

Integrating Artificial Intelligence into Information Retrieval

Presented By: Babur Nawaz Khan

Introduction

- ▶ The amount of available information is growing at an incredible rate, for example the Internet and World Wide Web.
- ▶ Information stored in many forms e.g. images, text, video, and audio.
- ▶ Information Retrieval, a way to separate relevant data from irrelevant.
- ▶ IR field has developed successful methods to deal effectively with huge amounts of information.
 - ▶ Common methods include the Boolean, Vector Space and Probabilistic models.

Introduction (cont'd)

- ▶ Artificial Intelligence
 - ▶ Study of how to construct intelligent machines & systems that can simulate or extend the development of human intelligence.
- ▶ Both IR and AI fields developed in parallel during the early days of computers.
- ▶ The fields of artificial intelligence and information retrieval share a common interest in developing more capable computer systems.
- ▶ Integration of Artificial Intelligence & Information Retrieval
 - ▶ Development of methods to learn user's information needs.
 - ▶ Extract information based on what has been learned.
 - ▶ Represent the semantics of information.

What is Intelligence?

► According to Cook et.al. [1988]

1. **Acquisition:** the ability to acquire new knowledge.
2. **Automatization:** the ability to refine procedures for dealing with a novel situation into an efficient functional form.
3. **Comprehension:** the ability to know, understand, and deal with novel problems.
4. **Memory management:** the ability to represent knowledge in memory, to map knowledge on to that memory representation, and to access the knowledge in memory.
5. **Metacontrol:** the ability to control various processes in intelligent behavior.

What is Intelligence? (cont'd)

6. ***Numeric ability***: the ability to perform arithmetic operations.
7. ***Reasoning***: the ability to use problem-solving knowledge.
8. ***Social competence***: the ability to interact with and understand other people, machines or programs.
9. ***Verbal perception***: the ability to recognize natural language.
10. ***Visual perception***: the ability to recognize visual images.

What are Intelligent IR Systems?

- ▶ The concept of 'intelligent' information retrieval was first suggested in the late 1970s.
- ▶ Not pursued by IR Community until early 1990s.
- ▶ Definitions
 - ▶ An intelligent IR system can simulate the human thinking process on information processing and intelligence activities to achieve information and knowledge storage, retrieval and reasoning, and to provide intelligence support.
 - ▶ In an Intelligent IR system, the functions of the human intermediary are performed by a program, interacting with the human user. (Belkin et al. '87)
 - ▶ Intelligent IR is performed by a computer program (intelligent agent), which acts on (minimal or no explicit) instructions from a human user, retrieves and presents information to the user without any other interaction. (Maes '94)

How to introduce AI into IR systems?

- ▶ Can't take the "Human Factor" completely out of the equation.
 - ▶ A program which takes a query as input, and returns documents as output, without affording the opportunity for **judgment, modification** and especially **interaction with text**, or with the program, is one which would not qualify as an IR system at all.
 - ▶ Ignores user interaction and relevance feedback.
- ▶ Some processes which can't be performed by any other component than the user.
 - ▶ "Judgment" is a process which can only be performed by the user.

How to introduce AI into IR systems? (cont'd)

- ▶ The question is, “where” should AI be introduced into the IR system?
- ▶ Levels of user and system involvement, according to Bates '90:
 - ▶ Level 0 - No system involvement (User comes up with a tactic, formulating a query, coming up with a strategy and thinking about the outcome)
 - ▶ Level 1 - User can ask for information about searching (System suggests tactics that can be used to formulate queries e.g. help)
 - ▶ Level 2 - User simply enters a query, suggests what needs to be done, and the system executes the query to return results.
 - ▶ Level 3 - First signs of AI. System actually starts suggesting improvements to user.
 - ▶ Level 4 - Full Automation. User queries are entered and the rest is done by the system.

Barriers to Intelligent IR Systems

- ▶ Common Sense Reasoning.
- ▶ Natural Language Processing.
- ▶ Knowledge Acquisition, Representation, and Maintenance.
- ▶ Difficulty in Scaling Up Prototypes to Operational Systems.
- ▶ Level of Effort, Technical Expertise, and Expense.

Some AI methods currently used in Intelligent IR Systems

- ▶ Web Crawlers (for information extraction)
- ▶ Mediator Techniques (for information integration)
- ▶ Ontologies (for intelligent information access by making semantics of information explicit and machine readable)
- ▶ Neural Networks (for document clustering & preprocessing)
 - ▶ Kohonen Neural Networks - Self Organizing maps
 - ▶ Hopfield Networks
 - ▶ Semantic Networks

Neural Networks in IR

- ▶ Based on Neural Networks
 - ▶ Document clustering can be viewed as classification in **document*document** space
 - ▶ Thesaurus construction can be viewed as laying out a coordinate system in the **index*index** space
 - ▶ Indexing itself can be viewed as mappings in the **document*index** space
 - ▶ Searching can be conceptualized as connections and activations in the **index*document** space
- ▶ Applying Neural Networks to Information Retrieval will likely produce information systems that will be able to:
 - ▶ recall memories despite failed individual memory units
 - ▶ modify stored information in response to new inputs from the user
 - ▶ retrieve "nearest neighbor" data when no exact data match exists
 - ▶ associatively recall information despite noise or missing pieces in the input
 - ▶ categorize information by their associative patterns

Conclusion

- ▶ AI offers us a powerful set of tools, especially when they are combined with conventional and other innovative computing tools. However, it is not an easy task to master those tools and employ them skillfully to build truly significant intelligent systems.
- ▶ By recognizing the limitations of modern artificial intelligence techniques, we can establish realistic goals for intelligent information retrieval systems and devise appropriate system development strategies.
- ▶ AI models like the neural network will probably not replace traditional IR approaches anytime soon. However, the application of neural network models can make an IR system more powerful.

Open Problems

- ▶ Which AI techniques are appropriate for which information retrieval problems?
- ▶ How should we compare AI approaches with traditional information retrieval techniques, and what is the best way to evaluate their performance?
- ▶ Is it desirable and productive to look for new AI paradigms for information retrieval, or should research be aimed at finding better symbolic representations of document retrieval activities by people?
- ▶ How to allow users, whether expert or not, using any language, to access information stored in any media, and across distributed sites?

References

- ▶ N.J. Belkin, H.M. Brooks and P.J. Daniels, “Knowledge acquisition using discourse analysis”. *International Journal of Man-Machine Studies*, 27: 127-144, 1987.
- ▶ P. Maes, “Agents that Reduce Work and Information Overload”. *Communications of the ACM*, 37(7): 31-40, 1994.
- ▶ M. Bates, “Where should the person stop and the information search interface start?” *Information Processing and Management*, 26(5): 575-59, 1990.
- ▶ K.S. Jones, “Information retrieval and artificial intelligence”, August 1998.
- ▶ Thomas Mandl, “Artificial Intelligence for Information Retrieval”, 2009.
- ▶ D. Fensel & M. Lalmas “Seeking Information: Methods from Information Retrieval and Artificial Intelligence”, 2000
- ▶ Nicholas J. Belkin, “Intelligent Information Retrieval: Whose Intelligence?”
- ▶ C.W. Bailey, “Intelligent Library Systems: Artificial Intelligence Technology and Library Automation Systems”, 1991.
- ▶ J. Cook, A.D. Whittaker, R.H. Thieme, O.R. Smith, and G. Salvendy, “Human Intelligence Models and Their Implications for Expert System Structure and Research”, 1988.

Questions