Exploiting Stereotypes to Eliminate Strategic Bias*

Matthew N. Dailey
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093-0114
(619) 534-5486
fax: (619) 534-7029
mdailey@cs.ucsd.edu

Gregory S. Miller
Glaxo Wellcome, Inc.
5 Moore Dr.
RTP, NC 27709
(919) 990-6890
fax: (919) 941-3704
miller˜gs@glaxo.com

James C. Lester
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206
(919) 515-7534
fax: (919) 515-7896
lester@adm.csc.ncsu.edu

Abstract

Because of the enormous complexity inherent in many data-intensive problem-solving tasks, even experts become overly-specialized in a restricted region of the problem space and fail to develop techniques that apply outside of this region. This results in the troubling phenomenon of *strategic bias*, which leads problem solvers down inappropriate paths, frequently causing premature abandonment of a problem and, ultimately, failure.

We propose *anti-bias stereotypes* as a mechanism to counteract strategic bias. By exploiting these stereotypes, an intelligent consultant can recognize situations in which the biases may negatively impact users' problem solving, and recommend alternative strategies that would likely have gone unconsidered. We have designed and implemented a prototype anti-bias stereotype modeling system in the context of the GENOME COLLABORATOR, an intelligent advisory system that assists molecular biologists in the Human Genome Project. Initial empirical studies conducted to assess the effectiveness of anti-bias stereotyping are encouraging.

Introduction

Providing an automated consultant to humans performing complex tasks has been a long-term goal of the AI community. With the promise of substantial theoretical and applied results, the field has witnessed a growing body of research in task-oriented dialogue systems (Carberry 1989; Quilici 1989; McCoy 1989 1990; Zukerman 1992), explanation systems (Moore & Paris 1993; Suthers 1993; Lester 1994), and and critiquing systems (Fischer *et al.* 1991). Automated consultants can play a particularly important role in *hyper-complex* problem-solving tasks.

Hyper-complex problems are common in data-intensive scientific domains such as molecular biology, where sophisticated data collection and intricate analyses are prerequisites to success. In a typical problem-solving session, a scientist will (1) analyze the data currently available, (2) hypothesize an action that will bring him or her closer to the overall goal, (3) perform the action, and (4) analyze the results. The scientist then iterates until either the goal is satisfied, all possible paths are exhausted without satisfying the goal, or the goal is abandoned. Moreover, the task is so immense that achieving (or failing to achieve) it may require several days or even weeks.

Because of the enormous complexity inherent in these problems, even the most advanced human problem solvers are overwhelmed. By necessity, they become overly-specialized in a restricted region of the problem space and fail to develop techniques that apply outside of this region. This results in the troubling phenomenon of

> *Strategic Bias*: The tendency of a problem solver to prefer particular, idiosyncratic problem-solving strategies over others.

Although strategic biases may be beneficial in dealing with problems confined to the familiar region of the problem space, they act as a serious impediment to solving problems which lie outside of the region. When confronted with unfamiliar situations, problem solvers rely on techniques that have proven to be successful in the past, even though alternative strategies that are less familiar would significantly increase their chances of success. In short, strategic bias results from domain and task complexity, and it leads problem solvers down paths which are inappropriate, frequently causing premature abandonment of a problem and, ultimately, failure.

Anti-bias Stereotypes

To counteract strategic bias, we propose the mechanism of *anti-bias stereotypes*, which represent aggregations of strategic biases that co-occur in particular segments of the user population. By exploiting these stereotypes, an intelligent consultant for a hyper-complex task can predict the biases affecting users' behavior, recognize situations in which the biases may negatively impact users' problem solving, and recommend alternative strategies that would likely have gone unconsidered. Users that come to understand the perspectives and strategies of other users inhabiting different localities of expertise benefit from a broader view of the problem space. By encouraging a user from user class X to develop empathy with users of class Y whose perspectives are different from their own, users from class X can become more effective problem solvers.

*Support for this work is provided by the North Carolina Biotechnology Center under grant #9513-ARG-0047 and by Glaxo Wellcome, Inc.

Like traditional stereotypes (Rich 1979; Chin 1989; Jameson 1992) that encode commonalities of user attributes, anti-bias stereotypes represent aggregations of strategic biases that co-occur in particular segments of the user population. Anti-bias stereotypes bear the following tripartite representation:

- **User Class:** The categories of users which are biased toward a given strategy.
- **Context:** The problem-solving context in which the biased strategy is applicable.
- **Action:** The action which the typical member of the user class would take in this context.

