focused on navigation systems that display maps and/or landmarks as part of their output, although a wider variety of graphical elements is considered in [53].

The graph comprehension work that has influenced our research was discussed in Section 5.1.2. Cognitive psychologists have continued to study graph comprehension, with an emphasis on how humans perceive and interpret graphical information and the factors that impact graph comprehension; examples of recent work include [64] which contends that spatial cognition is essential in a model of complex graph comprehension, [61] which explores visual and cognitive integration, and

[34] which investigates the affect on information processing of the layout of graphics with respect to related text. These research efforts are influencing our ongoing work on understanding more complex information graphics.

There has been little work on automating the summarization of information graphics. Futrelle and Nikolakis [25] developed a constraint grammar for parsing vector-based visual displays, and Futrelle has considered the problem of constructing a graphic that is a simpler form of one or more graphics in a document [24]. However, the end result is itself a graphic, not a representation of the graphic's intended message. Yu et al. [70] used pattern recognition techniques to summarize interesting features of time series data from a gas turbine engine. However, the graphs were automatically generated displays of the data points and did not have any intended message. Our work is the first to address the *understanding* of an information graphic by identifying its high-level communicative message.

Multimodal document generation has also addressed the problem of constructing captions for graphics in a multimodal document. Mittal et al. [52] employed a planner that applied discourse strategies captured in plan operators to generate captions for complex graphical displays; the purpose of the captions was to enable the viewer to understand how the graphic elements mapped to data and thus be able to extract information and make deductions from the graph. Fasciano and LaPalme [23] classified intentions for information graphics into five broad categories: reading, comparison, evolution, correlation, and distribution. In their work on caption generation, Corio and LaPalme [18] identified seven kinds of information provided by captions (such as focusing on particular raw data or drawing a conclusion from the graphic) and estimated how often each of the five broad categories of graphic intentions were associated with each of the seven kinds of captions. For example, a focusing caption was most often associated with a graphic whose intention fell into the comparison category. In our work, a focusing caption would most likely give the labels of one or two bars, thereby making them salient and thus leading to an intended message in which the bar or bars played a role (such as a Get-Rank or Relative-Difference-Degree message). Corio and LaPalme then developed empirically derived rules for generating captions. However, caption generation differs from our research in that the caption generator has access to the data that was used to generate the graphic and to the intention underlying the graphic, whereas our work is concerned with using the communicative signals in the graphic to identify the graphic's communicative goal.

12. Current and future works

We are currently extending our intention recognition methodology to pie charts as well as more complex information graphics such as line graphs and grouped bar charts. These graphics pose additional problems that must be addressed. For example, a line graph can consist of many short segments connecting individual data points; to reason about the graphic's message, we must first transform it into a sequence of visually distinguishable trends. We are using machine learning on attributes of the line graph to develop a model for performing such a segmentation [69]. In the case of grouped bar charts, the possible messages and perceptual tasks are more complex than in the case of simple bar charts, and a richer model of perceptual task effort is required [7]. Moreover, while simple bar charts appear to have a single intended message,[23] grouped bar charts often strongly convey more than one message, although one of the messages typically stands out and is more important to the context than the other. Thus our future work will address the recognition of both primary and secondary messages in grouped bar charts.

We are also pursuing two applications of our message recognition system. Our digital libraries project will utilize the recognized message to store graphics and facilitate their retrieval. We will also investigate how to integrate the intended message of a graphic with a summary of the article's text in order to produce a more complete summary of a multimodal document. Our assistive technology project is using our message recognition system to provide individuals with sight impairments with a brief summary of a graphic encountered in a multimodal document and with the capability for requesting more in-depth information about the graphic. Researchers such as Huff and Geis [37] and Mittal [51] have noted that information graphics can be intentionally deceptive; for example, the graph designer might truncate the y-axis in order to make the relative differences in bar height appear greater. Thus our work on conveying information graphics to sight-impaired individuals may need to take such phenomena into account.

In addition, the issue of "seeding" the construction of the Bayesian network with the lowest effort tasks and salient tasks, as described in Section 8.1, warrants additional investigation. For example, it would be interesting to explore, as system limitations are lifted, whether performance could be improved by including more tasks in the initial seed set used to construct the network. Tests could also be conducted to explore whether limiting the initial set further impacts system accuracy.

13. Conclusion

While identifying the intention of an utterance has played a major role in natural language understanding, this work is the first to extend intention recognition to the domain of information graphics. As Clark noted, language is more than just words. It is any "signal" (or lack of signal when one is expected), where a signal is a deliberate action that is intended to convey a message [15], and a tenet of this work is the belief that information graphics are a form of language.

