Can networked human computation solve computer language comprehension?

January 26 2009

Researchers at the University of Essex hope to answer this question by getting more volunteers to take part in their online game, Phrase Detectives.

Jon Chamberlain, from Essex's School of Computer Science and Electronic Engineering, explains: ‘Human language is not an unconnected series of words, phrases and sentences but a series of people, objects and ideas that refer to each other in different ways. The complexity of language makes it sound "natural" to a reader but it can be difficult to define the rules that allow us to understand it.

‘Consider the statement: "Mary is a teacher who is 25 years old. She lives in England." A human reader can easily ascertain facts about Mary's occupation, age and residence by, for example, knowing that the word "she" refers to the person "Mary". However, comprehending this type of language referencing is a challenge facing programmers when designing computer systems that try to understand text, such as search, translation and summarisation systems.’

This is where the work of those playing Phrase Detectives becomes important. The game, part of a larger project called AnaWiki, is an attempt to address the bottleneck in creating annotated linguistic resources. By initially investigating anaphoric references (as in the example above) the project aims to develop a resource larger than anything currently available.

Players (or detectives) register at: www.phrasedetectives.org and read through texts, making annotations to highlight relationships between words and phrases. They may be asked to 'name the culprit', so will be given a word or phrase and must look for it appearing earlier in the text. For example: 'Sherlink Holmes went to the shop. He got some tobacco for his pipe.' The word ‘he’ refers to 'Sherlink Holmes'.

Jon added: ‘Players of the game are helping to create a resource that is rich in linguistic information and improves future technology. This project aims to collect a significant amount of data and investigate the possibility of using mass collaboration to train computer systems.

‘The best way to understand a language is to have lots of examples where the meaning has been clarified. Unfortunately creating this type is resource is both time consuming and expensive but the new approach offered by Phrase Detective should address this resource shortage. The same methodology could also be used to create resources for machine translation, semantics and other linguistic phenomenon.’

So far, players have made over 40,000 annotations in four weeks. However, the researchers hope more will join as detectives and that people will add new text to the site for analysis.

Phrase Detectives can be defined as part of a genre of “games with a purpose” (GWAP) that collect data on images, texts and music. The crucial element of these games is that players receive points for agreeing with each other. They are motivated to collaborate with their partners in order to score maximum points. This ensures that players are attempting to provide good quality information, as this will result in the most agreement.

The Essex researchers believe Phrase Detectives is the first attempt to collect linguistic judgements using a fun, collaborative online game.

They aim to make the tasks and the texts interesting so it feels more like a computer game than a linguistic task. The data collected can then be used to improve computer systems that try to understand text. For example, it could help search engines find information more relevant to your searches.

So, can networked human computation really solve complex language comprehension tasks on computers? Initial results from the beta version of the game look promising and more detailed analysis will completed in early 2009.

Source: University of Essex