Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

Classifying Salient Textual Entities in the Headlines and Captions of Grouped Bar Charts

Richard Burns
Dept. of Computer Science
West Chester University
West Chester, PA, 19383

Sandra Carberry
Dept. of Computer Science
University of Delaware
Newark, DE, 19716

Stephanie Elzer Schwartz
Dept. of Computer Science
Millersville University
Millersville, PA, 17551

Abstract

Information graphics, such as grouped bar charts, generally have a communicative message that they are intended to convey when they appear in popular media. Communicative signals are typically designed into the graphic to help convey to the graph viewer these intended messages. We have designed and implemented a system that automatically hypothesizes the intended message of a grouped bar chart from communicative signals that are automatically extracted from the graph. Analysis of our system revealed that textual evidence, such as graph entities mentioned in the headline or caption of the graphic, was the most important piece of evidence in our system. This paper describes a support vector machine classifier that takes a graph and its headlines and captions and predicts whether an entity is linguistically salient.

Introduction

Information graphics, such as simple and grouped bar charts, pie charts, and line graphs are often incorporated into multimodal documents to convey one or more communicative goals (Iverson and Gergen 1997; Green et al. 2004). When these graphs appear in popular media they usually have a high-level message that they are intended to convey. For example the grouped bar chart[1] in Figure 1, which is taken from the popular media magazine *Technology Review*, ostensibly conveys the message that the *"the New Method performs better than the Standard XOM method"*.

We have implemented a system that takes a grouped bar chart as input and outputs the intended message of the chart by reasoning about the communicative signals contained in the graphic.

It is non-trivial to automatically infer the intended message of a graphic in popular media (even with sophisticated NLP techniques) because the graphic's message is often not contained in the graphic's caption or repeated in an article accompanying the graphic (Carberry, Elzer, and Demir 2006).

[1] *Grouped* bar charts differ from *simple* bar charts in that they have an additional grouping dimension.

Just as Secure, but Faster
Programs that, say, compress files and compile code run faster with Jun Yang's new method than they do on a standard decryption (XOM) chip.

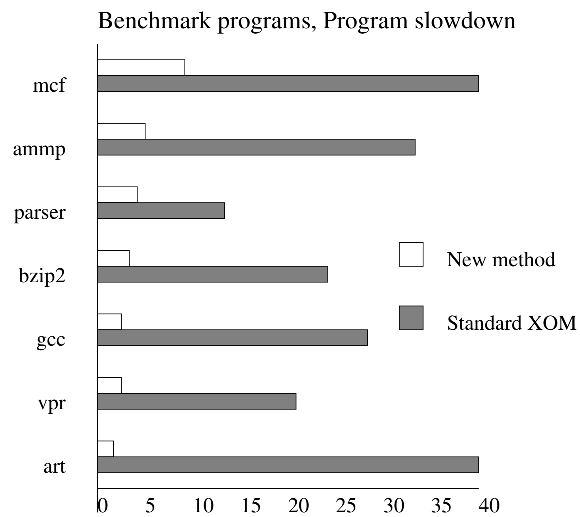

Figure 1: Comparators in the caption help signal the most linguistically salient entity. Graphic from *Technology Review*, July 2005.

Our analysis of our system's accuracy revealed that some types of communicative evidence were more important than others (Burns, Carberry, and Schwartz 2014). Specifically the *Linguistic Classifier* in our system, a module that analyzes the headline and caption of a graph and automatically identifies which entities are most salient linguistically, had the greatest impact.

This paper presents an overview and evaluation of the *Linguistic Classifier* module in our overall grouped bar chart intention recognition system.

Grouped Bar Chart Messages and Communicative Signals

We collected 330 grouped bar charts (including their captions) and examined them to identify the types of intended messages conveyed by grouped bar charts, as well as the communicative signals utilized by graphic designers to assist in communicating these messages.

Many different types of communicative signals can be designed into a grouped bar chart to help communicate some message, including: coloring a bar differently, ordering the groups in a meaningful way, or crafting the caption or headline to make one entity linguistically salient.

