Framework for quality assessment of knowledge

J. Brender\textsuperscript{a}, J. Talmon\textsuperscript{b}, P. McNair\textsuperscript{c}

\textsuperscript{a}Medical Informatics Laboratory ApS, Stengaards Allé 33d, DK-2800 Lyngby, Denmark. \textsuperscript{b}University of Limburg, Dept. of Medical Informatics, P.O.Box 616, 6200 MD Maastricht, The Netherlands. \textsuperscript{c}Hvidovre Hospital, Dept. of Clinical Chemistry339, DK-2650 Hvidovre, Denmark.

Abstract: One of the key issues in the development (and subsequent application) of medical knowledge - be it in terms of a KBS or otherwise - is the assessment of its quality. We present a framework for how to manage and make measurable the quality of the semantic as well as pragmatic aspects of the knowledge embedded in classification models during the development of such models.

1. Introduction

The KAVAS\textsuperscript{1} and KAVAS-2\textsuperscript{1} AIM Projects have developed a framework for the assessment of quality of (executable) classification models [Brender et al. 1990a, 1990b]. It has the purpose of managing and making measurable the quality of classification models in the semantic sense and in relation with its usefulness in application. Although the framework was designed for the assessment of Knowledge-Based Systems in medicine it is generally applicable. The emphasis in this contribution will be on the pragmatic issues of quality assessment during the development of classification models.

1.1 The KAVAS approach for development of classification models

KAVAS-2's main objective is to develop a toolbox, named KAVIAR ("Knowledge Acquisition, VIsualisation Assessment and Refinement"), to be used by medical experts for developing and assessing domain knowledge. Innovative elements of KAVIAR are the following key functionalities [Brender et al. 1993a]: 1) built-in quality assessment of semantic aspects of knowledge [Brender et al. 1993b], 2) features for controlling the machine learning process in order to make it focus on the desired qualities, a process called Conditioning [Talmon et al. 1991, McNair et al. 1993], and 3) coupling of machine learning techniques with knowledge elicitation techniques thereby enabling the support of the user in varying clinical situations where different data and knowledge sources are available.

The process of using KAVIAR compares to an iterative learning process:

1) Goal definition, i.e. the formulation of the medical problem approached and elicitation of conditioning information.

2) Selection (and parameter setting) of appropriate tools and techniques to be applied on the basis of among others the medical problem focused on and the conditioning information.

3) Application of the selected tool(s).
4) Measuring and exploring the characteristics and quality of the domain models.
5) Validation of the domain models against the defined goals, leading to either a new learning iteration or to an implicit or explicit acknowledgement or rejection of the outcome.

These five operational steps correspond to the conceptual steps for tasks' accomplishment in general: goal clarification, selection of means, carrying out the task, and assessment of task, see Figure 1.

Presently, three machine learning techniques are implemented in KAVIAR: 1) an induction tool providing decision trees (i.e. protocols), 2) a neural network tool for creating feed-forward networks, trained by the back-propagation mechanism, and 3) a density inference algorithm which provides a probabilistic estimate of which class(es) in a given clinical database the case at hand resembles most. Fur-
ther, tools are implemented for a) definition and formalisation of concepts e.g. by capturing them from a text, b) representation of (semantic) relations between concepts (Conceptual Graph tool), and c) representation of procedural knowledge (State-Transition-Diagram tool), supporting the modelling of knowledge of various epistemological types (causal, goal-oriented, dynamic, context and basic biomedical knowledge). The latter tools have a dual purpose: to enable the user in formulating his own knowledge as a means for validation of the classification models, and to support the other tools (conditioning tools, the machine learning tools and the quality assessment tool) in circumscribing the clinical problem and needs.

KAVIAR may be seen as a Decision Support System for the development of clinical protocols. The learning process can be managed at any level of automation. The most automated being the one that builds models purely on the basis of the optimality criteria implemented in the tools; this is the approach taken in most implementations of induction algorithms. At the other extreme, support in data inspection is provided and full user control of all parameter settings for the involved tools is feasible. The induction tool has a manual mode, where the tool supports the user in each decision during the tree building process. All intermediate levels of human intervention between these extremes are feasible and supported by KAVIAR.

