

Challenge: How IJCAI 1999 can Prove the Value of AI by Using AI

James Geller

Department of Computer and Information Sciences
New Jersey Institute of Technology
Newark, NJ 07102
U. S. A.

Abstract

The challenge formulated in this paper is a direct reaction to the reviewer selection process of IJCAI 1997, which relied on an unintelligent string matching program. We challenge the AI community to implement an intelligent knowledge-based reviewer selection program for IJCAI 1999.

1 Introduction

In this paper we will formulate a challenge that should result in the design and implementation of a reviewer selection program to be used by the organizing committee of IJCAI 1999. The motivation behind this problem is partially rooted in a dissatisfaction with the assignment process of papers to reviewers that was apparently used for IJCAI 1997. The challenge is that the process of assigning papers to reviewers should be automated and performed by a flexible and intelligent AI program. We will explain this motivation in more detail, based on a metaphor from the business of telephone system development. Then we will outline some of the major problems that are part of this challenge. Finally, we will recapitulate the importance of this challenge problem. Indeed, in our opinion, this challenge problem can well compete with any other challenge problem that is likely to be presented at IJCAI 1997.

2 Motivation

It is helpful to recount very briefly a personal experience of this author to create a clear picture of the motivation for this challenge problem. After graduating from college, this author looked for his first entry level job in software development. At the time and place where this occurred, the only game in town were the phone companies. There were four of them, let's call them AB, CD, EF, and YZ. The author was eventually hired by the AB phone company.

One of the main activities of all four companies was to build large PBX phone systems. These systems would connect to a number of outside lines and make them available to 300 or more local extensions. Such PBX systems would be bought by businesses in town. In fact, this was the accepted way businesses supplied their employees with a connection to the outside world. Naturally, there was fierce competition among the four phone companies to provide cheap and feature-rich PBX systems.

However, the AB company had a special secret. Something they would not have wanted their customers to know. The AB company used a PBX system made by the YZ company for their own in-house communication needs! This system was eventually replaced by an AB system, but still, there was this period when one could have imagined a TV advertisement of the YZ company, sounding like this: "YZ PBX systems – they are so good, even our competition is using them."

Recently, when writing a(nother) paper for IJCAI 1997, the author was reminded of this incident. After having finished the bulk of the paper, it was time to check the list of Content Areas contained in the Call for Papers. (Content Areas are more commonly called Keywords and denote the subareas of AI into which a given paper can be classified.) There were two obvious Content Areas, "semantic networks" and "parallel AI," describing our paper. Thus, the author consulted the IJCAI 1997 Call for Papers list of Content Areas in search of them.

However, both those areas were not contained in the official list of Content Areas. Obviously, "semantic networks" has been a main stream area in AI almost since the beginning of the field [Quillian, 1968]. Parallel AI is less common, but it has been a field of investigation for many years; see e.g., [Geller, 1991; Kanal *et al.*, 1994; Geller *et al.*, 1997].

A quick inquiry with the authorities of IJCAI confirmed that no new Content Areas could be added to the Call for Papers of IJCAI 1997 at such a late date.

The reader might think that this should cause no problems. Unfortunately, a preamble to the list of Content Areas stated:

Please use the exact phrases below; do not make up new keywords, since they will be ignored by the paper classification software.

What conclusions do we have to draw from this statement? First of all, papers will be assigned (partially?) to reviewers by a program. Secondly, this program is so unintelligent that it cannot even deal with variant spellings, much less Content Areas that are not known to it. What else could the admonishment to use "exact phrases" mean?

This sounds like a phone company that does not have a phone system itself.

3 Current Reviewer Assignment

After acceptance of this challenge paper, the author contacted the IJCAI 1997 program chair to find out more about the reviewer assignment program that had been used. The program was developed by Ramesh Patil at USC/ISI¹ for use by the National Conference on Artificial Intelligence, commonly known as "AAAI." Due to the different committee setup at IJCAI and AAAI, a considerable amount of human post-processing was necessary for IJCAI 1997 reviewer assignment [Pollack, 1997].

An inquiry at ISI revealed that the AAAI reviewer assignment program has been used for five years. It relies on a system of bidding for papers, in combination with stored profiles of expertise. Details of this program have not been published, as the author considers them "not algorithmically novel." However, future versions of the reviewer assignment program might be published [Patil, 1997].

4 The Challenge Problem

The following is a suggestion of what should appear in the Call for Papers of IJCAI 1999:

Authors must supply a list of Content Areas from any subarea of Artificial Intelligence that they consider appropriate. These Content Areas will be used by a flexible and intelligent program to assign papers to reviewers. Only in cases where this program reports insufficient confidence in its own results, will human editors make the assignment of papers to reviewers.

