

First Announcement and Call for Contributions.

Workshop

User Oriented Evaluation of Knowledge Discovery Systems

Centro Cultural de Belem, Lisbon, Portugal.
25th May, 2004, afternoon

In association with the 4th International Conference on Language Resources and Evaluation:

LREC 2004

Main conference May 26th – 28th, 2004

The problem area.

Knowledge discovery systems, such as intelligent information extraction and data mining, offer special challenges to the evaluation community. The only real measure of success with such a system is whether it really will help someone to achieve an objective efficiently, in safety and with satisfaction (to paraphrase ISO/IEC 9126 talking of ‘quality in use’).

With other softwares, a task can be identified such that producing the results specified will satisfy the needs of a wide range of users: for example, a speech recognition system must accurately recognize words, a spelling checker must identify all mistaken spellings, a machine translation system must produce good quality translation, and this remains true even if the system is embedded in some larger system. In all these cases, simply achieving the specified results will be enough to achieve a certain level of quality. Furthermore, there are accepted metrics which can be applied to the system to judge whether it is achieving the specified results. Evaluators therefore create and implement metrics whose job, even if the metric is applied to system design or to system behaviour independently of context of use, is to predict whether, at the end of the day, someone will want to use the system to get some useful job done.

The situation is considerably more complicated in the case of knowledge discovery systems, where the notion of utility to a specific potential user is much more complicated. The critical question is not, for example, whether a given piece of software identifies clusters with strong intra-cluster similarity and strong inter-cluster dissimilarity, but whether the end user finds the clusters identified useful in accomplishing his task. By definition, the task of each user is similar to that of other users only at a quite high level of generality, such as the search for new insights, so that it is hard, if not impossible, to tell during system design and subsequent development whether the ultimate user will be happy or not. Of course, it would be possible to manufacture and install the system and then to test for user satisfaction in situ, but that seems a less than satisfactory solution from the system designer’s or manufacturer’s point of view.

Even apart from the problem of accounting for potential user needs, definition of metrics for knowledge discovery systems poses special problems for several reasons. First, knowledge discovery systems are typically used in situations where a mass of data too large for thorough

human understanding has to be dealt with. Secondly, in at least some situations, the data to be treated is not homogeneous in kind or in reliability. Finally these and other factors make it very difficult if not impossible for an evaluator to define what might constitute a good result. For example, if a system is supposed to discover market trends or trends in teenage behaviour which were previously unknown, how can you find out whether it does so correctly or whether there are important trends which have gone undiscovered? This is, of course, only one example of a question which might be asked.

To summarize all this in concrete terms, we give the following typical scenario, which contributors to the workshop may take as a framework for their contribution if they choose.

An organisation has a very large number of reports produced over many years. These reports contain information in the form of text, graphics and tabular data which is potentially of considerable importance to current and future projects of the organisation. It is not feasible to search the mass of reports manually. If the organisation wants to deploy a knowledge discovery system to find and present information relevant to a specified context, what criteria should it look for in a potential system, and how can it evaluate whether the system performs satisfactorily in retrieving pertinent information? If the mass of documents to be searched is even larger and perhaps dynamically changing, for example the World Wide Web, how does this change the evaluation?

Workshop format.

The main purpose of the workshop is to launch discussion on this topic. The workshop will start with brief invited presentations setting out the points of view of

- the users
- the developers
- the evaluators

The rest of the workshop will be organised around brief presentations whose main purpose is to set out a problem in the user oriented evaluation of knowledge discovery and text or data mining systems. Each presentation will then serve as the basis for larger discussion with all the participants in the workshop. Thus the workshop will be divided up into one-hour sessions, each of which will start with a twenty to thirty minute presentation.

Proposals for presentations.

We invite proposals for presentations from representatives of all those concerned by the issues:

third party evaluators, specialists in evaluation, designers and manufacturers of knowledge discovery systems and most particularly users or potential users of knowledge discovery systems.

Since the purpose of the workshop is to launch discussion, we are not asking for full papers from those who wish to make a presentation. Rather, contributions should set out the problems to be presented and should state whether a solution will also be presented. Elegant prose is not required: contributions in note form will be acceptable. Proposals for contributions may be very brief, typically between two and five pages. Final versions of the contributions will be included in the workshop workbook, which will take the place of a more conventional set of proceedings.

Submission procedure.

Proposals for contributions should be sent to

Margaret.King@issco.unige.ch

Important Dates.

Deadline for proposals for contributions: March 1st 2004

Notification of acceptance: March 8th

Preliminary Programme: March 10th

Deadline for final version of contributions: April 8th

Workshop: May 29th 2004

The workbook will be published by the LREC Local Organising Committee. Final versions of contributions must therefore conform to the style sheet that will be adopted for the LREC proceedings. This style sheet will be made available in February.

Organising Committee

Maghi King, ISSCO/TIM, University of Geneva

Hilbert Bruins Slot, Unilever Nederland BV

Myra Spiliopoulou, University of Magdeburg

Agnes Lisowska, ISSCO/TIM, University of Geneva

Nancy Underwood, ISSCO/TIM, University of Geneva

Fabio Rinaldi, Institute of Computational Linguistics, University of Zurich

Michael Hess, Institute of Computational Linguistics, University of Zurich

Further information

For any further information, please contact

Maghi King

e-mail: Margaret.King@issco.unige.ch

ISSCO/TIM/ETI

University of Geneva

Uni-Mail

40 blvd du Pont d'Arve

CH 1211 Geneva 4

Phone: +41 +22 739 87 55

Fax: +41 +22 739 86 89