Computerized Adaptive Student Help: Experiments in Low-Cost Artificially Intelligent Automation of Student-Tutor Interaction

HUEY LING TOH and R.C. LACHER Department of Computer Science Florida State University 253 Love Building, Florida State University, Tallahassee FL, 32306-4530 UNITED STATES OF AMERICA toh@cs.fsu.edu, lacher@cs.fsu.edu http://www.cs.fsu.edu/~lacher/

Abstract: - The lessons attained from research in artificially intelligent learning systems demonstrate the complexity in their attempts to learn either broad categories of knowledge or explicit tasks and skills that require extensive use of resources. This paper describes an investigation into some energy effective and low-cost approaches to developing lightweight artificially intelligent learning systems that are specific to a knowledge domain, and yet, are easily portable and expandable to incorporate larger domains of knowledge. A prototype system was built to test the feasibility of such an approach. The resulting observations provide evidence that it is indeed feasible to create a low-cost system that is highly energy effective for both end-users and trainers with the use of relatively simple processing algorithms. Furthermore, the investigation has yielded several promising opportunities for improvements to the current prototype system.

Key-Words: - artificially intelligent learning systems, web-based education, machine learning.

1 Introduction

Computational models of human learning are typically designed to mimic human learning, processing, and querying activities. Often, the focus of these systems is to represent information in a meaningful context, and store, retrieve and effectively communicate the learned knowledge in a human environment.

In an effort to create artificially intelligent learning systems, researchers have turned to the workings of the most sophisticated form of learning system known to mankind; the human brain.

While some may argue that certain forms of human learning could occur in the absence of brain activity, this research is concerned mainly with learning that involves the human brain. In this context, we are interested in learning activities that occur by engaging the brain, the nervous system, the environment and the process in which the collaboration of all these factors is supported by information acquisition and skills.

An enormous amount of information from the environment is gathered through the five senses. Human learning and communication is facilitated by the filtering of information gathered through the five sensory inputs of the human body [1]. What changes the learning process when one of the sensory inputs is absent? An analogy for the assertion is discussed in the following situation. Given the absence of background knowledge of physical geometry, normally acquired through the sense of sight, the learner will not be able to understand the phrase, "*The fence ran along the side of the house*". Logically, the *fence*, being an object without wheels or legs is incapable of moving. However, if the learner had prior knowledge of physical geometry, she would understand the metaphorical implications of the phrase.

We contend that the limitations in artificially intelligent learning systems stem from the medium of representation of the processed information: the human language. In the context of the English language, the ambiguous nature of the language adds to the complexity of knowledge representation. In fact, it has been argued that all forms of human thought are derived from our physical experiences in the world we live in [2]. Therefore, in order for learned information to be expressed in a meaningful context that is understood, all five sensors should ideally be present.

Much research has been conducted to improve artificial sensors. Many artificially intelligent systems rely on these synthetic inputs to process information from their surroundings. Despite the achievements made to date, researchers acknowledge that there is still a long way to go before each artificial sensory input can be honed to match those that are bestowed on humans. Limitations in current technology make it economically unfeasible to attempt to create artificial systems that are as intelligent as the human brain.

What we do know, however, is that artificially intelligent learning systems have their own set of sensory inputs. Currently, humans have been able to effectively communicate with computers through input devices such as the keyboard and the mouse. Since the computer has such dissimilar and restricted sensory inputs, it is incapable of processing information in a manner similar to the human learner. This handicap results in the narrowed domain of knowledge that can be identified by the system.

The research described herein, limits the domain to one where knowledge can be completely conveyed through straightforward keyboard, mouse and/or monitor input/output devices. One of the goals of this investigation is to evaluate the practicality of creating an artificially intelligent learning system that is specific to a knowledge domain and yet, is easily portable and expandable to incorporate larger domains of knowledge.

2 Background

The following is a study of prior researches conducted, which we are considered to be within the same problem domain.

2.1 Natural Language Processing Systems

Natural Language Processing (NLP) systems use natural language as input and output for the system [3]. An impressive example of such a system is the START (SynTactic Analysis using Reversible Transformation) natural language system [4].

The user queries the database in English Language. If an answer is found, the system will produce a response in the English language and deliver the answer. The knowledge base of the system expands through what is known as a "virtual collaboration"; a technique that locates websites with interesting databases and identifies the sites' directory structure and argument conventions. Then, the system creates rules based on the English question, cites the appropriate uniform-resource locators and eventually presents the information to the user.

