Web Intelligence meets Brain Informatics at the Language Barrier: a Procrustean Bed?

Nick Cercone
Faculty of Science & Engineering
York University – Toronto, Canada
December 2006

YORK
UNIVERSITÉ
UNIVERSITY
redefine THE POSSIBLE.

A Procrustean bed is an arbitrary standard to which exact conformity is forced.

In Greek mythology, Procrustes (the stretcher), also known as Damastes (subduer) and Polypemon (harming much), was a bandit from Attica. He had his stronghold in the hills outside Eleusis. There, he had an iron bed into which he invited every passerby to lie down. If the guest proved too tall, he would amputate the excess length; if the victim was found too short, he was then stretched out on the rack until he fit. Nobody would ever fit in the bed because it was secretly adjustable: Procrustes would stretch or shrink it upon sizing his victims from afar. Procrustes continued his reign of terror until he was captured by Theseus, who "fitted" Procrustes to his own bed and cut off his head and feet (since Theseus was a stout fellow, the bed had been set on the short position). Killing Procrustes was the last adventure of Theseus on his journey from Troezen to Athens.

Computational Intelligence ☞ Web Intelligence

Systems that can communicate in natural ways and learn from interactions are key to long-term success in Web intelligence. By focusing directly on the Web, researchers in traditional computational (artificial) intelligence topics can help in developing intelligent, user-amenable Internet systems.

The demands of the interactive, information-rich World Wide Web will challenge the most skillful practitioner. The number of problems requiring Web-specific solutions is large, and solutions will require a sustained complementary effort to advance fundamental machine-learning research and to incorporate a learning component into every Internet interaction.

Language

Natural language embodies important modalities for human-computer interactions, from simple database interfaces and machine translation to more general answer-extraction and question-answering systems.

The Workshop organizers would have us believe that "the synergy between web intelligence (WI) with brain informatics (BI) will yield profound advances in our analyzing and understanding of the mechanism of data, knowledge, intelligence and wisdom, as well as their relationship, organization and creation process." Our use of language should put this hypothesis to the test.

Language

Language creates special worlds.

Most languages are inaccessible to most of us most of the time.

It is not so much *why* some-thing happens, or *how* it occurs, as it is why we *perceive* things to be the way that they are or how we *plan* activities to occur or even what we *ruminate* in between all other thoughts that holds our interest.

Interpretation

For most of us, the world is a world of matter - wysiwyg. The superiority of physics to, say, interpersonal communication, massage, etc. derives from the assumption that if we are able to explain the physical, we may be in a position of explaining everything else.

Science, then, sets a stern face against the ancient dogged desire to interpret physical events in terms of the categories of *intention* and *belief*. It is thus that *mechanism* has come to replace *animism* in contemporary thought. And yet, curiously enough, in the concept of a machine, mechanism has itself outlived mechanics - satisfying *evidence* in science, unlike life, nothing is ever given up for good.

AI, WI, BI – general I

Where are *artificial intelligence, web intelligence, brain informatics* situated?

In the days of good old fashioned AI, the quest was to find a *general Intelligence*. Representation schemes served as the structure(s) by which the systems we built could be extended to cover more and more of a particular domain and, if we were lucky, extend to another domain as well.

Then something curious appears to have happened. We began designing logics for specific purposes, never connecting them all up with the original quest for finding a more general intelligence (surely this is what Turing had in mind when he proposed that boring party game known now as the Turing test.

Generality

Is this what is happening to the quest for generality. Sadly, I think the situation is different in this case. I am becoming a skeptic, not of the value of artificial intelligence but of the time frame we have given ourselves to produce an artificial intelligence, hence the importance to web intelligence, and of the need to encourage multidisciplinary teams of researchers to tackle problems. Let me cite some examples from natural language understanding, admitting that in each case, we could probably devise a system to tackle the particular problem, but generalizing the solution may well prove elusive.

NL Example 1

It is common in the best of the Yiddish tradition to answer a question with a question, often the very same question with different emphasis and intonation. For example, in response to the question "Did you buy flowers for your mother on her birthday?", the response would be quite different if a different word were emphasized in the answer. Thus the answer "*Did* I buy flowers for your mother on her birthday?" is quite different from "Did *I* buy flowers for your mother on her birthday?" which is different from "Did I *buy* flowers for your mother on her birthday?" and so on and so on.

