International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems

Vol. 11 (Suppl. Issue Nov.), 25-49, 2003

Incremental Development of Diagnostic Set-Covering Models with Therapy Effects

Joachim Baumeister
baumeister@informatik.uni-wuerzburg.de

Dietmar Seipel
seipel@informatik.uni-wuerzburg.de

Frank Puppe
puppe@informatik.uni-wuerzburg.de

Department of Computer Science
University of Würzburg
Würzburg, Germany

Received September 2002
Revised February 2003

Although a lot of work in the field of knowledge acquisition has been done, the manual development of diagnostic knowledge systems by domain experts still is a very complex task. In this paper we will present an incremental approach for building diagnostic systems based on set-covering models. We start with a simple model describing the coarse structure between diagnoses and findings. Subsequently, this simple model can be enhanced by similarities, weights and probabilities to increase the accuracy of the knowledge and the resulting system. We will also show how these static set-covering models can be combined with dynamic set-covering models including higher level knowledge about causation effects. We will motivate how dynamic set-covering models can be used for implementing diagnostic systems including therapy effects. Finally, we report on two practical applications dealing with set-covering models from the geo-ecological and from the medical domain, respectively, that we have implemented.

Keywords: set-covering models; model-based diagnosis; abductive reasoning; applied uncertainty; applied causality

1. Introduction

This paper focuses on the problem of developing knowledge based diagnostic systems. The work was motivated by the experience[1] we have made with experts, who tried to build a complex diagnosis system from scratch. The available data samples were not large enough for using machine learning algorithms, and therefore we chose a manual development approach. In the first step we tried probabilistic models to represent the implied uncertainty of the domain. But even simplifications of models

did not help to cope with the complexity of the structure (e.g., handle numerical probabilities and conditional independence assumptions).

For this reason we developed an inverse approach: Starting with an extremely simple model the expert is able to formulate the basic structure of the diagnostic problem in a qualitative manner. For improving the quality he/she is able to extend this model by qualitative or/and quantitative knowledge without loosing the fundamental relations of the simple model. Our representation is able to handle diagnostic problems including the accounting of therapy effects, which can change observed findings but keep the original diagnosis in focus.
A *set-covering model* consists of a set of diagnoses (solutions), a set of findings (observations) and set-covering relations between the elements of these two sets. There exists a set-covering relation between a diagnosis and a finding, iff the diagnosis forces the observation of the finding. The basic idea of set-covering diagnosis is the detection of a reasonable set of diagnoses which can explain the given observations. For this task we propose an abductive reasoning step: Firstly, hypotheses are generated in order to explain the given observations. In a second step, we define a *quality measure* for ranking competing hypotheses.

Abductive reasoning with set-covering models has got a long tradition in diagnostic reasoning: One of the earliest approaches might be Patil's system ABEL[2], which describes abnormal behavior models with multi-level nets. Edges between the state nodes can describe causal, associative or grouping knowledge. However, ABEL cannot represent uncertain information about causal relationships. The assessment of a diagnosis is defined by the completeness with which it can explain the observations. Another significant approach are the set-covering models defined by Reggia et al.[3]. In [4] they introduce numeric probabilities to describe covering relations in more detail and discuss a transformation to the Theorem of Bayes. Similarly, Long[5] uses probabilistic covering models related to Bayesian networks, but allows the use of forward-loops and conditional links. The system MOLE[6] implements a similar covering-and-differentiate method, which solves the diagnostic task by first proposing candidates that will cover findings specified by the user and then trying to obtain more information that will differentiate the candidates. The system uses ordinal preferences instead of numerical probability measures for ranking the competing hypotheses. Recently, Lucas et. al.[7] have presented a diagnostic system for reprogramming pacemakers using a covering model with Horn formulas. Uncertainty is represented by an assumption literal in the precondition of the formula. But there is no qualifying assessment of the competing hypotheses, since a hypothesis can be either suggested or confirmed. Another interesting aspect is the combination of set-covering knowledge with other problem-solving methods in order to allow for knowledge reuse. Therefore Puppe[8] presents an inference structure for diagnostic problem solving, which integrates set-covering knowledge into other formalisms like heuristic rules or decision trees.
All of these approaches have one major shortcoming: They only provide the evalu-