In spite of its many advances, the system is still faced with the difficulty of dealing with the unstructured nature of information available on the World Wide Web (WWW). At present, there is no known reasoning technology or method of representation that is capable of ordering the information. In addition, the exponential growth of the WWW amplifies the complexity of the task.

2.2 Intelligent Tutoring Systems

Intelligent tutoring systems (ITSs) are systems that are aimed at providing individualized instruction to the student [5]. An ITS works by helping students learn through the problem-solving activities incorporated into the system.

An example of an ITS is ELM-ART II, an intelligent interactive textbook that helps students learn programming in LISP [6]. ELM-ART II's web-based implementation brought accessibility of ITS outside of the classroom, especially in distance-learning environments, where the instructor is not physically available.

Among the more prominent hurdles of ITSs are the portability and scalability of such systems. Since each ITS is tailored for a specific course, it often requires a tight and constant collaboration between the instructor and the design team. The total man hours in creating a complete ITS for a course may exceed the time a faculty is willing to spend on the development process. In most instances, the cost far outweighs the benefits.

2.3 Search engines

Search engines are made up of large databases of web page files. Although the search process remains similar in general, most search engines distinguish themselves from one another in their methods of implementation, such as ranking schemes, size of searchable domains, speed of search, analysis of Web content, and search options.

The popular Google search engine utilizes a search architecture that is supported by two sets of technology, known as *PageRank* and *Hypertext-Matching Analysis* [7]. *PageRank* allows Google to collect votes based on a collective input from the web sites throughout its search domain. *Hypertext-Matching Analysis* allows Google to analyzes a web site's content based on fonts, sub-divisions, and the exact location of all terms on a page and neighboring pages.

Another interesting example worth noting is the *Ask Jeeves* search engine. *Ask Jeeves* uses NLP technology to search through unstructured web content of up to 225 unique document types [8], [9].

The drawback lies in the high volume of indexed web sites, which causes hundred of irrelevant results



Fig. 1. Combinational Matching Algorithm

to be presented for a simple search request. Another shortcoming occurs when the keyword searched appears only once in a lengthy document.

2.4 Observation

Researchers of NLP, ITS and search engines have demonstrated that technology can be utilized to revolutionize the way humans communicate and transfer knowledge. However, there is still much debate as to whether the technology requires more effort that it is worth [10]. The fact remains that the process of automating query and answer sessions requires significant commitment in time and effort and the success of implementing such systems depends largely on the enterprise's willingness to allocate the necessary manpower and resources to prepare and maintain them.

3 Computerized Adaptive Student Help (CASH)

The challenges that were unique to NLP, ITS and search engine systems were examined and incorporated into the development of this experiment. prototype named А system Computerized Adaptive Student Help (CASH) was created to implement the concepts and ideas set forth in this research.

The current prototype system provides the following features:

Distribution of expertise. The system emulates the knowledge and skills possessed by the experts/trainers of the system.

Question-based processing. The answer is found only when the question if recognized. Unlike

START and search engines that looks for matching keywords in the answer, CASH will first determine if the query contains keywords that exist in the knowledge base. The query-to-answer matching process is achieved through the *combinational matching algorithm* illustrated in Fig. 1.

Narrow and synthetic knowledge domain. In order to effectively capture the semantics of a user query, the knowledge domain is confined within the perimeters of the specific topic. In addition, the knowledge domain needs to be synthetic to remove possible interaction with world knowledge. This would alleviate the complexity of requiring the system to maintain a set of knowledge for information not directly related to the topic.

Uninterrupted accessibility. The system, being deployed over the Internet, is able to provide its services twenty-four hours a day, seven days a week.

Ease of adaptation. The system can be easily updated to adapt to changes in the knowledge base. Unlike most ITS systems, CASH begins with an empty knowledge base and acquires knowledge through use. This feature overcomes the limitations of ITS systems, which require large number of manhours and tedious effort on the path of both development team and domain expert to prepare a system for deployment.

Lightweight implementation. CASH can be easily trained through a web-based interface. Once the system is setup correctly, the trainer will be able to update and add to the knowledge base without the intervention of the development team. Updates occur through amendments to existing date and new knowledge imparted by the trainer. As learning occurs dynamically in real time, the trainer need not spend a large number of hours preparing the system prior to its interaction with its users.