[23] This does not mean that other information cannot be gleaned from examination of the graphic. However, we argue that in almost every case, simple bar charts have one message that is the communicative goal of the graphic.

This paper has shown the relation between simple bar charts in popular media and more traditional forms of language such as natural language utterances. In doing so, this paper has

- identified the kinds of communicative goals that are achieved by bar charts;
- determined the kinds of communicative signals in bar charts that help achieve these communicative goals;
- shown how these communicative signals can be extracted and computationalized;
- demonstrated how plan inference techniques that have been used for understanding natural language utterances can also be used for recognizing the communicative message of a simple bar chart;
- provided an implemented Bayesian network methodology for reasoning about the communicative signals in a bar chart and hypothesizing its intended message;
- explored the practical constraints of framing a problem within a Bayesian network methodology and presented reasonable solutions, such as seeding the network with likely nodes and analyzing the impact of various evidence sources.

We have demonstrated the success of our methodology through evaluations of our implemented system. The intended message of a bar chart that is inferred by our system can be used to facilitate access to this information resource for a variety of users, including users of digital libraries and visually impaired users.

Acknowledgements

We would like to thank Daniel Chester for his work in implementing the Visual Extraction Module (VEM), and Seniz Demir and Peng Wu for their assistance with implementing various other parts of the system.

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

References

[1] D. Albrecht, I. Zukerman, A. Nicholson, Bayesian models for keyhole plan recognition in an adventure game, User Modeling and User-Adapted Interaction 8 (1–2) (1998) 5–47.

[2] J.F. Allen, A plan-based approach to speech act recognition, Ph.D. thesis, University of Toronto, Toronto, Ontario, Canada, 1979.

[3] J.F. Allen, C.R. Perrault, Analyzing intention in utterances, Artificial Intelligence 15 (1980) 143–178.

[4] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, A. Stolcke, Prosody-based automatic detection of annoyance and frustration in human–computer dialog, in: Proceedings of the International Conference on Spoken Language Processing, 2002, pp. 2037–2040.

[5] J. Bateman, J. Kleinz, T. Kamps, K. Reichenberger, Towards constructive text, diagram and layout generation for information presentation, Computational Linguistics 27 (3) (2001) 409–449.

[6] N. Blaylock, J. Allen, Generating artificial corpora for plan recognition, in: Proceedings of the International Conference on User Modeling, 2005, pp. 179–188.

[7] R. Burns, S. Elzer, S. Carberry, Modeling relative task effort for grouped bar charts, in: Proceedings of the Annual Meeting of the Cognitive Science Society, July 2009, pp. 2292–2297.

[8] C. Callaway, Non-localized, interactive multimodal direction giving, in: Proceedings of the Workshop on Multimodal Output Generation, 2007, pp. 41–50.

[9] S. Carberry, Plan Recognition in Natural Language Dialogue, ACL-MIT Press Series on Natural Language Processing, MIT Press, Cambridge, MA, 1990.

[10] S. Carberry, S. Elzer, Exploiting evidence analysis in plan recognition, in: Proceedings of the 11th International Conference on User Modeling (UM2007), June 2007, pp. 7–16.

[11] S.K. Card, T.P. Moran, A. Newell, The Psychology of Human–Computer Interaction, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1983.

[12] E. Charniak, R. Goldman, A Bayesian model of plan recognition, Artificial Intelligence 64 (1993) 53–79.

[13] D. Chester, S. Elzer, Getting computers to see information graphics so users do not have to, in: Proceedings of the 15th International Symposium on Methodologies for Intelligent Systems, 2005, pp. 6–68.

[14] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Borriello, B. Hemingway, P. Klasnja, K. Koscher, J. Landay, J. Lester, D. Wyatt, D. Haehnel, The mobile sensing platform: An embedded activity recognition system, IEEE Pervasive Computing 7 (2) (2008) 32–41.

[15] H. Clark, Using Language, Cambridge University Press, 1996.

[16] W.S. Cleveland, Graphical Perception, The Elements of Graphing Data, Hobart Press, 1994 (Chapter 4).

[17] P.R. Cohen, C.R. Perrault, J.F. Allen, Beyond question answering, in: W. Lehnert, M. Ringle (Eds.), Strategies for Natural Language Processing, Lawrence Erlbaum Associates, 1981, pp. 245–274.

[18] M. Corio, G. LaPalme, Generation of texts for information graphics, in: Proceedings of the 7th European Workshop on Natural Language Generation EWNLG'99, 1999, pp. 49–58.

[19] M.J. Druzdzel, L.C. van der Gaag, Building probabilistic networks: where do the numbers come from?, IEEE Transactions on Knowledge and Data Engineering 12 (2000) 481–486.

