Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

Collaborative Knowledge Discovery & Data Mining: From Knowledge to Experience

Timo Horeis and Bernhard Sick
Faculty of Computer Science and Mathematics, University of Passau, Germany
e-mail: {horeis, sick}@fmi.uni-passau.de

***Abstract*— Experts have important qualitative knowledge about interrelations between more or less abstract concepts in an application area. However, the knowledge of a single expert is typically quite uncertain (e.g., incomplete or imprecise). By fusing the knowledge of several experts it would be possible to obtain more certain and, therefore, more valuable knowledge. Conventional systems for Knowledge Discovery (KD) and Data Mining (DM) have the ability to extract valid rules from huge data sets. These rules describe dependencies between attributes and classes in a quantitative way, for instance. By fusing this kind of knowledge with the combined, qualitative knowledge of several experts it would be possible to obtain more comprehensive knowledge about an application area. In this article, we propose a concept for a new KD & DM technique based on Computational Intelligence: Collaborative Knowledge Discovery (CKD). These technique combines the uncertain knowledge of several experts using methods based on Dempster-Shafer theory. The combined human knowledge is again fused with automatically extracted, well interpretable knowledge (fuzzy rules embedded in a radial basis function neural network) of a conventional KD system. Thus, a CKD system not only acquires more comprehensive knowledge, but also experience (knowledge about knowledge), meaning that it is able to explain automatically extracted rules to the human experts and to assess the interestingness (e.g., novelty or utility) of these rules. This can be done by adapting inference mechanisms from the field of Probabilistic Argumentation Systems. A CKD system will comprise self-awareness mechanisms (it must know what it knows) as well as environment-awareness mechanisms (it must know what human experts know or what they want to now). In order to reduce the effort for knowledge acquisition, a CKD system must learn (pro-)actively. There are many application areas for such CKD systems, e.g., in the field of technical data mining (quality control, process monitoring, etc.).**

I. Introduction

In many technical applications, conventional systems for Knowledge Discovery and Data Mining (in the following briefly referred to as KD systems) are based on data with a non-human origin. That is, they typically utilize data from sources such as microphones, cameras, ultrasonic sensors, or displacement sensors. Only minimal information provided by humans (e.g., application experts) is involved, e.g., class labels for a supervised training of classifiers. Usually, the knowledge extracted from these data is not combined with the existing complementary human knowledge about a given application. Thus, conventional KD systems often use only a part of the actually available knowledge about an application area. With a fusion of human knowledge and automatically extracted knowledge it would be possible to gain a more comprehensive and a more valuable view of an application.

If at all, human knowledge about an application is mostly acquired from only one expert. In this case a validation by means of statements of other application experts is not possible and the knowledge of a single human is assumed to be certain. However, human knowledge is potentially uncertain (e.g., imprecise or faulty). With a fusion of the knowledge of several experts with various levels of expertise it would be possible to obtain more certain, high-quality knowledge concerning the various aspects of an application.

Conventional KD systems have little information about the human experts involved in the KD process, about their needs and their expertise. Thus, they are not able to assess various aspects of knowledge extracted from data. For example, in the field of KD knowledge is termed to be *interesting* if it is *valid*, *novel*, *useful*, and *understandable* [1]. In general, interestingness is equated with validity of knowledge and this aspect is rated with statistical methods (e.g., based on data by cross-validation or bootstrapping). With a fusion of human expert knowledge and knowledge which is automatically extracted from data it would be possible to identify application knowledge which is novel or unexpected (either in general or for a particular expert) or which allows the application experts to initiate useful actions.

This article introduces a concept for a new kind of KD systems which we call CKD systems (CKD: Collaborative Knowledge Discovery). These CKD systems ...

- ... fuse the potentially uncertain knowledge of several application experts in a collaborative approach (in particular with an active knowledge acquisition behavior),
- ... combine this fused human knowledge with automatically extracted knowledge of a conventional KD approach to obtain more comprehensive and more valuable knowledge about an application area, and
- ... allow the interestingness assessment of this knowledge (not only validity, but also novelty or utility, for instance).

Altogether, we can say that such a CKD system automatically gains experience, i.e., knowledge about knowledge (meta-knowledge).

The remainder of the article is structured as follows: Section II introduces some important terms, describes the abilities of a CKD system from an user's viewpoint, and discusses some related work. A detailed suggestion for the realization of a CKD system can be found in Section III. Finally, Section IV gives an outlook to additional ideas.

1-4244-0705-2/07/$20.00 ©2007 IEEE

II. Prerequisites

In this section we will first introduce some new, important terms. Then, a fictive application scenario shall give an overview of the desired properties of a CKD system. Finally, we discuss some related work in the field.

A. Knowledge and Uncertainty – Some Important Terms

Knowledge and *uncertainty* are two terms that play an important role for CKD systems. In the literature they are utilized in various ways (for the term *knowledge* cf. [1], [2], [3], for the term *uncertainty* of knowledge cf. [4], [5], [6], [7], for instance). That is, we have to explain these terms in our context. Two other terms – human-driven knowledge and data-driven knowledge – are introduced because the underlaying concepts cannot completely be characterized by existing terms.

For the term *knowledge* we adopt the semantics from KD which characterizes knowledge as "interesting patterns" (relationships) in data [1]. Here, it is important that knowledge can be more or less valid or even wrong. It may have subjective or objective aspects and it may be missing. If it is available (in an explicit or an implicit form), it may be uncertain. The meaning of the term *uncertainty* is adopted from [5]. That is, "uncertain" is a kind of generic term for other terms such as "likely", "doubtful", "plausible", "reliable", "imprecise", "inconsistent" or "vague". We address and model various kinds of uncertainty; this will become evident in the following sections.