3.1 Overview of the Cash Prototype

In essence, the CASH system interacts with the world through the following entities: the *user*, the *trainer*, the *processing engine* and the *knowledge base*.

The user is the student who submits a query. The query is delivered to the processing engine for further processing. The trainer is the instructor who adds and/or updates the knowledge stored within the knowledge base. The processing engine provides responses to the user's query and performs algorithms to facilitate the learning and training processes executed by the trainer. The knowledge base stores the knowledge imparted by the trainer.

3.2 The Processing Engine

The CASH prototype serves four key processes, namely, the *query*, *search*, *training* and *learning* processes.

3.2.1 The Query Process

When a user submits a query, the processing engine parses the query into separate words. Each word is matched against a set of keywords in the knowledge base and the answer is found when the keywordcode is found in the query-to-answer mapping table. The process is implemented with *combinational matching algorithm* illustrated in Fig. 1.

If an answer is found, the system is said to have found an existing question in the knowledge base and proceeds to retrieve the answer and display the question and answer set to the user. The resulting screen display is shown in Fig. 2.



Fig. 2. Answer to user's question found

If an answer if not found, the system will supply the user with a list of queries that is the closest possible match to the original query. An example of this scenario is shown is Fig. 5.

If none of the suggested possible matches presented to the user are helpful, the query will be

tagged as a *pending question* and forwarded to the trainer to be answered during the training process.

3.2.2 The Search Process

The search process is executed when the user performs a search on the available question-answer sets on the system. A screen display of the search page is shown in Fig. 3.

Hello, Ling. Please type in your keyword(s).				
Show all keywords				
KEYWORDS (separate keywords with commas):				
Match preference : Any word C All words Submit •Clear				

Fig. 3. Search page

3.2.3 The Training Process

The training process occurs at the Train page. Only the trainer of the system is allowed access to this utility. The trainer is presented with a list of unanswered questions.

The trainer may select the question she wishes to address. She will be presented with a training form as shown in Fig. 4.

The keywords for each question must be unique, however, the answer to each question need not be exclusive. This means that there must be one and only one set of keyword combination for a question, although it is possible to have more than one question that will lead to the same answer.

When a pending question has been answered by the trainer, the system will notify the user in the form of an email, detailing the original date and content of the query and the instruction to obtain the answer from the CASH system.



Fig. 4. Training form

Select	Question	Keywords	Date Modified	
©	How do you initialize an array?	how, initialize, array	Thu Jul 25 02:42:28 2002	
0	what is an array	what, array	Thu Jul 25 02:45:37 2002	
View record New search				

Fig. 5 List of possible query-and-answer set.

3.2.4 The Learning Process

The learning process is implemented to facilitate the artificial learning activity of the CASH system.

As mentioned earlier in the *query* process, the user is presented with a list of queries is the closest match to the original query. It is possible that an alternative match holds the same meaning as the user query.

At this point, the system keeps track of the number of such instances where a user query that is initially unrecognized by the system, is later accepted by the user as a correct match to another query in the knowledge base.

An unrecognized user query presented to the system that contains a match to an existing query in the knowledge base is recorded in a table, known as the *intermediate query* table. In the intermediate query table, the query is initially awarded the *strength* of one. The strength characteristic of the user query is used to indicate the popularity of the query. Each time that the same user query is presented to the system and the user accepts the identical matching knowledge base query; the strength of the user query is increased by one.

The strength is incremented until it reaches a threshold. The threshold is set during the initial setup process of the system. When the strength value reaches a threshold, the system would then contact the trainer through e-mail and notify her that a new query with a matching existing answer has been identified. The trainer should then log into the system and perform the appropriate updates by conducting the training process.

It is important to note that the system was not designed to update itself, even though that is a possible alternative. The decision to leave the update process to the trainer is based on the realization that the exact form of the query may need to be evaluated and rephrased to be precise and valid. In addition, it is also a form of security measure to avoid updating the knowledge base with false or incorrect queries.

4 Conclusions

The prototype system exhibits the following three key attributes:

- i. The system is energy effective for the trainer of the system;
- ii. The system is energy effective for the user of the system; and,
- iii. The system uses a relatively simple matching algorithm (CMA) to interact with the trainer and users of the system. It has been argued that intelligent behavior can be created without having explicit reasoning mechanism [11]. The CMA algorithm forms the basis of the processing engine within the system. Although relatively simple in implementation, it is currently able to respond to a user query by either providing a valid answer or a closest matching question. In addition, the CMA algorithm is used to facilitate the artificial learning process of the system. As such, an observer of the system who has no prior knowledge of the internal workings of the CASH prototype may easily assume that the system is composed of a more complex internal structure.