[20] 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.

[21] S. Elzer, N. Green, S. Carberry, J. Hoffman, Incorporating perceptual task effort into the recognition of intention in information graphics, in: Diagrammatic Representation and Inference: Proceedings of the Third International Conference on the Theory and Application of Diagrams, in: LNCS, vol. 2980, Springer, Berlin/Heidelberg, 2004, pp. 255–270.

[22] S. Elzer, N. Green, S. Carberry, J. Hoffman, A model of perceptual task effort for bar charts and its role in recognizing intention, User Modeling and User-Adapted Interaction 16 (1) (2006) 1–30.

[23] M. Fasciano, G. LaPalme, Intentions in the coordinated generation of graphics and text from tabular data, Knowledge and Information Systems 2 (3) (2000) 310–339.

[24] R. Futrelle, Summarization of diagrams in documents, in: I. Mani, M. Maybury (Eds.), Advances in Automated Text Summarization, MIT Press, 1999, pp. 403–421.

[25] R. Futrelle, N. Nikolakis, Efficient analysis of complex diagrams using constraint-based parsing, in: Proceedings of the Third International Conference on Document Analysis and Recognition, 1995, pp. 782–790.

[26] C. Geib, R. Goldman, Plan recognition in intrusion detection systems, in: Proceedings of DARPA Information Survivability Conference and Exposition, 2001, pp. 46–55.

[27] C. Geib, R. Goldman, Probabilistic plan recognition for hostile agents, in: Proceedings of the 14th International Florida Artificial Intelligence Research Society Conference, 2001, pp. 580–584.

[28] C. Geib, R. Goldman, A probabilistic plan recognition algorithm based on plan tree grammars, Artificial Intelligence 173 (11) (2009) 1101–1132.

[29] P. Gorniak, D. Roy, Probabilistic grounding of situated speech using plan recognition and reference resolution, in: Proceedings of the International Conference on Multimodal Interfaces, 2005, pp. 138–143.

[30] P. Gorniak, D. Roy, Situated language understanding as filtering perceived affordances, Cognitive Science 31 (2) (2007) 197–231.

[31] N. Green, G. Carenini, S. Kerpedjiev, J. Mattis, J. Moore, S. Roth, Atuobrief: An experimental system for the automatic generation of briefings in integrated text and graphics, International Journal of Human–Computer Studies 61 (1) (2004) 32–70.

[32] H.P. Grice, Utterer's meaning and intentions, Philosophical Review 68 (1969) 147–177.

[33] B. Grosz, C. Sidner, Attention, intentions, and the structure of discourse, Computational Linguistics 12 (3) (1986) 175–204.

[34] J. Holsanova, N. Holmberg, K. Holmqvist, Reading information graphics: The role of spatial contiguity and dual attentional guidance, Applied Cognitive Psychology (2008) 1215–1226.

[35] E. Horvitz, T. Paek, A computational architecture for conversation, in: Proceedings of the 7th International Conference on User Modeling, 1999, pp. 201–210.

[36] M. Huber, E. Durfee, M. Wellman, The automated mapping of plans for plan recognition, in: Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, 1994, pp. 344–351.

[37] D. Huff, I. Geis, How to Lie With Statistics, W.W. Norton & Company, 1993.

[38] A. Huntemann, E. Demeester, G. Vanacker, D. Vanhooydonck, J. Philips, H. Van Brussel, M. Nuttin, Bayesian plan recognition and shared control under uncertainty: Assisting wheelchair drivers by tracking fine motion paths, in: IEEE International Conference on Intelligent Robots and Systems, 2007, pp. 3360–3366.

[39] G. Iverson, M. Gergen, Statistics: The Conceptual Approach, Springer-Verlag, New York, 1997.

[40] B.E. John, A. Newell, Toward an engineering model of stimulus response compatibility, in: R.W. Gilmore, T.G. Reeve (Eds.), Stimulus-Response Compatibility: An Integrated Approach, North-Holland, New York, 1990, pp. 107–115.

[41] M. Johnston, S. Bangalore, Matchkiosk: A multimodal interactive city guide, in: Proceedings of ACL Poster and Demonstration Session, 2004, pp. 222–225.

[42] G. Kaminka, D. Pynadath, M. Tambe, Monitoring teams by overhearing: A multi-agent plan-recognition approach, Journal of Artificial Intelligence Research 17 (2002) 83–135.

[43] S. Kerpedjiev, S.F. Roth, Mapping communicative goals into conceptual tasks to generate graphics in discourse, in: Proceedings of Intelligent User Interfaces, 2000, pp. 157–164.

[44] S.M. Kosslyn, Understanding charts and graphs, Applied Cognitive Psychology 3 (1989) 185–226.