ation of existing and implemented covering relations between diagnoses and observations. Nevertheless it would be interesting to include therapies into set-covering models. Therapies combined with diagnostic covering models can change the observations in many ways. Also the handling of severities of diagnoses is not possible in the systems mentioned above, although it has been shown that they are helpful for real world applications. In this paper we will give an introduction to a new interpretation of set-covering models, which allows for the diagnosis of observations including therapy effects. We will show, how to build such models in an incremental fashion in order to provide a minimum effort of knowledge acquisition.

The rest of the paper is organized as follows: In Section 2 we will give a short formal introduction to set-covering models and to the ideas of reasoning with these models. We will motivate the idea of having the facility to build set-covering models in an incremental way. Beginning with the simple set-covering model in Section 3 we will focus on the knowledge enhancements. We will show, how to incrementally apply similarities, weights, and uncertainty. Beyond these *static* models we will introduce *dynamic* models using causal set-covering relations in Section 4. In Section 5 we report on two applications which we implemented using the set-covering approach. Section 6 summarizes the work presented in this paper and gives a short survey of further extensions we want to consider in the future.

2. A Framework for Diagnosis with Set-Covering Models

A set-covering model consists of set-covering relations of the following form:

> If a diagnosis D is true, then the parameters (attributes) $A_1, \ldots, A_n$ are observed with corresponding values $v_1, \ldots, v_n$.

A single set-covering relation r is denoted by $r = D \rightarrow A{:}v$; we call the assignment $A{:}v$ of a value v to a parameter A a *finding*, and we say that the finding $A{:}v$ is *covered* by the diagnosis D.
For example, Figure 1 shows a set-covering model $\mathcal{R}$ with 5 set-covering relations for two diagnoses D_1 and D_2. An edge from a diagnosis D to a finding $A{:}v$ with the label r indicates a set-covering relation $r = D \rightarrow A{:}v$.

The basic algorithm for *set-covering diagnosis* is very simple: Given a set of observed findings, it uses a hypothesize-and-test strategy, which picks a hypothesis (coined from diagnoses) in the first step and tests it against the given observations in a second step. The test is defined by calculating a quality measure, which expresses the covering degree of the hypothesis regarding the observed findings. The generation and evaluation of the hypotheses is an iterative process, which stops when a satisfying hypothesis has been found or all hypotheses have been considered. Normally the algorithm will look at single diagnoses, compute the corresponding quality measure, and then it will generate hypotheses with multiple diagnoses, if needed. In principle the detection of the most suitable hypothesis will be similar to a search

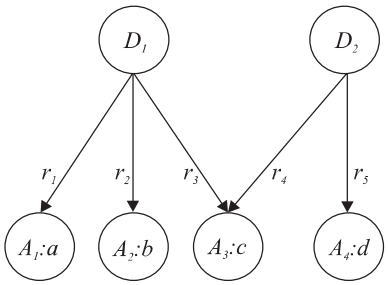

Fig. 1. Set-covering model for the diagnoses D_1 and D_2, and the parameters A_1, A_2, A_3 and A_4.

in an exponentially large search space, since there are 2^n possible hypotheses for n diagnoses. However the focus of this paper is the evaluation of a given hypothesis on the basis of a quality measure (see Section 4.3 for a short discussion).

2.1. *Basic Definitions and Notations*

In the following we define necessary sets that are considered for weight functions of quality measures.
The basic sets for this task are the following: We define $\Omega_{\mathcal{D}}$ to be the universe of all diagnoses and $\Omega_{\mathcal{A}}$ to be the universe of all observable parameters (attributes). To each parameter $A \in \Omega_{\mathcal{A}}$ a range $dom(A)$ of values is assigned, and $\Omega_{\mathcal{V}} = \bigcup_{A \in \Omega_{\mathcal{A}}} dom(A)$ is the set of all possible values for the parameters.