Actions speak louder than words. Having the major world-wide conference on Artificial Intelligence in 1997 issue a Call for Papers that admits that it relies on

¹University of Southern California - Information Systems Institute

exact string matching for reviewer preselection speaks very loudly about the state-of-the-art of Artificial Intelligence. We think that this is a statement that we as members of the AI community do not want to make.

What would be involved in the kind of processing required for an intelligent reviewer assignment program (which we will call REVIEWER-SELECT)? We need at least a Natural Language Processing system that is able to parse noun phrases. We need a knowledge base that associates in a flexible way content areas with reviewers. We need a second Natural Language Processing system that is powerful enough to extract possible content areas from given source text, as will be explained below. Finally, we need an expert system that selects reviewers, keeping in mind certain overall constraints which will be elaborated in more detail shortly.

To summarize, the challenge for IJCAI 1999 one more time succinctly:

Papers should be assigned to reviewers by an intelligent program that does not impose any limitations, except that the Content Areas supplied by paper authors need to be cogent to the eyes of a human expert.

5 Sources of Information

In order to create the same base-line conditions for all the teams that want to take on this challenge, the IJCAI organization is asked to do the following. Due to the electronic submission process of abstracts through the world-wide web, the paper abstracts of most IJCAI 1997 submissions will be available on-line. The relatively few abstracts that are not submitted electronically should be typed in by IJCAI volunteers. This set of text sources should be made available through the world-wide web to the world AI community immediately after the IJCAI 1997 conference. The (second) Natural Language Processing system mentioned above should be used by research teams to extract candidate Content Areas from the stored abstracts. Some human filtering and/or augmentation might still be necessary to create a comprehensive list of Content Areas.

An important issue to address in preparing this source of information is how to treat the abstracts of rejected IJCAI 1997 papers. Copyright reasons as well as the large number of rejected papers make it difficult to deal with them as part of the source data. On the other hand, omitting the abstracts of rejected papers might bias the reviewer selection process towards papers in the areas that were accepted at IJCAI 1997. A solution to this problem needs to be found and publicized quickly, to create an even playing field for all future paper submitters.

Secondly, after IJCAI 1997, authors of papers should be asked to supply a list of five to ten individuals that they would have considered good reviewers for their own

papers. Authors might also be asked to supply all the references cited in their own 1997 papers in electronic form to the IJCAI organization, adding an additional source of knowledge for the selection of reviewers. Another source of information is already there, and available to anybody with a browser, and that is the world-wide web itself. How the web could be used will be hinted at below.

One additional point should be made, although it is hopefully obvious at this stage. A response to this challenge will not consist of a Call for Papers containing a 10 page list of Content Areas. Rather, there might be no Content Areas listed at all. Anybody who is capable of submitting a paper to IJCAI should be aware of what constitutes an acceptable Content Area of AI. On the other hand, it appears useful to make the knowledge base of REVIEWER-SELECT itself publicly accessible at some point in time. Whether this should be before or after the IJCAI 1999 dead-line is open to discussion.

6 Problems

The problems posed by designing and implementing the REVIEWER-SELECT program are definitely non-trivial. This includes all aspects of it, but we want to stress the overall control that has to be exercised by this program. For instance, the REVIEWER-SELECT program should not come back with

1. Reviewers with the same affiliation as the author (conflict of interest);
2. The author of a paper as a suggested reviewer (extreme conflict of interest);
3. The doctoral advisor of the author as a suggested reviewer (another extreme conflict of interest);
4. Reviewers that the author has recently written a paper with (potential conflict of interest);
5. Assignment of fifty papers to one reviewer, and one paper each to fifty other reviewers (uneven load for reviewers);
6. Assignment of thirty reviewers to one paper, and no reviewers to thirty other papers (uneven load for papers).

The problem of identifying conflicts of interest could be attacked by using the world-wide web. As many institutions now have home pages on the world-wide web, information about affiliations could be retrieved by appropriate agents and softbots. Even potential conflicts of interest could be detected when authors have their vitae available on-line.

7 Limitations and Success Measurements

In order to make this challenge self-contained and its solution feasible in the allotted time, we must allow for

some Content Areas that will not be contained in the knowledge base of REVIEWER-SELECT. In such cases the system will need to ask a human editor for help. It will be a measure of the success of any program taking on this challenge, how many (or how few) requests for human intervention will be required in processing all IJCAI 1999 submissions.

To ensure both fairness to IJCAI 1999 submissions and an alternative evaluation process of several programs that might take on this challenge (assuming that there will be several) the reviewer assignments should be rated blindly by a committee of human experts. By computing precision and recall measures compared to the ratings of human experts, a numeric evaluation of different programs will be performed.