Data-driven knowledge is application-specific knowledge which is extracted from data by conventional KD systems. It can be represented by means of association rules, fuzzy rules, Bayesian networks, or Neural Networks, for instance. Data-driven knowledge describes, for instance, how certain classes depend on attributes (features) of a data set using a fuzzy decision rule with appropriate linguistic terms. Data-driven knowledge may be uncertain. Depending on the type of uncertainty, this may by modeled with probabilities (Bayes theory), membership degrees (Zadeh's fuzzy theory), or potentials (Dempster-Shafer theory), for instance. Uncertainty of data-driven knowledge is often reduced by providing a large number of samples.

Human-driven knowledge is application-specific knowledge, too, but this kind of knowledge originates from human experts. They have a certain expertise concerning an application area. For example, they have knowledge about causal relationships between various concepts, such as: "$Friction$ creates $scratches$". Human-driven knowledge often describes dependencies without using information about the (numeric) characteristics of attributes. Uncertainty is usually not stated explicitly. A quantification by numerical values is possible, e.g., by means of a hybrid representation based on statements expressed in propositional logic annotated with additional probability values. Uncertainty can be reduced by combining the human-driven knowledge of several experts.

If we compare data-driven and human-driven knowledge, we can notice that there are substantial differences concerning their origin, contents, and representation. Often, both address different aspects of an application at different levels of abstraction, e.g., a functional description of relationships between attributes and classes or influences of machine tools on product quality. However, there are often some overlaps in the contents which could be exploited.

If we want to describe data-driven and human-driven knowledge by means of some existing terms, (cf. [3], [8]), we can state that data-driven knowledge is often provided in an implicit way (it must be extracted from data). It typically has a quantitative nature and it is less abstract (with respect to the application) than human-driven knowledge. Human-driven knowledge is (at least in an initial phase of knowledge acquisition) explicitly provided by human experts. Both types of knowledge must be represented in a numerical or symbolic form in order to be processed further by any formal methods.

In the following, we will equal data-driven and human-driven knowledge with its corresponding representation.

B. Application Scenario of a Future CKD System

To illustrate our objectives, we will now outline a vision of a future CKD system and its interaction with human experts.

Anomalies (quality defects such as scratches, cracks, or adhesive small particles) occurring in a manufacturing process of silicon wafers must be classified automatically. Images showing examples of various anomalies are available and class labels – required here to apply supervised learning mechanisms for the training of classifiers – are acquired from several application experts involved in this CKD process. The CKD system actively selects anomalies which it knows as being difficult to classify and asks experts who it considers being competent. As human application knowledge is potentially uncertain, the CKD system identifies samples with uncertain class labels and presents the corresponding images to other experts for labeling. Thus, the labeled data set needed by a conventional KD system for the extraction of data-driven knowledge emerges from a collaboration of several application experts. Then, a conventional KD system being part of the CKD system extracts this data-driven knowledge from the labeled data set utilizing appropriate libraries with methods for feature extraction (application-specific), feature selection, and model selection. Data-driven knowledge is represented in the form of fuzzy decision rules which can – in principle – be understood by the human experts, e.g.,

$$\text{If } length = high \text{ and } width = low \text{ then } class = scratches,$$

with suitably defined membership functions.

Simultaneously, human-driven application knowledge is directly acquired from several experts (see Figure 1). This kind of knowledge is received in form of causal statements of human experts such as

$$Friction \text{ creates } scratches.$$

These statements are regarded as potentially uncertain. Therefore, human-driven knowledge of several experts must be fused (superimposed). Similar to the acquisition of class labels, the CKD system actively controls the acquisition of human-driven knowledge by specifically asking for statements that are still uncertain and by selecting competent application experts.

Then, the CKD system automatically fuses the basis of human-driven knowledge with the basis of data-driven knowledge (cf. Figure 2). This can be done by matching the terms used in both knowledge bases. Thus, it will be possible, for instance, to utilize the human-driven knowledge together with appropriate inference mechanisms for an analysis of the data-driven knowledge.

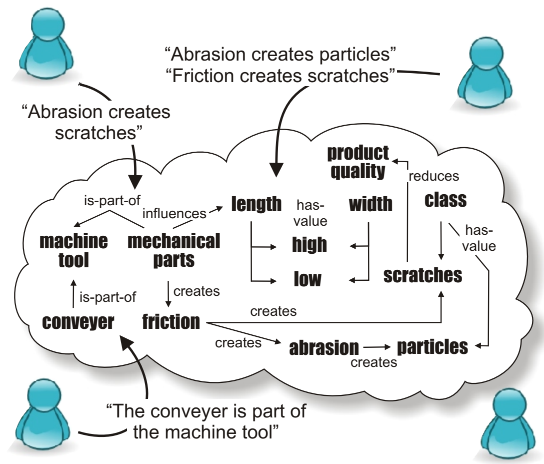

Fig. 1. Acquisition of human-based knowledge from experts.

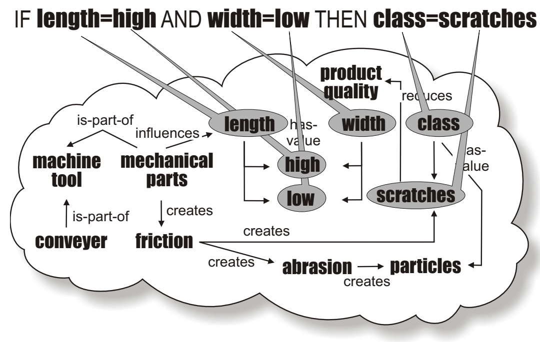

Fig. 2. Fusion of data-driven and human-driven application knowledge.