Definition 1. If a parameter $A \in \Omega_{\mathcal{A}}$ is assigned to a value $v \in dom(A)$, then we call $A{:}v$ a *finding*. The universe of all possible findings is defined as

$$\Omega_{\mathcal{F}} = \left\{ A{:}v \,\middle|\, A \in \Omega_{\mathcal{A}},\, v \in dom(A) \right\}.$$

A set of findings $\mathcal{F}$ is called *functional*, iff for all parameters $A \in \Omega_{\mathcal{A}}$ there exists at most one finding $A{:}v \in \mathcal{F}$, i.e., with A in the first component. In this case we can define the partial function $val_{\mathcal{F}} : \Omega_{\mathcal{A}} \to \Omega_{\mathcal{V}}$ which returns the assigned value v of a given parameter $A \in \Omega_{\mathcal{A}}$: $val_{\mathcal{F}}(A) = v$, if $A{:}v \in \mathcal{F}$, and $val_{\mathcal{F}}(A) = \perp$, otherwise. Furthermore, we define $a(\mathcal{F}) = \{ A \in \Omega_{\mathcal{A}} \mid \exists\, A{:}v \in \mathcal{F} \}$ to be the set of parameters in a set $\mathcal{F}$ of findings.

Definition 2. A *set-covering relation* r between a diagnosis D and a finding F is denoted by $r = D \to F$. Then, we say that "D covers F". $\Omega_{\mathcal{R}}$ denotes the universe of all set-covering relations. A *set-covering model* is a set $\mathcal{R} \subseteq \Omega_{\mathcal{R}}$ of set-covering relations.

Definition 3. We call a set $\mathcal{H} \subseteq \Omega_{\mathcal{D}}$ of diagnoses a *hypothesis*. A hypothesis $\mathcal{H} = \{ D_1, \ldots, D_n \}$ can be interpreted as a conjunction $D_1 \wedge \ldots \wedge D_n$ of diagnoses, which tries to explain a given observation.

Given a set $\mathcal{F}_\mathcal{O} \subset \Omega_\mathcal{F}$ of *observed findings*. The goal is to find a hypothesis $\mathcal{H}$ which is able to explain the observed findings. This is the case if $\mathcal{H}$ covers all observed findings, where a finding F is *covered* by $\mathcal{H}$, iff F is covered by at least one diagnosis $D \in \mathcal{H}$. If a finding is not covered by $\mathcal{H}$, then it is called *isolated* w.r.t. $\mathcal{H}$; the set of all observed findings that are not covered by (isolated w.r.t.) $\mathcal{H}$ will be denoted by $\mathcal{F}_\mathcal{O}^{\neg\mathcal{H}}$.

Definition 4. Given a set-covering model $\mathcal{R}$, a diagnosis $D \in \Omega_\mathcal{D}$, and a hypothesis $\mathcal{H} \subseteq \Omega_\mathcal{D}$. Then $\mathcal{F}_D = \{\, F \in \Omega_\mathcal{F} \mid D \rightarrow F \in \mathcal{R} \,\}$ is the set of all findings that are covered by D.

We select a functional set $\mathcal{F}_\mathcal{H} \subseteq \bigcup_{D\in\mathcal{H}} \mathcal{F}_D$ of findings that are covered by $\mathcal{H}$, i.e., $\mathcal{F}_\mathcal{H}$ should only contain one finding for each parameter A. If two diagnoses in $\mathcal{H}$ cover the same parameter A with different values, then $\bigcup_{D\in\mathcal{H}} \mathcal{F}_D$ is not functional. In this case, we select one of the conflicting findings $A{:}v \in \bigcup_{D\in\mathcal{H}} \mathcal{F}_D$ as follows:

(1) If one of the conflicting findings is contained in $\mathcal{F}_\mathcal{O}$, then we select it.
(2) If probabilities are defined (see Section 3.3), then we select the finding with the most probable set-covering relation in $\mathcal{F}_\mathcal{H}$.
(3) Otherwise, we randomly select one finding.