A second, more expensive, way to evaluate the success of a REVIEWER-SELECT program would be to perform the reviewer selection process exactly as it was done in 1997, and to use the results of this process as the gold standard. This appears attractive for IJCAI 1999 because confidence in a completely automated review process might still be low in the AI community at large. If the evaluation is successful, then IJCAI 2001 might opt for completely automatic reviewer selection.

A third way to assess the success of a REVIEWER-SELECT program would be to compare its results with the names of reviewers suggested by the authors of IJCAI 1999 papers.² If this route is chosen, then the IJCAI 1999 Call for Papers must state that authors need to propose reviewers for their papers at submission time.

Whatever choice is made for evaluation, two conditions must be maintained. (1) If there are several teams taking on this challenge, then all of them must get the same starting conditions, including access to relevant materials collected by AAAI. (2) It must be assured that no REVIEWER-SELECT program has a built-in bias towards papers of its own designers. The author of this challenge paper will gladly serve as a coordinator of such evaluation activities throughout the whole life-time of this effort.

8 Acknowledgment of Debt to Previous Research

The idea of doing "reviewer selection driven research" is not new. Gio Wiederhold, a leading researcher in databases and other fields, and previously the editor of the most prestigious database journal, has mentioned the problem of automatically finding reviewers for his journal in conference presentations [Wiederhold, 1992]. However, we feel that in our formulation as a challenge problem, the difficulties that will be encountered are by

²This idea is due to one of the anonymous reviewers of this paper.

an order of magnitude larger. On the other hand, this challenge problem appears still small enough and well defined enough to lead to real practical results within the given time frame.

9 Why this Problem is Important

We have no doubt that there are many other challenge problems that make sense at the current state-of-the-art. Kitano's Grand Challenge problems have not been solved [Kitano *et al.*, 1993]. Those problems are, however, in a different league, as opposed to the relatively small and well defined problems sought in this year's Challenge Paper track. The Turing Test, in whichever guise, is still the grand challenge problem to all of AI, but there is little hope to see it solved by IJCAI 1999.

Among problems in the same league as the reviewer assignment problem, we feel that our challenge is very important for a simple reason. It is the same reason why a phone company should first build a phone system for in-house use, before starting to sell it to other people. It looks very bad when an enterprise does not produce products that are good enough for its own needs. If IJCAI 1999 will have to rely on string matching for assigning reviewers to papers, it might be better to return to manual assignments, and to never even mention the use of a program in the Call for Papers. On the other hand, any solution to this challenge problem would considerably improve the credibility of AI as a whole.

10 The Real Challenge – A Final Remark

The challenge formulated in this paper is very limited, compared to the real challenge of making AI programs superior to average human practitioners in a field. From this point of view, why are authors of papers asked to supply Content Areas at all? Why can't an intelligent program "just read" the abstract of a paper and decide what the Content Areas are? A program that can assign papers to reviewers without being given Content Areas and that can do this for the whole range of AI problems would be a supreme achievement, as very few human individuals can perform this task. Such a program would probably build on many current subareas of AI, including data mining, machine learning, advanced natural language processing, knowledge representation, reasoning, etc. Research on summarizing text might be a valuable starting point. Maybe we will be able to pose this task as a challenge problem for IJCAI 2009.

Acknowledgment

Mike Halper has improved the English of this paper.

References

- [Geller *et al.*, 1997] J. Geller, H. Kitano, and C.B. Suttner, editors. *Parallel Processing for Artificial Intelligence 3*. North-Holland Elsevier, Amsterdam, 1997.
- [Geller, 1991] J. Geller. Upward-inductive inheritance and constant time downward inheritance in massively parallel knowledge representation. In *Proceedings of the Workshop on Parallel Processing for AI at IJCAI 1991*, pages 63–68, Sydney, Australia, 1991.
- [Kanal *et al.*, 1994] L. Kanal, V. Kumar, H. Kitano, and C. Suttner, editors. *Parallel Processing for Artificial Intelligence*. Elsevier Science Publishers, Amsterdam, 1994.
- [Kitano *et al.*, 1993] H. Kitano, L. Hunter, B. Wah, W. von Hahn, R. Oka, and T. Yokoi. Grand challenge AI applications. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 1677–1683. Morgan Kaufmann, San Mateo, CA, 1993.
- [Patil, 1997] R. Patil. Personal communication, 1997.
- [Pollack, 1997] M. Pollack. Personal communication, 1997.
- [Quillian, 1968] M. R. Quillian. Semantic memory. In M. L. Minsky, editor, *Semantic Information Processing*, pages 227–270. The MIT Press, Cambridge, MA, 1968.
- [Wiederhold, 1992] G. Wiederhold. Model-free optimization. In *DARPA Software Technology Conference*, pages 83–96. Meridian Corp., Arlington, VA, 1992.