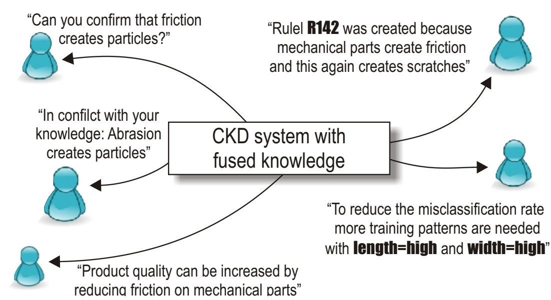

Fig. 3. Utilization of the CKD system with fused application knowledge.

The CKD system is now able to deal with the fused knowledge as set out in Figure 3. For example, the CKD system can rate the interestingness of knowledge. That is, it can decide whether certain knowledge is useful, interpretable, novel, or even unexpected for an particular expert. The fuzzy rule (data-driven knowledge) mentioned above can be explained automatically by inferring from the human-driven part of the fused knowledge base that long and small scratches originate from friction within a machine tool etc. This understanding of data-driven knowledge can now be used to select experts who are known to be interested in this kind of knowledge (e.g., an application engineer responsible for product quality) as they are able to initiate appropriate actions that solve a problem (e.g., to reduce friction in order to increase product quality again). Furthermore, it is possible to detect novel knowledge, for instance, if it not possible to infer from the human-driven knowledge base that long and small anomalies must be classified as scratches. Altogether, the CKD system acquires knowledge about knowledge, that is, it gains experience.

C. Related Work

Semantic Web Mining (SWM) [9] and Knowledge Discovery for Ontologies (KDO) [10], [11] focus on combining conventional KD systems with ontologies. Both fields differ mainly by the structure of the data being analyzed. SWM analyzes the WWW by means of web mining techniques (e.g., [9]), KDO focuses on the analysis of datasets stored in relational database management systems (e.g., [12], [13]).

Adibi et al. [14] developed KOJAK, a hybrid system for the analysis of dependencies in databases which uses techniques from the fields of knowledge management and cluster analysis. A framework concept for a knowledge-based analysis of patterns extracted by KD systems was proposed by Pohle [15], [16]. The sketched framework suggests to combine techniques form the fields of knowledge representation, automatic reasoning, and Soft Computing.

Current research also covers the use of ontologies for a data-driven analysis of large datasets: Bloehdorn et al. [17] apply self-learning ontologies to the categorization of text documents. Froehner et al. [18] focus on the fusion of knowledge acquired from different sources (specific software agents) with knowledge stored in ontologies. Nazeri and Bloehdorn [12] integrate ontologies into KD algorithms (A-Priori and C4.5) to reduce the search space and to support the extraction of interesting rules. Phillips and Buchanan support human experts in solving the feature construction and selection problem by using domain-specific ontologies [19]. Svatek et al. [13] adopt ontologies to support the understandability of extracted association rules by manually linking the attributes of a rule to a domain-specific ontology. Vanzin and Becker [20] apply ontologies in the field of Semantic Web to the analysis of web usage patterns. Their work focuses on increasing the understandability of extracted patterns and supporting an explorative analysis.

III. Realization Concept

In this section we will discuss how a CKD system could be realized. This CKD system will have the architecture set out in Figure 4. The tasks of the components depicted in this figure will be described in the following sections.

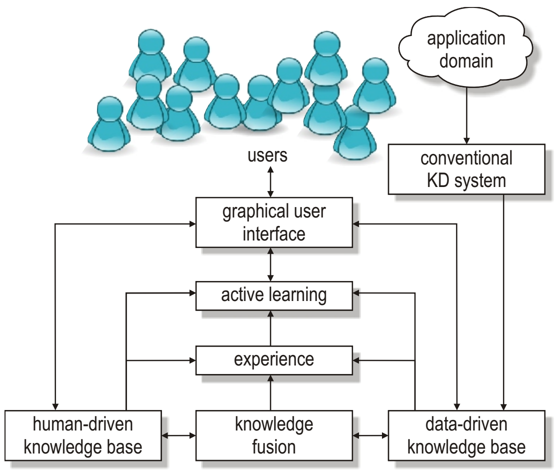

Fig. 4. Architecture of a CKD system.

A. Extraction and Representation of Data-Driven Knowledge

For the extraction and representation of data-driven knowledge we need a classifier paradigm together with appropriate techniques for feature and model selection. The classifier must adjust its parameters automatically utilizing labeled data. Furthermore, it must represent the extracted knowledge in a form which can be understood by human experts.

Appropriate classifier paradigms can be found in the field of Soft Computing: E.g., Neural Networks which can be trained with (labeled) data or Fuzzy Systems consisting of interpretable decision rules. Even if there is some work on rule extraction from Neural Networks (cf. [21], [22]) and the training of fuzzy classifiers from data (cf. [23], [24]), the natural way would be to define a paradigm which is both, a Neural Network (NN) and a Fuzzy System (FS).