KAVAS's philosophy is to develop and validate classification models by means of several supplementary (and integrated) tools and techniques. Classification models based on machine learning techniques are by nature executable, but it is up to the user to implement them in clinical practice either as computer-based or paper-based classification models.

1.2 Quality Assessment during the development of clinical protocols
The purpose of the framework for quality assessment and its implementation in the KAVIAR tool is to make it an integrated part of the entire process of development of clinical protocols and other classification models, and to enable users to fully assess the resulting models. The overall approach is that of integrating development and evaluation to enable iterative refinement cycles leading to a satisfactory result. The user requirements (i.e. the conditioning information) as regards the required classification model are quality aspects like predictive values, coverage as well as ethical and economical constraints [McNair et al. 1993]. They serve 1) as a basis for tool parameter settings and/or 2) in a multivariate criteria analysis as a basis for selection among numerous candidate classification models. Finally, they serve as means for the quality assessor functionality in providing the user with information on the degree of fulfilment of user requirements and on potential reasons for experienced deviations, see Figure 1.

It is clearly seen that the quality management with an active and explicit quality assurance is integrated in the entire set-up of the learning process in KAVIAR.
from the point of specification of goals and requirements to the final assessment of outcome.

2. The Quality Assessment Framework

Most of the traditional quality measures of semantic aspects of medical knowledge are confined to 2 class problems (like Predictive Values, Sensitivity and Specificity) and/or have the assumption that all cases are labelled. These assumptions are not always fulfilled and one often is confronted with multiclass problems. The framework has served not only as a design principle but also as a means for structuring the development of new and generalised metrics and measures.

The KAVAS Framework for quality assessment constitutes a high level semantic description of quality management. It is composed of a set of orthogonal perspectives [Brender et al. 1993b] containing several concepts. Only a subset of these perspectives and their concepts will be addressed here:

a) The Quality Concepts Perspective, providing a comprehensive view on various aspects of quality.

b) The Quality Management Perspective, structuring the activities of a process ensuring the fulfilment of quality needs.

c) The Metrology Perspective, covering concepts in measurement of quality.

d) The Perception Perspective dealing with conceptual viewpoints on the quality concepts.

2.1 Quality Concepts Perspective

A literature study revealed that many quality concepts are defined at a measurement level, while only a few are defined at a conceptual level. Moreover, when a quality concept embraces only one measure this is often mistaken as identical with the concept itself. For instance, “correctness” is the term used everywhere in the literature when dealing with quality of KBS application. However, the literature does not discriminate between the quality concept itself and the measures (percentage of correct classifications) which characterise it and which are concepts in themselves.

In [Brender et al. 1990 a and b] we defined a preliminary minimum set of meta-knowledge (quality) concepts by generalisation and harmonization of quality concepts used in various domains. The notion of meta-knowledge stems from the perspective of quality of knowledge as knowledge-about-knowledge (“meta-X” is defined as “X about X” [Aiello et al. 1986]).

The quality concepts were acquired by review of the literature in four medical domains, computer science and medical informatics and subsequently generalised and abstracted. The concepts identified are connected on the one hand to quality assurance of semantics of the knowledge-base and properties of the KB(S). On the
other hand, they provide the basis for the assessment of possible consequences of application of the KB(S) in a given environment. For some of these concepts, one can define metrics and measures which can be obtained by examination of the KB(S) or by exercising the KB(S) with one or more sets of cases, while other measures require human intervention to achieve their measurement. Examples of the former type of quality concepts are e.g. correctness, coverage, robustness, model complexity, problem solving strength, and goodness of fit, and examples of the latter are e.g. usefulness and reliability.