We remark, that additional knowledge can enhance the conflict resolution strategy. E.g., in [9] abnormality information was used to select the most abnormal parameter value. If abnormality information for parameter values is defined in the model, then we can apply this easily for conflict resolution.
Given a set $\mathcal{F}_\mathcal{O}$ of observed findings and a set $\mathcal{F}_\mathcal{H}$ of predicted findings. We say that a parameter $A \in \Omega_\mathcal{A}$ is observed, if there exists a finding $A{:}v \in \mathcal{F}_\mathcal{O}$ for A. For a comparison between $\mathcal{F}_\mathcal{H}$ and $\mathcal{F}_\mathcal{O}$ we introduce two subsets of $\mathcal{F}_\mathcal{H}$:

(1) The set $\mathcal{F}_{\mathcal{H},\mathcal{O}}$ of *parametrically predicted findings* consists of all predicted findings $A{:}v \in \mathcal{F}_\mathcal{H}$ for which the parameter A is observed.
(2) The set $\mathcal{F}^+_{\mathcal{H},\mathcal{O}} = \mathcal{F}_\mathcal{H} \cap \mathcal{F}_\mathcal{O}$ of *positively predicted findings* consists of all predicted findings which are observed with the predicted value.
Analogously, we define $\mathcal{F}^-_{\mathcal{H},\mathcal{O}} = \mathcal{F}_{\mathcal{H},\mathcal{O}} \setminus \mathcal{F}^+_{\mathcal{H},\mathcal{O}}$ to be the set of *negatively predicted findings*.

2.2. *Quality Measures*

In this paragraph we introduce quality measures for evaluating a hypothesis for a given set of observed findings. The quality measures are based on a binary difference function and on an unary weight function

$$\begin{array}{c} \| \,.,. \,\| : 2^{\Omega_\mathcal{F}} \times 2^{\Omega_\mathcal{F}} \rightarrow \quad , \\ \| \,. \,\| : 2^{\Omega_\mathcal{F}} \rightarrow \end{array}$$

for sets of findings. Given $\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O} \subseteq \Omega_\mathcal{F}$, then $\|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\|$ measures the difference between the predicted and the observed findings.

(1) In the simplest case, $\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = |\mathcal{F}^{+}_{\mathcal{H},\mathcal{O}}| = |\mathcal{F}_{\mathcal{H}} \cap \mathcal{F}_{\mathcal{O}}|$ is the number of findings, for which the predicted value is also observed.
(2) In Section 3 we will discuss more refined difference functions, which can assign different weights to the different parameters A, which can take into account the similarity between the predicted value v and the observed value v' of a parameter A, and which can deal with probabilities.

Given $\mathcal{F} \subseteq \Omega_{\mathcal{F}}$, then $\|\mathcal{F}\|$ is the weight of the set $\mathcal{F}$ of findings. In the simplest case, $\|\mathcal{F}\| = |\mathcal{F}|$ is the number of parameters A for which findings $A{:}v$ are given in $\mathcal{F}$. This means that all parameters A have the same weight $w(A) = 1$. In Section 3 we will assign different weights $w(A)$ to different parameters A.

Definition 5. For a given hypothesis $\mathcal{H}$ and a non-empty set of parametrically predicted findings $\mathcal{F}_{\mathcal{H},\mathcal{O}} \neq \emptyset$ and a non-empty set of observed findings $\mathcal{F}_{\mathcal{O}} \neq \emptyset$, we define

(1) The *precision* $\pi(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ relates the predicted findings, for which the values are as observed, to the weight of all parametrically predicted findings:

$$\pi(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \frac{\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\|}{\|\mathcal{F}_{\mathcal{H},\mathcal{O}}\|} \tag{2}$$

(2) The *covering rate* $\kappa(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ relates the weight of the parametrically predicted findings to the weight of the observed findings:

$$\kappa(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \frac{\|\mathcal{F}_{\mathcal{H},\mathcal{O}}\|}{\|\mathcal{F}_{\mathcal{O}}\|} \tag{3}$$

(3) The *quality* $\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ of a hypothesis $\mathcal{H}$ is given by

$$\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \pi(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) \cdot \kappa(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \frac{\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\|}{\|\mathcal{F}_{\mathcal{O}}\|}. \tag{4}$$

The quality of a hypothesis is defined as the ratio of observed and predicted findings and the overall number of observed findings. Ideally, all observed findings are also predicted and we would obtain a quality value $\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = 1$. Sometimes it might be interesting to have more subtle measures than the quality value (e.g., for the candidate generation of hypotheses). Therefore we define the quality ϱ as a product of the covering rate κ and the precision π. The precision π is optimal, i.e., $\pi(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = 1$, if all predicted observations are assigned to the same values as noted in the covering relations. On the other hand the covering rate κ determines how many of the observed parameters are currently predicted by the given hypothesis and which therefore can explain the observation.