Therefore, we define – in accordance with the discussions in [25], [26], [27] – the radial basis function classifier RBF̶S as a hybrid system that can be seen as both, an RBF NN and a Mamdani-type FS (cf. [28]). From the viewpoint of a NN, the RBF̶S may be defined as follows (cf. Figure 5):

1) The RBF̶S has three layers of neurons: input layer U_I, hidden layer U_H, and output layer U_O. Feed-forward connections exist between U_I and U_H as well as between U_H and U_O. A scalar weight ($w^{(I,H)}_{(i,j)}$ or $w^{(H,O)}_{(j,l)}$) is associated with each connection.
2) The activation of each hidden neuron $j \in U_H$ is determined using a multivariate Gaussian function:

$$a^{(H)}_j(k) \stackrel{\text{def}}{=} \frac{a'_j(k)}{\sum_{m=1}^{|U_H|} a'_m(k)}$$

with

$$a'_j(k) \stackrel{\text{def}}{=} e^{\left(-\sum_{i=1}^{|U_I|}\frac{(w^{(I,H)}_{(i,j)}-x_i(k))^2}{r_{(i,j)}{}^2}\right)} = \prod_{i=1}^{|U_I|} e^{\left(-\frac{(w^{(I,H)}_{(i,j)}-x_i(k))^2}{r_{(i,j)}{}^2}\right)},$$

where $\mathbf{x}(k) \stackrel{\text{def}}{=} (x_1(k), \ldots, x_{|U_I|}(k))$ is the input vector (sample) and $k = 1, 2, \ldots$ denotes its identifier. The activation function is parameterized by the weight vector $\mathbf{w}^{(I,H)}_j \stackrel{\text{def}}{=} (w^{(I,H)}_{(1,j)}, \ldots, w^{(I,H)}_{(|U_I|,j)})$ and a parameter vector $\mathbf{r}_j \stackrel{\text{def}}{=} (r_{(1,j)}, \ldots, r_{(|U_I|,j)})$.
3) Each output neuron $l \in U_O$ computes its activation as a weighted sum:

$$a^{(O)}_l(k) \stackrel{\text{def}}{=} \sum_{j=1}^{|U_H|} w^{(H,O)}_{(j,l)} \cdot a^{(H)}_j(k).$$

The external output vector of the network, $\mathbf{y}(k) \stackrel{\text{def}}{=} (y_1(k), \ldots, y_{|U_O|}(k))$, consists of the activations of output neurons, i.e. $y_l(k) \stackrel{\text{def}}{=} a^{(O)}_l(k)$.

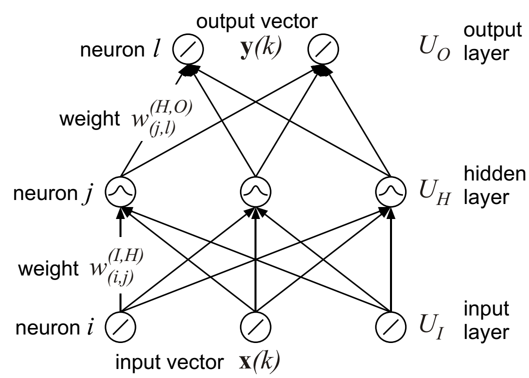

Fig. 5. Structure of a radial basis function neural network classifier.

Note that with an abbreviation for univariate Gaussians $a'_j(k) \stackrel{\text{def}}{=} \prod_{i=1}^{|U_I|} \varphi_{(i,j)}(k)$. The $\varphi_{(i,j)}$ are called basis functions; $w^{(I,H)}_{(i,j)}$ is the center of such a basis function and $r_{(i,j)}$ is its radius. The vectors $\mathbf{w}^{(I,H)}_j$ and $\mathbf{r}_j$ describe an axes-oriented hyperellipsoid in the input space of the RBF̶S. Thus, $\mathbf{w}^{(I,H)}_j$ can be regarded as a center of a hyperellipsoidal cluster – big x in Figure 6 – and $\mathbf{r}_j$ defines the shape of the cluster – ellipses in Figure 6. The activation of a hidden neuron describes the similarity between an input vector $\mathbf{x}(k)$ and a center based on a matrix norm (Mahalanobis distance measure).

The parameters of an RBF̶S must be determined by means of training algorithms such as gradient-based techniques or clustering techniques in combination with methods for the solution of linear least-squares problems (see, e.g. [29]).

For a classification problem, each class is typically assigned its own output neuron using an orthogonal representation of

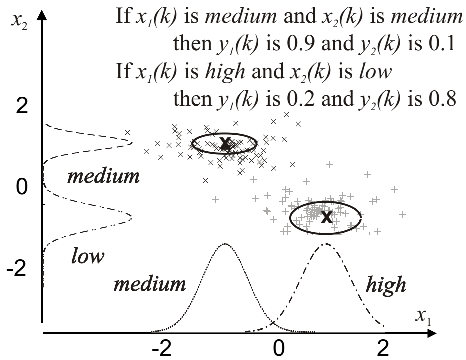

Fig. 6. Example of a classifier consisting of two rules operating in a two-dimensional input space ($|U_I| = 2$ and $|U_H| = 2$).

classes for training. A winner-takes-all approach is used for the final decision on class membership.

From the viewpoint of FS we can say that we have defined an FS with $|U_I|$ inputs, $|U_H|$ rules, and $|U_O|$ outputs (here: classes). The membership functions of the input variables correspond to the Gaussian basis functions of the hidden neurons, singletons are used for the output variables. That is, a fuzzy rule j ($j = 1, \ldots, |U_H|$) has the form

$$\text{if } x_1 \text{ is } \varphi_{(1,j)} \ \ldots \text{ and } x_{|U_I|} \text{ is } \varphi_{(|U_I|,j)}$$
$$\text{then } y_1 \text{ is } w^{(H,O)}_{(j,1)} \ \ldots \text{ and } y_{|U_O|} \text{ is } w^{(H,O)}_{(j,|U_O|)}.$$

The conjunction of variables in the premise of a rule as well as the implications are realized by the product operator. The sum operator is taken to combine the rules (i.e., we use sum-prod-inference). For defuzzification, the height method is applied. The usage of rules with Gaussian premises is motivated by the *generalized central limit theorem*: Processes with multi-causal origination tend to be normally distributed.