2.2 The Quality Management Perspective

Quality Management (QM) is the act of dealing with quality in a goal oriented and structured way. Within QM there are three concepts constituting a layered operational structure: quality assurance (QA), quality control (QC) and quality inspection (QI).

From the ISO definition\(^1\) it may be concluded that implementation of a quality policy implies that a goal and an approach/strategy for the goal’s accomplishment have to be established; the goal serving as a frame of reference for further activities. This process is called QA. QA is concerned with strategic issues (objectives, premises) and planning activities (e.g. selection of methods) to ensure the desired level of quality.

The operationalisation of QA on the concrete problem case is the QC; the pragmatic set-up of relevant work processes is the QI. QC is concerned with combining and setting up the selected methods on the concrete problem case to meet the goal and fulfil other strategic decisions. QI is concerned with practical aspects in the measurement of individual measures. These three aspects function in a conceptual hierarchy of increasing abstraction from QI to QC and further on to QA. Each of these levels are operationally linked to the other levels.

The above description highlights the establishment part of QM. However, there is implicitly also a follow-up part of the QA, QC and QI in the ISO definitions to maintain the fulfilment of the quality policy: the most simple quality assessment constitutes the measurement part of the QI alone, i.e. providing measures only. QI together with a frame of reference for the quality measures lead to the level of QC. When the interpretation of the measures is related to the question of “good enough for the purpose?” the level of QA emerges.

When viewing the learning cycle in this perspective (see Figure 2) it is easily seen that the quality management is actively and explicitly dealt with in KAVIAR.

2.3 The Perception Perspective

From the definition of *correctness*\(^4\) is it clear that there are at least two viewpoints in semantic aspects of quality of knowledge: i) the success characteristics (“general validity of the system’s output when applied on a population of cases of inter-
est”), and ii) the assumptions characteristic ("it concerns only those cases that are classified").

In clinical biochemistry, biostatistics and other domains that formally apply metrology, it is also common practice to express characteristics of failures, for example in terms of the measures of dispersion (randomness in measurement) and bias (systematic error) (see e.g. [Magid 1992]). Hence, it is obvious also to look for failure characteristics in semantic aspects of quality of knowledge. An example from statistics is: the variation around a linear regression line expresses how good the model fits the data, i.e. it constitutes a success estimate. The existence (and nature) of a pattern within the residual variation indicates the failure of fitting the model chosen (e.g. a 2nd order polynomial) with the data, i.e. a failure characteristic. An assumption is that the error of an individual datum is normally distributed.

Note, that the perception of success and failure is always relative to the user needs and requirements (see also below), and constitutes a basis for supporting the user during his quality assessment of the outcome, see Figure 2.

Figure 2: The KAVIAR learning cycle in view of the quality assessment framework.
2.4 The Metrology Perspective

The hierarchy within the concepts of measurement is *methodology, methods, metrics and measures*. A *methodology* comprises the global approach (goal definition and strategies) together with a consistent and coherent set of methods from which one may select and combine a subset to achieve the goal; a *method* is a formalized description (including theories, application range and assumptions) for accomplishing a task; a *metric* is a formalized procedure or technique for measuring a quality or characteristic, therein also its basic theoretical framework; and a *measure* is a named concept that may bear a value.

The methodology is related with the information need and serves as the context for fully interpreting the measures. Methods have assumptions and therefore add to the context for interpretation of measures. One methodology may have numerous alternative methods to be applied for a task, and one method may be applicable within numerous methodologies.

The idea is that for each of the quality concepts identified (defined preliminary in [Brender et al. 1990 a and b]) there shall exist a methodology for establishing and interpreting their value, by means of the methods, their operational formula (the metrics) and the measures to which they give a value.

Viewed in KAVIAR's perspective, the methodologies and methods are implicitly implemented in the quality assessment and the conditioning, while the metrics are the actual calculations to be performed. The measures within their context for interpretation are the pieces of information to visualize towards the user during the quality assessment (see Figure 2).