For grouped bar charts, headline and caption evidence is tricky to process because multiple entities in the graphic may each be present textually. If there is only one graph entity that is mentioned in the caption, then it is linguistically salient. The more interesting graphs are those instances where multiple entities occur in the text, and one of them is the most linguistically salient. Although linguistic evidence is present in both cases, the evidence is communicatively different: in one case an entity is most linguistically salient by default, in the other case an entity is most linguistically salient by having greater linguistic salience than the other mentioned entities. (Ultimately, our overall intention recognition system differentiates the value of the linguistic evidence between the two cases.)

Examples: Entities with Linguistic Salience

In Figure 1, two entities are mentioned in the caption of the graphic, *"New method"* and *"Standard XOM"*. The comparator *than* in the caption helps signal that *"New method"* is being compared with *"Standard XOM"* and is more linguistically salient because it precedes the comparator.

In another example, the grouped bar chart in Figure 2 contains all three series entities in the headlines and captions of the graphic, *"Reebok"*, *"Adidas"*, and *"Nike"*. The graphic ostensibly conveys a comparison of *"the growth of Adidas against Reebok and Nike"*. While both *Adidas* and *Reebok* are contained in the headline and caption description, *Nike* is only contained in the sub-headline and caption description. The prominence of *Adidas* in the headline rather than in the sub-headline, while preceding the other companies in the headline and caption description, gives it more linguistic salience.[2]

Extracting Linguistic Evidence

In this section, we describe a *Linguistic Classifier* module in our system that processes the caption and headline text of a grouped bar chart and predicts the linguistic salience of each entity. This output is used as part of the communicative evidence in the overall processing of the graphic.

Features

We analyzed the grouped bar charts in our corpus to identify the features and attributes that could intuitively support the process of automatically determining the most linguistically salient entity. These features are described in the following list alongside our intuitions.

[2]"German company" in the sub-headline also refers to *Adidas* but this relation is difficult to automatically extract because of the required domain and world knowledge.

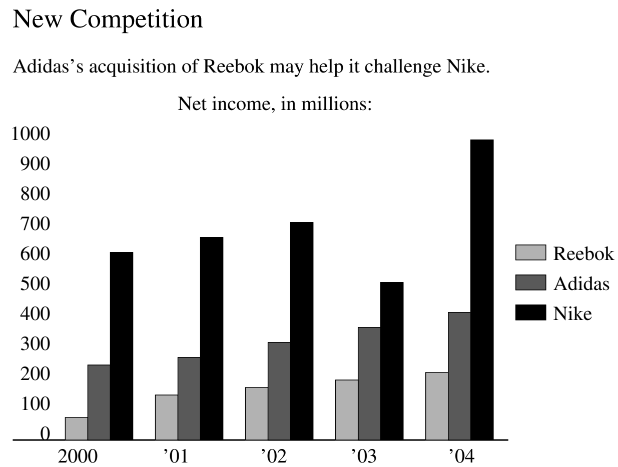

Figure 2: Graphic from *Wall Street Journal*, August 4, 2005.
Article Headline: Leap Forward: For Adidas, Reebok Deal Caps Push to Broaden Urban Appeal
Sub-Headline: Known for Its Engineering, German Company Takes on Nike in Lifestyle Market; Teaming Up With Missy Elliot
Caption Headline: New Competition
Caption Description: Adidas's acquisition of Reebok may help it challenge Nike.

1. **Precedes** - whether the entity in a headline or caption precedes all other entities and is never preceded elsewhere in a different headline or caption.
2. **Numeric** - whether the entity is numeric and is the only entity in the headline and caption text that is numeric.
3. **Prominence** - whether the entity has the greatest *prominence* compared to all entities, where *prominence* considers *which* headlines and captions (headline, sub-headline, etc.) that an entity was present in.
4. **Possessive** - whether the entity is possessive and is the only possessive entity.
5. **First/Last Clause** - whether the entity is the first clause of any headline or caption while other entities are in the last clause, or vice versa.
6. **Main/Contrastive Clause** - whether the entity is in the main clause or a clause that is contrastive in any headline or caption while other entities are in clauses outside of the main or contrastive clause (such as subordinate clauses).
7. **Subject** - whether the entity occurs as a subject in any headline or caption as frequently as any other entity.
8. **Object** - whether the entity occurs as an object in any headline or caption as frequently as any other entity.
9. **Comparator** - whether the entity occurs first in a comparator if there is also another entity on the opposite side of the comparator.