Potential *feature selection* algorithms are described in [30]. In general, *filter* and *wrapper* approaches can be distinguished. Filter approaches decide on the selection of certain features by means of an analysis of the structure of the data (i.e. in the input space). Wrapper approaches take the classification performance of classifiers for different feature subsets into account. The problem of *model selection* for Neural Networks is discussed in [31] in greater detail. Usually, these techniques are categorized as being either *constructive* (growing techniques), *destructive* (pruning techniques), or *hybrid*. Constructive techniques start with small network structures and gradually add new neurons and connections. Destructive techniques go the other way: Initially large structures are pruned gradually. Hybrid techniques are iterative techniques that allow arbitrary search directions in each step.

In [32] we have shown how feature and model selection for conventional RBF networks can be realized by means of an evolutionary algorithm (EA). This EA could be adapted here. It selects appropriate features from a given set of possible features and optimizes the network architecture (e.g., number of rules / hidden neurons). A set of possible features must be extracted by means of an application-specific algorithm library. In general, it can be expected that wrapper approaches for feature selection and hybrid techniques for model selection may be slower but also yield better results in terms of classification performance.

The uncertainty that is considered here is the uncertainty of an input vector being classified correctly. For example: The uncertainty of a correct classification is high near the decision boundary, because a vector is mapped to different classes with almost equal degree of membership or probability.

B. Acquisition and Combination of Human-Driven Knowledge

CKD systems require an appropriate management component for human-driven knowledge that

1) acquires knowledge from single users,
2) combines their knowledge, and
3) answers queries.

The knowledge acquired from different users must be combined to obtain a more comprehensive knowledge base and to reduce uncertainties. That is, the quantity and the quality of the human-driven knowledge play a fundamental role.

Most work related to these tasks is based on symbolic approaches such as propositional logic, description logic, first-order logic, modal logic, production systems, or ontologies. Symbolic knowledge may ease the interaction with human experts, but is not able to deal with uncertain knowledge. If knowledge is assumed to be uncertain, a hybrid (symbolic / numeric) approach [6] is required which annotates symbolic knowledge with numerical values that quantify uncertainty. Relevant hybrid knowledge management solutions are Probabilistic Argumentation Systems (PAS) [33], [34], [35] and Markov Logic Networks (MLN) [36], [37]. PAS extend conventional argumentation systems (see, e.g. [38]) by processing uncertain logical assumptions which can contain terms annotated with probabilities. These uncertain terms can be used to evaluate symbolic arguments that either support or refute a hypothesis numerically. An MLN can be regarded as a knowledge base consisting of logical statements weighted with their probability. To effect inference, an MLN is transformed into a Markov Network and then an appropriate inference mechanism such as Markov Chain Monte Carlo, Gibbs Sampling, or Loopy Belief Propagation is applied. Comparing PAS and MLN we can state that both paradigms process symbolic and numeric knowledge by annotating logical statements with probabilities. While cyclic dependencies can be processed by both paradigms, only PAS are intended to use – in addition to probabilities – additional numerical criteria to evaluate the uncertainty of a hypothesis (support, possibility, doubt, etc).

In our CKD system, knowledge management will be based – as far as possible – on existing techniques based on a sound mathematical theory. To simplify knowledge acquisition we will restrict our work to causal relationships between concepts in the application domain (e.g., $friction \Rightarrow scratches$) which can be modeled using a subset of propositional logic. When acquiring knowledge from a single user we will check it for inconsistencies using a simple model checker (e.g., based on

Binary Decision Diagrams [39]) and, thus, prevent inconsistent knowledge to be entered into a user-specific knowledge base. When considering the knowledge of a single user, this approach reduces the knowledge revision problem to a simple knowledge update problem, because consistent knowledge can simply be added. Inconsistencies will be presented to the user and resolved manually.

The knowledge of different users will be combined into a single global knowledge base with an hybrid (symbolic / numeric) knowledge representation scheme. This approach has several advantages: We are able to evaluate the uncertainty of knowledge numerically and this evaluation will not result from an error prone process (such as users who enter numerical values) but rather from a combination of knowledge from different users. Here, Dempster-Shafer (DS) theory is a natural choice. DS theory can be seen as an extension of traditional probability theory [40] which has several advantages within the context of CKD systems: DS theory

1) offers a combination rule to combine knowledge from different sources,
2) is able to process cyclic dependencies (e.g., $A \Rightarrow B$, $B \Rightarrow C$, $C \Rightarrow A$), and
3) is not only able to evaluate the validity of an hypothesis, but also the amount of absent information (referred to as degree of ignorance).

Based on Dempster-Shafer theory, uncertainty will be modeled by so-called potentials. Potentials model the belief (also referred to as support), plausibility, and ignorance of a hypothesis and can be interpreted as shown in Table I. From the viewpoint of probability theory, belief and plausibility can be interpreted as a lower (belief) and upper (plausibility) probability bound. Ignorance is specified as the distance between belief and plausibility and, thus, can be interpreted as the amount of absent information. Accordingly, a belief of zero and a plausibility of one model total ignorance when no supporting or refuting information is available at all. Therefore, in our CKD system the ignorance of a causal relationship must decrease with the number of different users from which consistent knowledge about the relationship was acquired.