2.5 Practical aspects in assessment of quality of knowledge

From the above four sections it is clear how the mentioned perspectives within the quality assessment framework relate with the actual learning process in KAVIAR. The user of KAVIAR normally operates with a learning database and one or more test databases, of which some may be derived artificially to assess e.g. the vulnerability towards added noise on the attribute values (the robustness quality concept). Application of the generated classification algorithms on these populations of cases provides a lot of information which may be further processed into the actual quality measures.

KAVAS-2 has experimented with metrics and measures for a number of quality estimates [Brender et al. 1994, Egmont-Petersen et al. 1994]. For instance, we have identified several quality measures for the quality concept correctness:
- overall conformance (concerned with all classes and cases), label respectively class conditional conformance (concerned with conformance of the labelling respectively the true class)
- overall and conditional Kappa conformance (conformance corrected for the proportion of outcome classifications that appear by chance)
- bias.
In medicine, it is often not equally important to distinguish class A from class B as it is to distinguish class B from class A. In a situation where the distinction shall be made between say normals and cancer cases it is important not to overlook a cancer and may be unimportant to include a (small) proportion of normals into a group of cancer candidates which anyway will be prone to further clinical examinations and selections before initiation of a treatment. This means that the information theoretic optimality is not the same as the ethical optimality. The user (clinician) will require a classification model that has a bias reflecting his clinical situation. Hence, even large biases may be desirable, and therefore, what is relevant as the frame of reference in the quality assessment of a model is the deviation from this user requirement. Hence, what constitutes the success and the failure characteristics in a given quality assessment session entirely depends on the user requirements (the goal specification) for the classification models.

3. Conclusion

The major quality of the framework for quality assessment of knowledge is the structure which supports the development of quality measures. However, the framework has not only enabled the development of such quality measures. More important, it constitutes an integrated part of the design of the KAVIAR tool, supporting the user and enabling an active quality management in his development of classification models. As it stands now the framework constitutes a starting point for explicit quality management during the development of knowledge-based systems.

Note, that there is a distinction between quality assessment in the development phase and the application phase of a classification model. The former is population-oriented, while the latter is case-oriented. However, the framework may serve as a basis also for the development of metrics and measures in this respect.

Notes

1. KAVAS is an acronym for “Knowledge Acquisition, Visualization and Assessment System”

2. Meta-knowledge is “knowledge about knowledge”, and as knowledge is a perception of causal relations and associations, meta-knowledge becomes “knowledge about causal relations and associations”, which we interpret as “knowledge about qualities and characteristics of the perception of causal relations and associations”. When expressed in an operational form, meta-knowledge may be defined as “knowledge about when, how, why and to what extent the knowledge can be applied and with what level of quality and confidence”, i.e the concepts of meta-knowledge “bear” the values and the context for interpretation of quality of knowledge.
According to [ISO 8402, 1986] Quality Management is "that aspect of the overall management function that determines and implements the quality policy"; Quality Assurance is the "establishment and follow-up of strategies, including objectives, premises and methods applied to achieve the desired level of quality" (adapted from the ISO definition to achieve the same grammatical form from); Quality Control is the "establishment and follow-up of procedures, thereby operationalizing the strategies for a concrete case" (adapted from the ISO definition to achieve the same grammatical form; and Quality Inspection is the "establishment and measurement of individual quality measures" (ISO has no definition for quality inspection, but defines inspection as "activities such as measuring, examining, testing, gauging one or more characteristics of a product or service and comparing these with specified requirements to determine conformity".

Correctness is defined as "the general validity of the system's output when applied on a population of cases of interest"; correctness is meaningless, when (part of) input information is clinically inconsistent or when the knowledge is not applicable. Hence, it concerns only those cases that are classified.

Metrology is the science of measurement.

Acknowledgements

This work was partly funded by the Commission of the European Communities under the AIM Exploratory Action (KAVAS (A1021) Project) and the AIM Telematics in Health Care Program (KAVAS-2 (A2019) Project).

References