Method

The annotated grouped bar chart corpus provides data for training a linguistic classifier. Along with its intended message, each grouped bar chart was additionally annotated with its most linguistically salient entity if it existed. From the 330 graphs in the corpus, 111 grouped bar charts had multiple entities that occured in the headline or caption text, and 73 had an entity that was identified as the most linguistically salient.

Since the number of grouped bar charts with multiple entities occurring in the text of the headline and caption is limited, leave-one-out cross-validation was applied, where each instance is used once as a test case and instances from all other graphics were used as training cases. The results are averaged together to obtain the accuracy of the model.

The trained model is a support vector machine (SVM)[3].

The most interesting case of analyzing this linguistic classifier is when it is trained and evaluated on the subset of grouped bar charts with multiple entities that occur in the headline or caption text where one entity is the most linguistically salient—the previous set of 73 grouped bar charts.

Within these 73 graphs, there are 189 entities that occur in the text of the headline or caption. (Thus, $\frac{73}{189} = 38.6\%$ entities are the most linguistically salient entity in their graph, and $\frac{116}{189} = 61.4\%$ entities are not the most linguistically salient.)

SVM Model 1: Binary Classification

One way of predicting the most linguistically salient entity for a graph is to treat each entity as an individual instance, and then build a classifier that predicts whether each entity is the most linguistically salient—a binary classification.

Leave-one-out cross-validation was performed to evaluate the classifier; the results are presented in Table 1. In this setup, all entities within a graph were left out while the classifier was trained on only entities from the other graphs. The confusion matrix in Table 2 shows the classifier's performance on both classes (whether an entity was / was not the most linguistically salient). The classifier correctly predicts 54 of the 73 most linguistically salient entities (74.0%), suggesting that the classifier output should be useful evidence for our overall intention recognition system.

Table 1: Cross-validation results of the linguistic binary classifier.

system	*accuracy*
Baseline: classifying entity as not most salient	61.4%
SVM Linear Binary Classifier	80.4%

[3]A benefit of support vector machines is that they work well with high-dimensional data and relatively small training sets—the properties of our training data. Other learning methods with various parameters were also tested including Naïve Bayes and decision trees. The linear SVM performed equally as well as or better than the others.

Table 2: Confusion matrix of the linguistic binary classifier.

	Classified as *not* most salient	**Classified as most salient**
Actual not most salient	98	18
Actual most salient	19	54

SVM Model 2: Entity Ranking

Another way of predicting the most linguistically salient entity for a graph is to rank each entity by its value as output by the linear SVM function. The output of the linear SVM classifier for each entity instance is a real-number value. In the binary classification, the sign of the value determines the predicted class (whether an entity is most linguistically salient). In this case, the ordering of entities by value results in a rank of the entities.

In this case, the prediction of the most linguistically salient entity would be to simply take the top ranked entity. Note that in this setup, the classifier always ranks one and only one entity for each graph as the most linguistically salient, whereas the binary classifier could predict *multiple* entities as the most linguistically salient for a graph or *no* entities as most linguistically salient.

Leave-one-out cross-validation was performed to evaluate the rank classifier and the results are shown in Table 3. Each graph has at least two entities, but some have three, or even more than four entities; thus, a baseline prediction of the expected accuracy of predicting a random entity as the most linguistically salient is less than 50%. The classifier correctly predicts the most linguistically salient entity in 53 of the 73 graphs, an improvement over the baseline by approximately 30%.

Table 3: Cross-validation results of the linguistic rank classifier.

system	*accuracy*
Baseline: expected random prediction of most linguistically salient entity	42.9%
SVM Rank Classifier	72.6%

Grouped Bar Chart Intention Recognition System

Our overall grouped bar chart intention recognition system is implemented as a Bayesian network (Burns et al. 2012). A top-level node at the top of the network represents the intended message of a graphic and the leaves of the network represent the communicative evidence that may be present or absent in the graphic. The network is trained on the corpus of grouped bar charts that we collected. To hypothesize the intended message of a new, unseen graphic, communicative evidence is automatically extracted from the graphic—including processing the graphic through the *Linguistic Classifier* to identify the most linguistically salient entity in its headlines and captions—and then entered in the

Bayesian network. The Bayesian network then outputs the most probable intended message for the graphic.