The resulting CASH system is an energy effective web-based application that employs a simple matching algorithm, to respond to a user's query and learn new knowledge through the input supplied by the trainer. Once the setup procedure is complete, the CASH system requires minimal interaction between the design team and the trainer. Because the processing engine is distinct from the web-based interface, the system is highly portable. In addition, the prospect of being able to build networks of CASH systems within the same domain presents an excellent opportunity to enhance the creation of a highly responsive artificial intelligent and learning system.

4 Future Work

The current CASH implementation is merely a prototype system. It is created with the goal of testing the practicality of using its services and discovering new approaches to the problem of creating artificially intelligent and learning systems.

In addition to helping this research address those issues, the creation of the CASH prototype also presents a new set of more exciting questions. It is the hope of this research that these opportunities can be investigated and implemented in future generations of CASH.

One of the first improvements that can be made to CASH is the enhancement of the parsing method in its Processing Engine. By incorporating a more sophisticated parsing algorithm to determine the grammatical structure of the query, the system would be able to achieve a higher level of effectiveness in the query processing procedure.

Secondly, the current CASH prototype requires that the user and trainer interact with the system through a web-interface. Theoretically, it would be more desirable if the training and querying processes were transparent from the web-based application. Would the human computer interaction be enhanced if the query and training process occur through electronic mail? Despite the advantages, the question remains as to the feasibility of such an approach, that is; will the system be able to interpret the e-mail messages correctly?

Finally, the most appealing prospect of all would be the possibility of creating a network of CASH multi-agent systems that facilitates communication between individual CASH nodes. This would allow a local CASH system to extract answers from other CASH systems to provide valid responses to the user. Therefore, instead of waiting for the trainer, the local CASH system can reactively seek answers to satisfy user queries.

In addition, it is foreseeable that the research can take it one step further and connect CASH networks of multiple domains. The result would be a learning universe where the user can query a local CASH system with questions that may not be within the immediate knowledge domain. Clearly, the most ambitious question of all is whether this goal can be attained and at what cost to the research team and the users of the system.

References:

[1] Soussa, David A., *How the Special Needs Brain Learns*, Corwin Press, Inc., 2001.

- [2] Johnson, Mark, *The Body in the Mind*, University of Chicago Press, 1987.
- [3] Dale, R., Moisl H., and Somers H., *Handbook* of Natural Language Processing, Eastern Hemisphere Distribution, 2000.
- [4] Katz, Boris, From Sentence Processing to Information Access on the World Wide Web [World-Wide Web site], Cambridge, MA: Massachusetts Institute of Technology, 1997: http://www.ai.mit.edu/people/boris/webaccess/ sergei-html.html
- [5] Thomas, Eric, Intelligent Tutoring Systems
 [World-Wide Web site]. San Diego, CA: San Diego State University, 2001; http://coe.sdsu.edu/eet/Articles/tutoringsystem/ start.htm
- [6] Weber G, Specht M, User Modeling and Adaptive Navigation Support in WWW-based Tutoring Systems, Proc. 6th International Conference on User Modeling, 1997. pp.289-300.
- [7] Google, Inc., *Life of a Query*; [World-Wide Web site]. Mountain View, CA: Google Press Center, 2001; http://www.google.com/press/query.ht ml
- [8] Jeeves Solutions, *Natural Language Processing*; [World-Wide Web site]. Emeryville, CA: Technology, 2001; http://www.jeevessolutions.com/techn ology/natural_language.asp
- [9] Weizenbaum, Joseph, ELIZA--A Computer Program For the Study of Natural Language Communication Between Man and Machine, *Communications of the ACM Vol. 9, Issue*, January 1966. pp. 36-45.
- [10] Aberdeen Group, Inc., Advanced Customer Service and Insight Through Automated Question Answering, An Executive White Paper, [World-Wide Web site]. Boston, MA: Search 2001; http://www.aberdeen.com/ab_company/search/ search.htm
- [11] Brooks, Rodney A., Intelligence Without Reason, Proceedings of 12th Int. Joint Conf. On Artificial Intelligence, Sydney, Australia, August 1991, pp. 569-595.