The quality measure is only considered meaningful, if the hypothesis is *sufficiently complete*: For this, the expert has to define a *completeness value* $cv(D) \in [0, 1]$ for each diagnosis D. Then, we call a hypothesis $\mathcal{H}$ sufficiently complete, if for each diagnosis $D \in \mathcal{H}$ the ratio between the number of observed parameters which

are covered by D and the number of parameters which are covered by D exceeds $cv(D)$:

$$\frac{|a(\mathcal{F}_D) \cap a(\mathcal{F}_\mathcal{O})|}{|a(\mathcal{F}_D)|} \geq cv(D) \tag{5}$$

This definition is motivated by the fact, that a covering model for a diagnosis will contain more findings than the diagnosis will cover on average. Nevertheless in most cases the observation of a percentage of the modeled findings will legitimate the validation of this diagnosis. Only if $cv(D)$ is reached by the observation set in the current case, D can be explained by a sufficiently large number of observed findings and therefore is considered as a valid subset of a hypothesis.

2.3. *Extensions*

The rest of the paper will concentrate on extensions of the simple set-covering model presented in Figure 1. Starting with simple set-covering relations we can apply additional information to improve the diagnostic quality. Each addition forms a supplementary component of the set-covering model. The extensions were motivated by real world problems in the area of *chronical polyarthritis* (pcP) and so we will use simplified medical examples for each single extension.

Suppose we have the covering relation "pcP → pain:very high" and we can observe "pain:high". In the simple covering model this observation would be negatively observed and be treated exactly like "pain:normal", although the values "very high" and "high" are much more similar than "very high" and "normal". As a consequence we will introduce similarities between findings as the first extension to set-covering models.

Another knowledge component adds weights for findings. For example, if diagnosis "pcP" covers the findings "morning stiffness:true" and "rheumatic blood parameter:high", then the "high blood" parameters may be more valuable to the diagnostic process than the parameter "morning stiffness". Because of this we will add (global) weights to parameters to emphasize their diagnostic power in comparison to other parameters. Though it might be more accurate to define local weights for each set-covering relation (diagnosis–finding relation), we took a trade-off between knowledge acquisition costs and diagnostic accuracy and decided to introduce global weights (we refer to Section 3.4 for a discussion).

Furthermore in most real world applications only uncertain knowledge is available. Uncertainty can be expressed through probabilistic covering relations, e.g., it *usually holds* that "pcP → pain:very high", which will be introduced as a single knowledge component.

Beyond that, one can build dynamic models by defining causal relationships between severities of diagnoses and parameter values. For example, if the disease "pcP" has the severities "normal", "weak", "strong", then we can easily model a set-covering relation to a parameter "pain" that increases the value of "pain" depending on the severity of "pcP".

Thereby all extensions can be combined with each other, but it is recommendable to start with similarities, weights and uncertain set-covering relations. To deal with more complex models including therapies one should add severities and effect relations. The following sections will introduce these concepts in more detail and show how the weight functions $\| \, . \, \|$ and $\| \, . \, , \, . \, \|$ will vary upon the available knowledge.

3. Static Set-Covering Models

In this section we will introduce the knowledge components available to static set-covering models. We will call these models *static* because the structure of the set-covering relations is specified by the expert in advance and will not change during the problem solving process.

Starting with the *simple covering model* the engineer is only able to define ordinary set-covering relations between diagnoses and findings. An example for a simple covering model is depicted in Figure 1. The functions are then defined as simple cardinality functions

$$\begin{aligned} \|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| &= |\mathcal{F}^{+}_{\mathcal{H},\mathcal{O}}| \, , \\ \|\mathcal{F}\| &= |\mathcal{F}| \, . \end{aligned}$$

Further extensions will build upon this model.