Using leave-one-out cross-validation, the overall system accuracy for recognizing the intended message of a grouped bar chart is 65.6%[4].

We conducted two experiments to discover the effect that each type of of communicative evidence had on the accuracy of our Bayesian network. When we separately removed different pieces of evidence and measured the degradation in accuracy, the removal of the *Linguistic Classifier* component had the greatest effect: accuracy decreased to 56.1%. Conversely, when we compared the addition of evidence sources to a system configuration with no communicative evidence, the addition of the *Linguistic Classifier* provided the greatest gain: accuracy increased from the baseline 18.8% to 33.6%. Both effects were statistically significant using a two-tailed McNemar test, $p < .01$.

Related Work

Previous work in our research group focused on message recognition for simple bar charts (Elzer, Carberry, and Zukerman 2011) and line graphs (Wu et al. 2010). Grouped bar charts, the type of information graphic in this paper, are another type of information graphic. However, the types of messages communicated in grouped bar charts as well as the communicative signals often used to convey messages are very different than in other information graphic types (Burns et al. 2012). For example, the most statistically important piece of evidence for simple bar charts was not linguistic evidence, but rather a cognitive model that estimated the relative effort required to recognize a message (Carberry and Elzer 2007). While intention recognition systems have been implemented for all three information graphic types, only the grouped bar chart system includes a *Linguistic Classifier* module that contains a model for classifying salient textual entities in headline or captions.

Conclusion

In this paper, we presented our *Linguistic Classifier* module that predicts whether or not an entity that occurs in the headline or caption text of a grouped bar chart is linguistically salient. This output is part of a set of communicative evidence that is used by our system for hypothesizing the intended message of a grouped bar chart. The *Linguistic Classifier* is essential for grouped bar charts.

The work presented here contributes to the following applications: (1) providing alternative access to information graphics for sight-impaired individuals by conveying their high-level content to the user via speech (Demir et al. 2010), (2) retrieving information graphics from a digital library where the graphic's message is used to capture its high-level content (Carberry, Elzer, and Demir 2006), and (3) summarizing multimodal documents, where the summary takes into account information graphics rather than ignoring them or merely considering only their captions (Wu and Carberry 2011).

[4]This accuracy far exceeds a baseline of 18.8% which is selecting the most commonly occurring possible message in the annotated corpus.

References

Burns, R.; Carberry, S.; Elzer, S.; and Chester, D. 2012. Automatically recognizing intended messages in grouped bar charts. In *Proceedings of the Seventh International Conference on the Theory and Application of Diagrams*, 8–22.

Burns, R.; Carberry, S.; and Schwartz, S. E. 2014. Analyzing the effect of communicative evidence in a bayesian system for grouped bar chart message recognition. In *In Proceedings of the 27th Florida Artificial Intelligence Research Society Conference*, 14–17.

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

Carberry, S.; Elzer, S.; and Demir, S. 2006. Information graphics: An untapped resource of digital libraries. In *Proceedings of 9th International ACM SigIR Conference on Research and Development on Information Retrieval*, 581–588.

Demir, S.; Oliver, D.; Schwartz, E.; Elzer, S.; Carberry, S.; and McCoy, K. F. 2010. Interactive sight into information graphics. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 16:1–16:10.

Elzer, S.; Carberry, S.; and Zukerman, I. 2011. The automated understanding of simple bar charts. *Artificial Intelligence* 175(2):526–555.

Green, N. L.; Carenini, G.; Kerpedjiev, S.; Mattis, J.; Moore, J. D.; and Roth, S. F. 2004. Autobrief: an experimental system for the automatic generation of briefings in integrated text and information graphics. *International Journal of Human-Computer Studies* 61(1):32–70.

Iverson, G., and Gergen, M. 1997. *Statistics: The Conceptual Approach*. New York: Springer-Verlag.

Wu, P., and Carberry, S. 2011. Toward extractive summarization of multimodal documents. In *Proceedings of the Workshop on Text Summarization at the Canadian Conference on Artificial Intelligence*, 53–61.

Wu, P.; Carberry, S.; Elzer, S.; and Chester, D. 2010. Recognizing the intended message of line graphs. In *Proceedings of the 6th International Conference on Diagrammatic Representation and Inference*, 220–234.