NL Example 2

Consider the following excerpt from Erle Stanley Gardner's "The Case of the Demure Defendant":

"Cross-examine," Hamilton Burger snapped at Perry Mason.

Mason said, "Mr. Dayton, when you described your occupation you gave it as that of a police expert technician. Is that correct?"

"yes sir."

"What is an expert technician?"

"Well, I have studied extensively on certain fields of science that are frequently called upon in the science of criminology."

"That is what you meant by an expert technician?"

"Yes sir."

"Now what is a police expert technician?"

"Well that means that ... well, it all means the same thing."

"What means the same thing?"

"An expert technician."

:

:

:

NL Example 2 (cont.)

Mr. Dayton needs to understand the subtleties of noun phrases such as "police expert technician", to answer Mr. Mason's questions. Understanding such phrases are troublesome to automate since "police", "expert" and "technician" are all nouns.

Imagine further, the processing required by Perry Mason. The subtleties of language understanding realized by Mr. Dayton must be mastered but also the reasoning capabilities of Mr. Mason and extraction of relevant, salient conversation features be identified to generate the appropriate next question. Actually, Mr. Mason's task is much simpler than Mr. Dayton's - to generate an utterance which
conveys a presumably preexisting thought. Mr. Dayton's task as listener is to decide what Mr. Mason must have been thinking in order to motivate his utterance in the particular context in which he uttered it.

Commentary – Parsing NL

Based on HPSGs (head driven phrase structure grammars), a Java based parser is part of a larger project to implement a NL system for Internet information retrieval. This IR task requires Java applets that can parse natural language. Java supports the dynamic class loading and object serializations that our concept of distributed NL processing requires. The approach is similar to the filtering techniques that improve traditional parser performance, but it differs by requiring the filtered knowledge to be in the form of a grammar.

Motivation for this work includes potential improvements in managing the complexity of language expressions, parsing efficiency, and context based disambiguation.

Commentary – Recommender Systems

Recommender systems suggest information sources, products, and services by learning from user preferences. Two methodologies dominate:

- *Collaborative (social) filtering* methods base recommendations on the preferences of users who made choices similar to the current user's; Amazon.com has used this method for years.
- *Content-based methods* use item-specified information that represents a unique niche across different product domains; wealth demographics is an example.

Future work in collaborative filtering can merge information sources to refine the analyses and subsequent recommendations.

Commentary – Recommender Systems (cont.)

Content-based recommender systems provide a unique application for embedded classifications system. An application could use the information extracted from a set of documents—Web pages, newsgroup messages, and so on—during a word extraction phase to develop a set of examples that serve as a user training set. The rule-induction process could then extract a user profile and rank the rest of the examples accordingly. The top ranked examples then serve as items for recommendation. This process can help to personalize the recommender to individuals.

Commentary – Agent Systems

The Internet is a large, distributed, heterogeneous source of information.

Users perceive it through a set of applications based on point-to-point TCP/IP communication links. Many of its applications require finding a relevant document. In NL processing, we try to match the meaning of user queries to the meaning of retrieved documents. This involves deciding what a concept is, how to extract it from a natural language text, and how to match a concept with others that are similar. Existing systems are inefficient, but combining NL processing with multiagent systems on the Internet may yield a new way to distribute processing costs—enabling NL IR to keep pace with the growing wealth of information and resources currently available.

Final Word

Adapting existing computational intelligence solutions may not always be appropriate for Web intelligence, but when it is, the solutions must incorporate a more robust notion of learning that will scale to the Web, adapt to individual user requirements, and personalize interfaces.

For Web Intelligence to meet Brain Informatics, yet another dimension needs to be breached. Language is one medium that ensures the task is not to be taken trivially.

Post Analysis … Moving forward

Stewardesses is the longest word typed with only the left hand and lollipop with your right.

No word in the English language rhymes with month, orange, silver, or purple.

Dreamt is the only English word that ends in the letters mt.

The sentence: "The quick brown fox jumps over the lazy dog" uses every letter of the alphabet.

There are only four words in the English language which end in "dous": tremendous, horrendous, stupendous, and hazardous.

Acknowledgements

I thank the Natural Sciences and Engineering Research Council of Canada for its support of this work.

I also thank my many fine students who contributed to so many of the ideas expressed today: thanks Vlado Keselj, Xiangdong An, Jiye Li and many others