Anti-bias stereotypes can be exploited to help users avoid the pitfalls they encounter when operating outside of their area of expertise by informing them about the strategy biases prevalent in *complementary* areas of expertise. Suppose that a user U_x belonging to class X has opted to perform some action A_x. A consulting system can exploit anit-bias stereotypes in the following way:

1. Determine if U_x is attempting to perform A_x in a known problem-solving context C.
2. If so, determine if there is an action A_y ($A_y \neq A_x$) that would typically be performed by users of class Y when confronted by C.
3. If U_x has not selected A_y, advise U_x that users of class Y would typically perform A_y, and give U_x the option of performing A_y, either instead of or in addition to A_x.

By recognizing critical problem-solving contexts in which the user may go astray and then encouraging him or her to consider how users from complementary classes would solve a given (sub-)problem, an intelligent consultant can counteract bias.

Anti-Bias Stereotypes in Advisory Systems

We are developing an anti-bias stereotype modeling system for hyper-complex problem-solving tasks in the context of the GENOME COLLABORATOR (Figure 1), an intelligent advisory system that assists molecular biologists working on the Human Genome Project (Watson 1986). The GENOME COLLABORATOR is being designed to advise scientists about biological databases and bioinformatics tools by making recommendations about analysis strategies, offering advice about the significance and interpretation of database searches, dynamically planning the presentation of search and analysis results by automatically invoking the most appropriate visualization tools, and helping scientists track the myriad of sub-goals in hyper-complex tasks.

The GENOME COLLABORATOR's stereotypes represents the strategic biases of the two principal communities of molecular biologists, DNA specialists and protein specialists. Scientists in each of these classes bring very different perspectives to bear on a sequence analysis problem than do their colleagues in the other class. When the GENOME COLLABORATOR is invoked, its first action is to ask the scientist to classify himself or herself as belonging to one of the two categories. The tripartite representation of anti-bias stereotypes in the GENOME COLLABORATOR encodes

- **User Class:** A scientific "sub-community" of molecular biologists, e.g., DNA specialists or protein specialists.
- **Context:** A state of a biological sequence analysis session that may be either partially or fully instantiated.
- **Action:** A particular type of information retrieval or analysis activity.

A tight coupling between user actions and problem-solving strategic elements facilitates (1) the representation of anti-bias stereotype models, (2) the recognition (at runtime) of user strategies, and (3) the recommendation (at runtime) of alternate strategies. Because the GENOME COLLABORATOR's domain model encodes user actions and strategic biases in the same ontology—in fact they are encoded in the same knowledge base *objects*—knowledge engineers may easily attach biases for each user class to any concept in the knowledge base. It then becomes a straightforward problem to recognize problem-solving contexts of interest and to suggest alternative courses of action that are applicable in those contexts.

Suppose a scientist is on a fact-finding mission to gather information about a particular piece of DNA whose nucleotide sequence is stored in a local file. When she invokes the GENOME COLLABORATOR, she is presented with an introductory control panel. Its purpose is to collect data about the user and act as a launchpad for the system's functionalities. Users first rate their experience in molecular biology, computer usage, and the GENOME COLLABORATOR; then they classify themselves as DNA or protein specialists.

The system encodes analysis strategies on the "DNA Sequence" concept's `DNA bias` and `Prot bias` slots. For example, DNA specialists are biased toward performing a nucleotide database similarity search by querying the similar-objects slot. This bias was first revealed by our protocol studies with DNA specialists. The `Prot bias` slot contains the actions a protein specialist is likely to take in this context: translating the sequence into a set of polypeptides and then performing a similarity search against protein databases.

As the scientist follows particular paths of analysis, the GENOME COLLABORATOR notes the problem-solving context and determines if one of the current actions being considered is associated with the stereotype of another user class. If so, and if the user-selected action is different from the course of action that a member of the complementary class would take, then it suggests that the user consider an alternative action. To present alternative strategies in a manner useful to the user, the information explorer employs simple explanation templates. The phrases that are used to instantiate template variables are attached to each of the slots associated with a bias, and the information explorer pieces this text together to form an understandable explanation of the strategy under consideration.

