How to get the right systems to the right users?

Nancy L. Underwood
Nancy.Underwood@issco.unige.ch
ISSCO / TIM, Ecole de traduction et d'interprétation,
Université de Genève, Boulevard du Pont-d'Arve,
CH-1211 Genève 4,
http://www.issco.unige.ch

Abstract

This paper considers a number of issues in evaluating Knowledge Discovery systems from the users' point of view. It is intended as a contribution to the discussion on how to ensure truly user-oriented evaluation of such systems. As such it puts forward a number of problems and open questions to be discussed during the workshop..

1. Introduction

This paper is intended to provide input to a discussion during the workshop on how to ensure trully user-oriented evaluation of KD systems. For the purposes of discussion we first present a working definition of KD at a general level (Section 2). Section 3 considers the nature of user-oriented evaluation and points out some of the shortcomings of more traditional metrics which have been applied to similar, but simpler systems for extracting information from data. In Sections 4 and 5 we present some more concrete suggestions for better metrics and the beginnings of a checklist for modelling users of KD systems. We conclude with a number of open questions concerning the next steps which should be taken to achieve truly user-oriented evaluation of KD systems.

2. Knowledge Discovery: a working definition

In this paper we take the goal of Knowledge Discovery (KD) to be the uncovering of knowledge or rather information from data (in the form of patterns and relationships) which was not previously apparent or readily accessible because of the amount and complexity of data available to the user. Thus KD systems are typically applied to very large amounts of data which may also be heterogenous (e.g. police intelligence gathered from different sources) and/or constantly updated (e.g. newswires, customer purchase records, the web).

In order to reveal such knowledge, rather complex systems involving, for example, data mining and intelligent information extraction techniques have been developed. In general, data mining techniques produce results which by their very nature are statistical and indicative rather than factual and which need to be interpreted by the user in order to be useful. Thus a KD system is very much a supporting technology to aid the user in carrying out a higher level task based on the knowledge which is uncovered.

These properties of KD give rise to particular for problems specifically user-oriented evaluation.

3. User oriented evaluation of KD systems

Carrying out a user-oriented evaluation of a system involves evaluating how well that system addresses the needs of the user. In ISO terms (ISO/IEC, 2001), the question we need to address concerns how we can reliably evaluate the "Suitability" of a system or piece of software. This should not be confused with evaluating the "Accuracy" of a system which will only tell us how well the system performs the tasks which it has been designed to carry out in the first place. Although accuracy may sometimes be a good inicator of suitability, it is perfectly possible to have an well designed and implemented system which fulfills all its design specifications but which is nonetheless not suitable for a particular user or group of users.

3.1 Shortcomings of traditional metrics

Much previous work in evaluation (see e.g. EAGLES (1996), the MUC and TREC web sites) has focussed on developing metrics for system attributes which are also assumed to predict the utility of a system for a typical user or group of users. In the field of KD evaluation we do not rule out the possibility of defining some such metrics. However, the open-ended nature of KD systems calls this approach into question.

In the fields of IE and IR evaluation for example, the standard metrics have traditionally been precision and recall. These metrics rely crucially on the construction of gold standards, which specify in advance the correct result which a system should produce. However by definition, the results obtained from KD should be new and unpredictable. So in principle, it should not be possible to pre-determine the results a system should produce and therefore such metrics seem to be difficult, if not impossible, to apply in the case of KD. It may of course be possible to try and plant