Usable AI Requires Commonsense Knowledge

Henry Lieberman
Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
lieber@media.mit.edu

CHI 2008, April 5 – April 10, 2008, Florence, Italy
Workshops and Courses: Usable Artificial Intelligence.

Abstract

Artificial Intelligence techniques are increasingly being applied to the user interface, as evidenced by growing numbers of CHI papers which have some AI aspect, and standalone conferences on the subject, such as the Intelligent User Interfaces (IUI) Conference (and this workshop!). I argue that an important, but underappreciated component for assuring the adherence of AI interfaces to CHI principles for usable interfaces, is capturing Commonsense knowledge. Commonsense knowledge can be viewed as a collection of simple facts about people and everyday life, such as "Things fall down, not up", and "People eat breakfast in the morning". One reason that conventional interfaces are stupid and frustrating to use, is that they lack such knowledge. At the MIT Media Lab, we have a large body of experience in creating applications across a wide variety of domains that make use of such knowledge [6]. We distill from our experience some principles for application of Commonsense knowledge to make interfaces more usable.

Keywords

Commonsense reasoning, Artificial Intelligence, Usability

ACM Classification Keywords

I.2 Artificial Inteliigence, I.2.8 Problem Solving, I.2.4 Knowledge Representation, I.2.7 Natural Language, H.5.2, Interfaces.

Common Sense in Interfaces

Why can't a cell phone know that it shouldn't ring if you're at a concert? In theory, it could. GPS could tell where you are; maps could say that it is a concert hall; schedules on the Web could realize a concert is taking place. A conditional could check for this specific case and initiate that specific action. One reason why this kind of thing is not implemented is that it would have to be programmed across a huge range of potential situations. It would be better to be able to figure it out from general knowledge, such as "People feel interrupted when a phone rings", and "Concerts take place in concert halls". This is what humans call Common Sense. AI pioneers such as John McCarthy, Marvin Minsky and Doug Lenat have conjectured that the central problem of Artificial Intelligence is collecting enough of this Commonsense knowledge and figuring out how to make use of it. We agree.

Projects like Lenat's Cyc and our own Open Mind Common Sense have developed large collections, on the order of a million assertions. A human lives for about 3 billion seconds; assuming someone can learn at most a few facts a second, and that most of this knowledge is acquired by the teens, these databases represent between a few tenths of a percent to a few percent of the total. Put another way, a pocket dictionary contains about 50K words; it is like knowing 10-100 things about each word in the dictionary – substantial, but not exhaustive.

Since about 2003, the Software Agents group at the MIT Media Lab has developed a wide range of interfaces using Commonsense knowledge [1-9, 6]. Areas of application include: predictive typing; speech recognition; natural language understanding; computing intangible qualities such as affect from text; end-user programming; personal information management such as calendars; search; managing photo libraries; storytelling; video editing; financial advice; language and cultural translation; electronic commerce; activity recognition from sensor data; interfaces to networks of consumer electronics devices; online help and education, and many more. Quite a few have been published at past CHIs, referenced below. We are now convinced that Commonsense knowledge is a practical, powerful tool for improving interfaces.

Intelligent Defaults

User interface designers rarely appreciate how *underconstrained* interfaces are. At any given moment, the user interface has an enormous variety of possibilities for how to react to user input. Usually, they just pick one of the possibilities, more or less arbitrarily. Wouldn't it be better if they could examine the situation and propose a *plausible*, if not necessarily correct, alternative? Common Sense can be used, therefore, to supply *intelligent defaults*.

When a user asks to "Open a File", for example, which one? Contemporary systems simply open the last folder used, etc. What if the computer had even a little understanding of what you were working on and could suggest *relevant* files? Or I could say, "Transfer the files I need for my trip to my laptop", and get my travel reservations, maps, talk slides, etc. without having to name the files one by one?

Natural Language Interfaces

Natural Language and especially, speech interfaces are often the easiest to use for non-expert users. Users can

leverage their knowledge about how to communicate with other people, to communicate with a computer.

Natural language interfaces are plagued with the problem of *ambiguity*. Words have different word senses according to how they are used. Speech interfaces have the problem of homonyms, words that sound the same but have different spellings and meanings. How do humans deal with these problems? They disambiguate input based on *context*.

Common Sense, therefore, represents the base level of context that one can assume by default, exclusive of specialized domain knowledge and user-personalized details. We have successfully used Commonsense knowledge for disambiguation in a number of natural language and speech interfaces. For example, in a speech interface, if the user says, "My bike has a squeaky {brake/break}", we go into the Commonsense knowledge base and ask, "What do you know about bikes?", making it easy to choose the right one. [4]

Understanding Goals and Actions in Interfaces

Users come to systems with *goals*: Plan a trip, figure out next year's budget, edit a video of their kids' birthday party [8]. Unfortunately, systems tend not to have goals, only functions: Open a spreadsheet, browse to a particular URL, copy a video from times 1:45 to 2:17 to an Mpeg2 file. We now leave it up to the user to figure out how to express their goals in terms of the system's concrete operations. People often screw this up.