TABLE I

SEMANTICS OF POTENTIALS.

Degree of Belief	Degree of Plausibility	Interpretation
low	high	high degree of ignorance; few supporting and refuting knowledge
low	low	low degree of ignorance; few supporting knowledge and much refuting knowledge
high	high	low degree of ignorance; much supporting knowledge and few refuting knowledge
medium	medium	low degree of ignorance; much supporting and refuting knowledge

To process the uncertain knowledge stored in our global knowledge base we use ABEL [41] – an implementation of a PAS – as inference mechanism. ABEL uses so-called assumptions (statements expressed in propositional logic) as input which can contain propositional variables annotated with a probability value modeling their uncertainty. To integrate ABEL into our CKD system, an appropriate mechanism is needed which transforms the statements stored in our global knowledge base by mapping their potential to a probability value. Based on the interpretation of belief and plausibility as probability bounds, an appropriate probability between the upper and lower probability bound must be selected. For example, the probability value could be chosen depending on the application of the CKD system. A probability near the lower bound or the upper bound corresponds to a pessimistic (high-risk applications) or optimistic (fault-tolerant applications) decision, respectively. If only few information is available about an assumption (corresponds to an high ignorance) then the assumption is not entered into the PAS and, thus, not available for inference.

The uncertainty that is considered here is uncertainty regarding the validity of relationships between concepts in the application domain. For example: If many experts provide information about a relationship that is contradictory (e.g., $A \Rightarrow B$ and $A \not\Rightarrow B$) then the validity is uncertain (i.e., modeled with a belief and plausibility near 0.5). In the case that only very few users provide information about a relationship then information is considered being absent (e.g., according to DS theory modeled by a high ignorance) and a certain decision about the validity of the relationship cannot be made.

C. Fusion of Human-Driven and Data-Driven Knowledge

A central component of our CKD system is responsible for the fusion of human-driven and data-driven knowledge. This fusion component can be seen as an interface that enables the system to use human-driven knowledge in order to analyze, verify, and validate data-driven knowledge and vice versa. In a first step, we will focus on analyzing data-driven knowledge (fuzzy classification rules) by using human-driven knowledge (causal relationships between concepts in the application domain). To solve this problem, the fuzzy variables and terms of the fuzzy classification rule are mapped onto the corresponding concepts entered by human experts. A human-driven knowledge base fused with a fuzzy classification rule is set out in Figure 7.

D. Interestingness and Explication of Data-Driven Knowledge

One important feature of a CKD system is the evaluation of the *interestingness* of data-driven knowledge. In our case, this evaluation is not restricted to some statistical measures. Rather, it can benefit from the knowledge of human domain experts. New opportunities arise with the fusion of data-driven and human-driven knowledge which allows a better and more detailed analysis.

In a first approach human-driven knowledge can be utilized as follows to evaluate the interestingness of a fuzzy classification rule: The dependency represented by the fuzzy rule is *novel* if no or only few human-driven knowledge exists about this dependency in the global knowledge base. It can

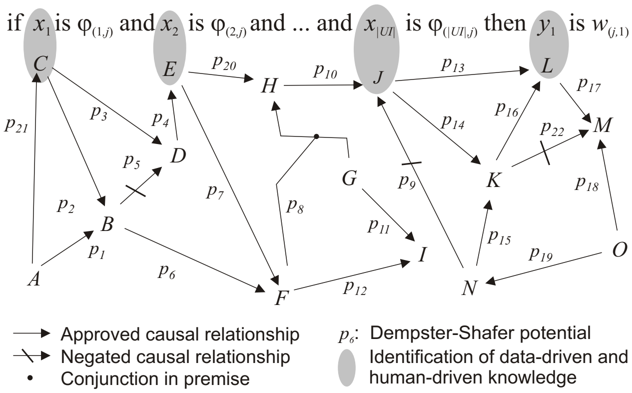

Fig. 7. Example of a fused knowledge base.

be regarded as *unexpected* if knowledge exists in the global knowledge base which implies a different class than the one specified by the fuzzy classification rule. In the case that a dependency is unexpected or novel and related to concepts marked as useful by an expert – because they will allow him to initiate valuable actions to improve product quality, for instance – than this knowledge can be referred to as *useful*. A further step would be to consider – in addition to the global knowledge base – the user-specific knowledge bases to tailor the evaluation to the needs and expertise of specific users.

In our CKD system, the interestingness of knowledge can be evaluated adopting the DS theory as follows: A dependency expressed by a fuzzy classification rule is unexpected if few knowledge exists in favor and much against (low degree of belief and plausibility) at the side of the human-driven knowledge. It can be referred to as novel if few knowledge exists in favor and against (low degree of belief, high degree of plausibility, and, thus, a high degree of ignorance). In the case that a dependency is unexpected or novel and also marked as useful by an expert then it can be referred to as useful.

In order to explain data-driven knowledge by means of human-driven knowledge (e.g., "Friction generates scratches because of abrasion from mechanical parts within a machine tool."), similar techniques can be used.

Both features – interestingness assessment and explication of data-driven knowledge – can be realized by means of the inference mechanisms provided by ABEL.

E. Proactive Behavior

The bottleneck of any knowledge based system is knowledge acquisition. Thus, knowledge acquisition has to be effected as efficient as possible to exploit the advantages of a CKD system. Including multiple human experts – which is a central aspect of a CKD system – is one approach to expand the bottleneck. Another approach is a proactive behavior of the CKD system. This means, that the CKD system must be able to identify required knowledge and to adapt the knowledge acquisition process accordingly. The prerequisite of proactive behavior is a *self-awareness* mechanism – in our case the self-contained ability to identify uncertain (missing, inconsistent, etc.) knowledge. *Environment-awareness* – in our case the self-contained ability to identify the needs and expertise of human experts – can further enhance the knowledge acquisition process.