[45] S.M. Kosslyn, Elements of Graph Design, W.H. Freeman, 1994.

[46] L. Lambert, Recognizing complex discourse acts: A tripartite plan-based model of dialogue, Ph.D. thesis, University of Delaware, June 1993.

[47] J. Larkin, H. Simon, Why a diagram is (sometimes) worth ten thousand words, Cognitive Science 11 (1987) 65–99.

[48] G.L. Lohse, A cognitive model for understanding graphical perception, Human–Computer Interaction 8 (1993) 353–388.

[49] M. Maragoudakis, A. Thanopoulos, N. Fakotakis, Meteobayes: Effective plan recognition in a weather dialogue system, IEEE Intelligent Systems 22 (2007) 67–77.

[50] Merriam-Webster, Merriam-Webster On-Line Thesaurus, http://www.webster.com, last accessed Sept. 8, 2009.

[51] V. Mittal, Visual prompts and graphical design: A framework for exploring the design space of 2-d charts and graphs, in: Proceedings of the Fourteenth National Conference on Artificial Intelligence, 1997, pp. 57–63.

[52] V. Mittal, J. Moore, G. Carenini, S. Roth, Describing complex charts in natural language: A caption generation system, Computational Linguistics 24 (3) (1998) 432–467.

[53] J. Muller, P. Poller, V. Tschernomas, Situated delegation-oriented multimodal presentation in SmartKom, in: Proceedings of AAAI Workshop on Intelligent Situated-Aware Media Presentations, 2002, pp. 1–8.

[54] Norsys Software Corp., Netica, http://www.norsys.com/netica.html, last accessed on June 23, 2010.

[55] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufman, San Mateo, CA, 1988.

[56] D. Peebles, P.C.-H. Cheng, Modeling the effect of task and graphical representation on response latency in a graph reading task, Human Factors 45 (2003) 28–46.

[57] M. Pollack, A model of plan inference that distinguishes between the beliefs of actors and observers, in: Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics, New York, NY, 1986, pp. 207–214.

[58] M. Pollack, L. Brown, D. Colbry, C. McCarthy, C. Orosz, B. Peintner, S. Ramakrishnan, I. Tsamardinos, Autominder: An intelligent cognitive orthotic system for people with memory impairment, Robotics and Autonomous Systems 44 (3–4) (2003) 273–282.

[59] D. Pynadath, M. Wellman, Probabilistic state-dependent grammars for plan recognition, in: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, 2000, pp. 507–514.

[60] X. Qin, W. Lee, Attack plan recognition and prediction using causal networks, in: Proceedings of the 20th Annual Computer Security Applications Conference, 2004, pp. 370–379.

[61] R. Ratwani, J. Trafton, D. Boehm-Davis, Thinking graphically: Connecting vision and cognition during graph comprehension, Journal of Experimental Psychology: Applied 14 (2008) 36–49.

[62] J. Ruppenhofer, M. Ellsworth, M. Petruck, C. Johnson, J. Scheffczyk, FrameNet II: Extended theory and practice, http://framenet.icsi.berkely.edu, 2006.

[63] J.E. Russo, Adaptation of cognitive processes to eye movement systems, in: J.W. Senders, D.F. Fisher, R.A. Monty (Eds.), Eye Movements and Higher Psychological Functions, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1978, pp. 89–109.

[64] S. Trickett, G. Trafton, Toward a comprehensive model of graph comprehension: Making the case for spatial cognition, in: Proceedings of the 4th International Conference on the Theory and Application of Diagrams, 2006, pp. 286–300.

[65] K. Van Deemter, R. Power, High-level authoring of illustrated documents, Natural Language Engineering 9 (2) (2003) 101–126.

[66] W. Wahlster, Smartkob: Fusion and fission of speech, gestures, and facial expressions, in: Proceedings of the International Workshop on Man–Machine Symbiotic Systems, 2002, pp. 213–225.

[67] W. Wahlster, E. Andre, W. Finkler, H.-J. Profitlich, T. Rist, Plan-based integration of natural language and graphics generation, Artificial Intelligence 63 (1–2) (1993) 387–427.

[68] WordNet, http://wordnet.princeton.edu/main/, last accessed on Sept. 8, 2009.

[69] P. Wu, S. Carberry, S. Elzer, Segmenting line graphs into trends, in: Proceedings of the Twelfth International Conference on Artificial Intelligence, vol. II, 2010, pp. 697–703.

[70] J. Yu, E. Reiter, J. Hunter, C. Mellish, Choosing the content of textual summaries of large time-series data sets, Natural Language Engineering 13 (2007) 25–49.