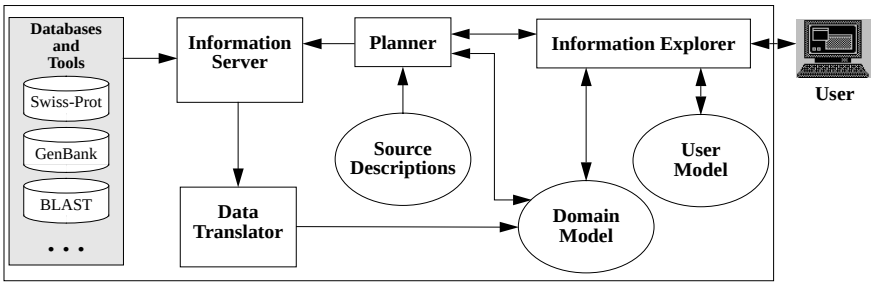

Figure 1: Genome Collaborator Architecture

Assessing the Utility of Anti-Bias Stereotypes

To obtain a preliminary assessment of the utility of anti-bias stereotyping, we conducted a user study with two molecular biologists. The biologists were presented with a hypothetical sequence analysis scenario in which a DNA specialist had obtained an anonymous DNA sequence in his laboratory. Together, the scientists and the authors proceeded through a series of steps required to perform a sequence similarity search. The scientists appreciated the notion of anti-bias stereotypes, and they opined that anti-bias recommendations will be of great assistance to members of the DNA and protein communities, both of which exhibit idiosyncratic analysis biases. An interesting controversy arose during discussion of how the system should gather the user profile at the beginning of the session. One scientist believed that the straightforward presentation of the current system was best because molecular biologists take great pride in their specialty and identify strongly with that group; the other believed that a better way of getting at the same profile data was to ask the user which speciality about which he or she would most like to become better informed.

Conclusion

Even experts suffer from strategic bias when confronted with hyper-complex, problem-solving tasks. A promising mechanism for counteracting strategic bias is the anti-bias stereotype, which represents an aggregation of strategic biases that co-occur in particular segments of the user population. By endowing an intelligent advisory system with anti-bias stereotypes, it is possible to diminish the deleterious effects of strategic bias. Initial studies with an advisory system that has been equipped with anti-bias stereotypes suggest that users' problem-solving effectiveness may be increased by the system's ability to recognize problematic situations and recommend alternative strategies that might have gone unconsidered.

Acknowledgements

The authors would like to thank: Rainer Fuchs, for insightful discussions on the design of the GENOME COLLABORATOR; Mike Agostino and Paul Furdon, for excellent feedback on the software; all of the scientists who participated in the protocol studies; and Charles Callaway for comments on an earlier draft of this paper.

References

Carberry, S. 1989. Plan recognition and its use in understanding dialog. In Kobsa, A., and Wahlster, W., eds., *User Models in Dialog Systems*. Berlin: Springer-Verlag. 133–162.

Chin, D. N. 1989. KNOME: Modeling what the user knows in UC. In Kobsa, A., and Wahlster, W., eds., *User Models in Dialog Systems*. Berlin: Springer-Verlag. 74–107.

Fischer, G.; Lemke, A. C.; Mastaglio, T.; and Morch, A. I. 1991. The role of critiquing in cooperative problem solving. *ACM Transactions on Information Systems* 9(3):123–151.

Jameson, A. 1992. Generalizing the double-stereotype approach: A psychological perspective. In *UM92: Third International Workshop on User Modeling.*, 69–83.

Lester, J. C. 1994. *Generating Natural Language Explanations from Large-Scale Knowledge Bases*. Ph.D. Dissertation, The University of Texas at Austin, Austin, TX.

McCoy, K. F. 1989 1990. Generating context-sensitive responses to object-related misconceptions. *Artificial Intelligence* 41:157–195.

Moore, J. D., and Paris, C. L. 1993. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics* 19(4):651–694.

Quilici, A. 1989. Detecting and responding to plan-oriented misconceptions. In Kobsa, A., and Wahlster, W., eds., *User Models in Dialog Systems*. Berlin: Springer-Verlag. 108–132.

Rich, E. 1979. User modeling via stereotypes. *Cognitive Science* 3:329–354.

Suthers, D. D. 1993. *An Analysis of Explanation and Its Implications for the Design of Explanation Planners*. Ph.D. Dissertation, University of MA.

Watson, J. 1986. The human genome project: Past, present, and future. *Science* 231:1055–1056.

Zukerman, I. 1992. What you infer might hurt you—a guiding principle for a discourse planner. In *Proceedings of the Fourteenth Cognitive Science Society Conference*, 201–206.