At the side of data-driven knowledge conventional classifiers can be extended to autonomously detect regions in their input space where classifications cannot be made with a satisfying reliability. These so-called active learners can enhance the supervised learning process by actively selecting unlabeled samples which are labeled by human experts. Related research in the field of active learning is mostly concerned with support vector machines (e.g., [42]). This research shows that an active learner can rigorously reduce the amount of labeled samples which are required to achieve a classification rate which is equal or even higher than the classification rate of a conventional classifier.

At the side of human-driven knowledge uncertain knowledge about causal relationships can be identified easily by examining the associated potentials. The challenge here is to identify valuable knowledge which can be applied to better explicate or rate the interestingness of data-driven knowledge. Therefore, human-driven knowledge about the relationship between concepts in the premise and conclusion of fuzzy classification rules must be acquired actively. A suitable learning approach must decide whether to acquire new knowledge or consolidate existing knowledge by asking additional experts. Therefore, an active learning approach must be able to rate causal relationships according to their utility to explicate or rate interesting data-driven knowledge. Related research about actively acquiring human-driven knowledge exists in the field of commonsense reasoning (e.g., [43], [44]). Research on rating the utility of human-driven knowledge for analyzing data-driven knowledge is not known to us.

IV. Summary and Outlook

In this article, we have shown how valuable (certain, comprehensive, interesting) knowledge about an application can be acquired with a collaborative approach. CKD systems go even beyond this point: They gain experience about an application area (see Figure 8) in order to explain knowledge and to assess its interestingness for the application expert.

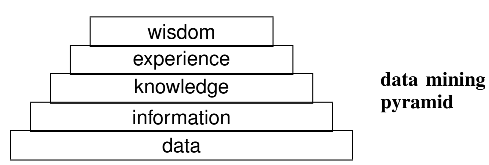

Fig. 8. Data mining wisdom pyramid (adopted from [2]).

The techniques outlined in this article could be used in various applications. One has already been mentioned: Product quality improvement in a wafer production process where several experts are involved. Another example would be the detection of rare errors in car electronics where test drivers, electronic engineers, and software analysts cooperate. Other examples could be given where process improvement is an important issue, e.g., in mechanical or chemical engineering. In all those technical applications, the human experts are highly motivated to be involved in such a CKD process because they greatly profit from that cooperation.

References

[1] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge discovery and data mining: Towards a unifying framework," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD 1996),* Portland, 1996, pp. 82–88.

[2] M. J. Embrechts, B. Szymanski, and K. Sternickel, "Introduction to scientific data mining: Direct kernel methods & applications," in *Computationally Intelligent Hybrid Systems: The Fusion of Soft Computing and Hard Computing*, S. J. Ovaska, Ed. John Wiley & Sons, Chichester, New York, 2004, ch. 10, pp. 317–362.

[3] G. Görz, *Handbuch der Künstlichen Intelligenz*. Oldenbourg Verlag, München, 2000.

[4] J. Y. Halpern, *Reasoning about uncertainty*. MIT Press, Cambridge, 2003.

[5] A. Motro and P. Smets, Eds., *Uncertainty Management in Information Systems – From Needs to Solutions*. Springer Verlag, London, UK, 1997.

[6] S. Parsons and A. Hunter, "A review of uncertainty handling formalisms," *Lecture Notes in Computer Science*, vol. 1455, pp. 8–38, 1998.

[7] L. A. Zadeh, "Toward a generalized theory of uncertainty (GTU): an outline," *Information Sciences – Informatics and Computer Science: An International Journal*, vol. 172, no. 1–2, pp. 1–40, 2005.

[8] A. Endres, "Forschungs- und Entwicklungsprojekte: Wissen bei Menschen und Maschinen – eine informatikbezogene Betrachtung." *Informatik, Forschung und Entwicklung*, vol. 18, no. 3–4, pp. 201–206, 2004.

[9] B. Berendt, A. Hotho, and G. Stumme, "Towards semantic web mining," in *Proceedings of the First International Semantic Web Conference on The Semantic Web (ISWC 2002),* Sardinia, 2002, pp. 264–278.

[10] "ECML/PKDD-2004 knowledge discovery and ontologies workshop (KDO 2004)," http://olp.dfki.de/pkdd04/cfp.htm, last visited: January 8th 2006.

[11] "ECML/PKDD-2005 knowledge discovery and ontologies workshop (KDO 2005)," https://webhosting.vse.cz/svatek/KDO05/, last visited: January 8th 2006.

[12] Z. Nazeri and E. Bloedorn, "Exploiting available domain knowledge to improve mining aviation safety and network security data," http://lp.dfki.de/pkdd04/nazeri-final.pdf, last visited: January 8th 2006.

[13] V. Svatek, J. Rauch, and M. Flek, "Ontology-based explanation of discovered associations in the domain of social reality," http://olp.dfki.de/pkdd04/vanzin-final.pdf, last visited: January 8th 2006.

[14] J. Adibi, H. Chalupsky, E. Melz, and A. Valente, "The KOJAK group finder: Connecting the dots via integrated knowledge-based and statistical reasoning," in *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence (AAAI/IAAI 2004),* San Jose, 2004, pp. 800–807.