3.1. *Similarities between Parameter Values*

Consider a parameter A with the value range $dom(A) = \{\, no,\ si,\ mi,\ hi \,\}$, with the meanings normal (no), slightly increased (si), medium increased (mi), and heavily increased (hi), where A:hi is predicted. We clearly see that the observation A:mi deserves a better covering rate than the observation A:no. Nevertheless the simple covering rate considers both observations as negatively observed parameters and makes no difference between the similarities of the parameter values.
For this reason we want to introduce *similarities* as the first extension to set-covering models. We define the similarity function

$$sim : \Omega_{\mathcal{V}} \times \Omega_{\mathcal{V}} \rightarrow [0, 1]$$

to capture the similarity between two values assigned to the same parameter. The boundary value 0 means no similarity and the value 1 indicates two equal values. Cluster analysis problems consider the closely related concept of *distance functions* (c.f. [10]). We obtain the quality for models with similarities, when we define the functions $\| \, ., . \, \|$ and $\| \, . \, \|$ as follows:

$$\begin{aligned} \|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| &= \textstyle\sum_{A \in a(\mathcal{F}_{\mathcal{H}\ \mathcal{O}})} sim\big(val_{\mathcal{F}_{\mathcal{H}}}(A), val_{\mathcal{F}_{\mathcal{O}}}(A)\big) \\ \|\mathcal{F}\| &= |\mathcal{F}| \end{aligned}$$

As a special case, we obtain the simple covering rate, if we define $sim(v, v') = 0, \text{ if } v \neq v'$, and $sim(v, v') = 1, \text{ if } v = v'$, for $v, v' \in \Omega_{\mathcal{V}}$.

3.2. *Weighted Findings in Set-Covering Models*

The introduction of weights for parameters is another common generalization of the simple set-covering model. Given a weight function

$$w : \Omega_{\mathcal{A}} \to I\!N_+ \tag{6}$$

for parameters, we define

$$\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = \textstyle\sum_{A \in a(\mathcal{F}^+_{\mathcal{H}\ \mathcal{O}})} w(A)\,,$$
$$\|\mathcal{F}\| = \textstyle\sum_{A \in a(\mathcal{F})} w(A)\,.$$

If all findings have the same utility, i.e., $w(A) = 1$ for all $A \in \Omega_{\mathcal{A}}$, then the weighted set-covering model reduces to the simple set-covering model.

3.3. *Uncertainty in Set-Covering Relations*

An important enhancement of set-covering models is the possibility to state uncertain set-covering relations. Thereby the diagnosis causes observations of specified findings not in any case but only with a specified probability.
We explain the handling of uncertainty in set-covering relations by introducing events and causation events, which were inspired by Peng and Reggia[4].

Definition 6. For a set-covering model we define the following basic events:

(1) A *cause event* $D \in \Omega_{\mathcal{D}}$ is defined as the event, that a diagnosis D is present. This event contains no information about the presence/absence of any other diagnosis $D' \in \Omega_{\mathcal{D}} \setminus \{D\}$. We analogously define $\neg D$ as the event, that D is absent.
(2) An *effect event* $F \in \Omega_{\mathcal{F}}$ defines the event, that a finding F is present (observed). The effect event F contains no information about the presence/absence of any other finding $F' \in \Omega_{\mathcal{F}} \setminus \{F\}$. We analogously define $\neg F$ as the event, that F is absent (not observed). In this way, an observation set $\mathcal{F}_{\mathcal{O}} = \{F_1, \ldots, F_n\}$ can be interpreted as a conjunction of effect events $F_1 \wedge \ldots \wedge F_n$.
(3) A *causation event* $D \mapsto F$ denotes the event, that the finding $F \in \Omega_{\mathcal{F}}$ is actually observed and caused by the diagnosis $D \in \Omega_{\mathcal{D}}$. For a set-covering relation $r = D \to F$ the causation event $D \mapsto F$ is an instantiation of r. Analogously we define $\neg(D \mapsto F)$ as the absence of $D \mapsto F$.