Systems that understand the relationship between goals and actions can help users, providing intelligent context-sensitive help, and debugging facilities. We use AI plan generators, plan recognizers, along with Commonsense reasoning, to a variety of help, educational, and debugging facilities [9]. Commonsense knowledge can also help knit together disparate applications in service of a single user goal, eliding unnecessary steps such as cutting and pasting, loading and saving of files.

Especially in help and education, the *relation between expert knowledge and novice knowledge* is important. Expert knowledge tends to be well codified in the form of user manuals, textbooks, etc. But what about novice knowledge? Common Sense can be used to model "what somebody knows if they don't know anything in particular" and systems can find analogies to help explain expert knowledge in Commonsense language. [3,7]

Knowledge-based and Statistical Heuristics

Currently popular in AI for tasks such as natural language understanding, speech recognition, vision, and other tasks are statistical algorithms such as Bayesian reasoning and Hidden Markov Models. Such algorithms are mathematical ways of combining large numbers of weak bits of evidence to form an overall judgment. They typically generate a list of possible hypotheses for the interpretation of data, often weighted with a confidence factor. Typically, these produce a selection of hypotheses that a human would consider plausible, together with some completely off-the-wall alternatives, due to artifacts of the data or the algorithms. The mathematics can't tell the difference; it must be presented to the human for judgment.

We have had considerable success with using Commonsense knowledge for *filtering* the output of statistical algorithms, eliminating those possibilities that "don't make sense" according to our Commonsense knowledge base. This has the effect of greatly improving the hypothesis selection. For example, a predictive typing system proposes for a continuation of typing "tennis p", "tennis player" where a statistical algorithm predicts "tennis place" [2].

That also emphasizes the importance of *making good mistakes*. Statistical algorithms, when they do make mistakes, tend to make arbitrary mistakes that confuse the user. Commonsense algorithms make *plausible* mistakes, which are much more forgivable by the user. The field of Human-Computer Interaction should put much more research effort into understanding how to have systems make better mistakes, not just to get the right answer.

Fail-Soft Interfaces

The CHI community has been eloquent at pointing out the possible dangers of too much reliance on AI and, by extension, Commonsense interfaces. What if the knowledge is wrong or incomplete? We agree that these concerns are legitimate and should be addressed. But excess caution in the HCI community has led to paralysis; rather than conclude AI can't be used, we should investigate its strengths and weaknesses, and use it where it works best.

In most of our applications, we are careful to pick applications that are not critical; we don't work on nuclear power plants or air traffic control. We design interfaces to be *fail-soft*. We cast the computer in the role of an *agent* interface, where its role is to make suggestions, or to adapt the interface to the current situation, making the most likely thing the easiest to do in the interface. We try to make sure to keep the *human in the loop*, where the user can review and approve or modify system suggestions.

REFERENCES

1. Liu, H., H. Lieberman, T. Selker, Visualizing the Affective Structure of a Text Document. CHI 2003, Ft. Lauderdale, Florida.
2. Stocky, T., A. Faaborg, H. Lieberman, A Commonsense Approach to Predictive Text Entry, CHI 2004, Vienna, Austria, April 2004.
3. Lieberman, H., A. Kumar, S. Sundararajan, Common Sense Investing: Bridging the Gap Between Expert and Novice, CHI 04, Vienna, Austria, April 2004.
4. Chung, J., R. Kern, H. Lieberman, Topic Spotting Common Sense Translation Assistant, CHI 2005, Portland, OR, April 2005.
5. Liu, H., H. Lieberman, Programmatic Semantics for Natural Language Interfaces, CHI 2005, Portland, OR, April 2005.
6. Lieberman, H. Liu, P. Singh, B. Barry, Beating Common Sense into Interac*tive Applications, AI Magazine*, Winter 2004.
7. Lieberman, H., A. Kumar, Providing Expert Advice by Analogy for On-Line Help, IEEE/ACM Conference on Web Intelligence & Intelligent Agent Technology, Compiègne, France, September 2005.
8. Faaborg, A., H. Lieberman, A Goal-Oriented Web Browser, CHI 2006, Montréal, April 2006.
9. Lieberman, H., J. Espinosa, A Goal-Oriented Interface to Consumer Electronics using Planning and Commonsense Reasoning, International Conference on Intelligent User Interfaces (IUI-06), Sydney, Australia, January 2006.