[15] C. Pohle, "Integrating and updating domain knowledge with data mining," in *PhD Workshop co-located with the Conference on Very Large Databases (VLDB 2003),* Berlin, 2003.

[16] ——, "Integrating domain knowledge for data mining post-processing," in *Lernen, Wissensentdeckung und Adaptivität (LWA 2004),* Berlin, 2004, pp. 76–83.

[17] S. Bloehdorn, P. Cimiano, A. Hotho, and S. Staab, "An ontology-based framework for text mining," *Zeitschrift für Computerlinguistik und Sprachtechnologie*, vol. 20, pp. 87–112, 2005.

[18] T. Froehner, M. Nickles, and G. Weiß, "Open ontologies – the need for modeling heterogeneous knowledge," in *Proceedings of the International Conference on Information and Knowledge Engineering (IKE 2004),* Las Vegas, 2004, pp. 91–97.

[19] J. Phillips and B. G. Buchanan, "Ontology-guided knowledge discovery in databases," in *Proceedings of the 1st International Conference on Knowledge Capture (K-CAP 2001),* Victoria, 2001, pp. 123–130.

[20] M. Vanzin and K. Becker, "Exploiting knowledge representation for pattern interpretation," https://webhosting.vse.cz/svatek/KDO05/paper10.pdf, last visited: January 8th 2006.

[21] A. S. d'Avila Garcez, K. Broda, and D. M. Gabbay, "Symbolic knowledge extraction from trained neural networks: A sound approach," *Artificial Intelligence*, vol. 125, no. 1–2, pp. 155–207, 2001.

[22] A. Darbari, "Rule extraction from trained ANN: A survey," Department of Computer Science, Dresden University of Technology, Dresden, Tech. Rep. WV-2000-03, 2000.

[23] A. A. Freitas, *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer Verlag, Berlin, Heidelberg, New York, 2002.

[24] A. Piegat, *Fuzzy Modeling and Control*. Physica-Verlag, Heidelberg, New York, 2001.

[25] J.-S. Roger Jang and C.-T. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems," *IEEE Transactions on Neural Networks*, vol. 4, no. 1, pp. 156–159, 1993.

[26] H. C. Andersen, A. Lotfi, and L. C. Westphal, "Comments on "functional equivalence between radial basis function networks and fuzzy inference systems"," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1529–1531, 1998.

[27] J.-S. Roger Jang, "Author's reply," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1531–1532, 1998.

[28] O. Buchtala and B. Sick, "Techniques for the fusion of symbolic rules in distributed organic systems," in *Proceedings of the 2006 IEEE Mountain Workshop on Adaptive and Learning Systems (IEEE SMCals 2006),* Logan, 2006, pp. 85–90.

[29] O. Buchtala, A. Hofmann, and B. Sick, "Fast and efficient training of RBF networks," in *Artificial Neural Networks and Neural Information Processing – ICANN/ICONIP 2003 (Joint International Conference ICANN/ICONIP 2003),* Istanbul, ser. Lecture Notes in Computer Science, O. Kaynak, E. Alpaydin, E. Oja, and L. Xu, Eds., no. 2714. Springer Verlag, Berlin, Heidelberg, New York, 2003, pp. 43–51.

[30] H. Liu and H. Motoda, Eds., *Feature Extraction, Construction, and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, Boston, 1998.

[31] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[32] O. Buchtala, M. Klimek, and B. Sick, "Evolutionary optimization of radial basis function classifiers for data mining applications," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 35, no. 5, pp. 928–947, 2005.

[33] B. Anrig, R. Bissig, R. Haenni, J. Kohlas, and Lehmann, "Argumentation systems: Introduction to assumption-based modeling with ABEL," University of Fribourg, Department of Theoretical Computer Science, Tech. Rep., 1999.

[34] R. Haenni, "Modelling uncertainty with propositional assumption-based systems," in *Applications of Uncertainty Formalisms*, 1998, pp. 446–470.

[35] R. Haenni and N. Lehmann, "Probabilistic argumentation systems: a new perspective on Dempster-Shafer theory," *International Journal of Intelligent Systems*, vol. 18, no. 1, pp. 93–106, 2003.

[36] S. Kok and P. Domingos, "Learning the structure of Markov logic networks," in *Proceedings of the 22nd International Conference on Machine Learning (ICML 2005),* Bonn, 2005, pp. 441–448.

[37] M. Richardson and P. Domingos, "Markov logic networks," *Journal of Machine Learning*, vol. 62, no. 1–2, pp. 107–136, 2006.

[38] K. D. Forbus and J. D. Kleer, *Building problem solvers*. MIT Press, Cambridge, 1993.

[39] P. Molitor and C. Scholl, *Datenstrukturen und effiziente Algorithmen für die Logiksynthese kombinatorischer Schaltungen*. Teubner Verlag, Stuttgart, 1999.

[40] C. Beierle and G. Kern-Isberner, *Methoden wissensbasierter Systeme*, 2nd ed. Vieweg Verlag, Braunschweig, 2003.

[41] R. Haenni, "Argumentative reasoning with ABEL," in *Proceedings of the European Conference on Logics in Artificial Intelligence (JELIA 2002),* Cosenza, 2002, pp. 505–508.

[42] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *The Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2002.

[43] H. Lieberman and H. Liu, "Adaptive linking between text and photos using common sense reasoning," in *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2002),* Malaga, 2002, pp. 2–11.

[44] T. Chklovski, "Learner: a system for acquiring commonsense knowledge by analogy," in *Proceedings of the 2nd international conference on Knowledge capture (K-CAP 2003),* Sanibel Island, 2003, pp. 4–12.