We remark, that a causation event $D \mapsto F$ may still be false, even if the effect event F is true (observed) and the cause event D is true (hypothesized). In this case, F is caused by another diagnosis D' possibly contained in the hypothesis, too. On the other side, a true causation event $D \mapsto F$ implies the cause event D and the effect event F to be true.
As we mentioned before, we are able to attach uncertainty to set-covering relations, which are called *conditional causal probabilities*.

Definition 7. The *conditional causal probability* $P(D \mapsto F|D)$ is the probability, that $F \in \Omega_{\mathcal{F}}$ is caused by $D \in \Omega_{\mathcal{D}}$ utilizing the causation event $D \mapsto F$. The probability of the absence of a causation event $\neg(D \mapsto F)$ is defined as

$$P\big(\neg(D \mapsto F)|D\big) = 1 - P(D \mapsto F|D)\,.$$

When we attach uncertainty to set-covering models, we define conditional causal probabilities to set-covering relations.

Definition 8. Let $D \mapsto F$ be a causation event for $D \in \Omega_{\mathcal{D}}$ and $F \in \Omega_{\mathcal{F}}$. Then we call C a context of $D \mapsto F$, if C is a conjunction of arbitrary cause events and causation events other than $D \mapsto F$ and $\neg(D \mapsto F)$.

To facilitate such propositions and their integration in the computation of covering models we need to make the following assumptions.

Additional Knowledge and Assumptions

(1) *Independence of cause events:* For each diagnosis $D \in \Omega_{\mathcal{D}}$ the apriori probability $P(D) \in (0;1]$ is given and the cause event D is independent from any other cause event. As a consequence, the apriori probability for a hypothesis $\mathcal{H} = \{D_1, \ldots, D_n\}$ is the product of the single probabilities, i.e.

$$P(\mathcal{H}) = \prod_{D \in \mathcal{H}} P(D) \cdot \prod_{D' \in \Omega_{\mathcal{D}} \setminus \mathcal{H}} \big(1 - P(D')\big)\,.$$

We remark, that for a hypothesis $\mathcal{H}$ all diagnoses $D' \in \Omega_{\mathcal{D}} \setminus \mathcal{H}$ are assumed to be absent.

(2) *Independence of causation events:* For each covering relation $r = D \rightarrow F$ defined in the set-covering model the non-zero conditional causal probability $P(D \mapsto F|D) \in (0;1]$ is given. If a cause event $D \in \Omega_{\mathcal{D}}$ happens, then the causation event $D \mapsto F$ occurs independently of any context C with $P(D \wedge C) > 0$, i.e.,

$$P(D \mapsto F|D \wedge C) = P(D \mapsto F|D).$$

(3) *Completeness assumption:* For each parameter $A \in \Omega_{\mathcal{A}}$ there need to exist at least one set-covering relation $r = D \rightarrow A{:}v \in \mathcal{R}$ with $D \in \Omega_{\mathcal{D}}$ and $v \in dom(A)$.

It is easy to see, that a set-covering model applying uncertainty information requires $|\Omega_{\mathcal{D}}| + |\mathcal{R}|$ probabilities. With the completeness assumption we can guarantee, that there exists a causation event for any observed finding.
For handling multiple faults (i.e. a set of diagnoses is explaining an observation set), we have to consider *composite causation events*.

Definition 9. Let $F \in \Omega_{\mathcal{F}}$ be an observed finding and let $\mathcal{H} \subseteq \Omega_{\mathcal{D}}$ be a hypothesis. Then, $\mathcal{H} \mapsto F$ is defined as the *composite causation event*. The meaning of $\mathcal{H} \mapsto F$ is, that F is caused by at least one diagnosis $D \in \mathcal{H}$.

When computing the probability of a composite causation event $\mathcal{H}\mapsto F$ we have to take the disjunction of every subset of $\mathcal{H}$ for causing F into account.
For example, for the set-covering model given in Figure 1, the hypothesis $\mathcal{H} = \{D_1, D_2\}$, and the finding $F = A{:}v_3$ we want to compute the conditional probability of the composite causation event $\mathcal{H}\mapsto F$. Thus, we have to consider

$$\begin{aligned} P(\mathcal{H}\mapsto F|\mathcal{H}) &= P(D_1\neg D_2\mapsto F|\mathcal{H}) + P(\neg D_1 D_2\mapsto F|\mathcal{H}) + P(D_1 D_2\mapsto F|\mathcal{H}) = \\ &= P(D_1\mapsto F|\mathcal{H})\cdot\big(1 - P(D_2\mapsto F|\mathcal{H})\big) + \\ &\quad + P(D_2\mapsto F|\mathcal{H})\cdot\big(1 - P(D_1\mapsto F|\mathcal{H})\big) + \\ &\quad + P(D_1\mapsto F|\mathcal{H})\cdot P(D_2\mapsto F|\mathcal{H}) \end{aligned}$$

In general, the probability for a composite causation event is computed as follows:

$$P(\mathcal{H}\mapsto F|\mathcal{H}) = \sum_{\mathcal{H}'\subseteq\mathcal{H}\setminus\{\}} \Big(\prod_{D\in\mathcal{H}'} P(D\mapsto F|D)\cdot \prod_{D\in\mathcal{H}\setminus\mathcal{H}'} \big(1 - P(D\mapsto F|D)\big)\Big) \qquad (7)$$

Now we can define the quality measure for probabilistic set-covering models.

Probabilistic Set-Covering Models

The quality measure for a probabilistic set-covering model is defined given the following functions $\|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\|$ and $\|\mathcal{F}\|$:

$$\begin{aligned} \|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\| &= \textstyle\sum_{F\in\mathcal{F}_\mathcal{H}\ \mathcal{O}} P(\mathcal{H})\cdot P_{\mathcal{H},\mathcal{F}_\mathcal{O}}(F)\,, \\ \|\mathcal{F}\| &= |\mathcal{F}|. \end{aligned}$$

where function $P_{\mathcal{H},\mathcal{F}_\mathcal{O}}$ is defined as follows:

$$P_{\mathcal{H},\mathcal{F}_\mathcal{O}}(F) = \begin{cases} P(\mathcal{H}\mapsto F|\mathcal{H}) & \text{for } F\in\mathcal{F}_\mathcal{O}, \\ 1 - P(\mathcal{H}\mapsto F|\mathcal{H}) & \text{otherwise.} \end{cases}$$

For the expert these equations serve as an intuitive understanding of the model. A more appropriate procedure for handling uncertainty would be the introduction of a *leak-diagnosis*. A leak-diagnosis D_l captures the idea, that no model can be a complete view of the domain and that there are always other reasons that can cause a given finding. These "other reasons" are collected in the leak-diagnosis, which is categorically connected to all available findings. To shrink the emerging number of probabilities, we can assume a constant probability for all set-covering relations between the leak-diagnosis and a finding. If the model contains weights it is easy to see, that the leak probabilities can be adapted with respect to the weights. Large weights will decrease the leak probability whereas small weights increase the probability. As a consequence, for every hypothesis we have to consider the leak-diagnosis to be included in the hypothesis as well. It is easy to understand that with the usage of the leak-diagnosis there will be no isolated observed parameters because the leak-diagnosis holds covering relations to all findings by default. So $\mathcal{F}_\mathcal{O}^{\neg\mathcal{H}}$ will be empty for all $\mathcal{H}$ and all $\mathcal{F}_\mathcal{O}$.

3.4. *Combining the Different Components of Knowledge*

In the previous sections we introduced different components of knowledge to include in the covering model. Each component supplies an additional support for the calculation of the covering rate, if it is available in the given model. However, if one component does not appear, it cannot contribute to the quality of a hypothesis and therefore will not appear in the calculation. For this reason we will introduce the abbreviations *wc*, *sc*, and *pc*: if the corresponding knowledge is available, then we set

$$\begin{aligned} wc(A) &= w(A)\,, \\ sc(A) &= sim\big(val_{\mathcal{F}_\mathcal{H}}(A), val_{\mathcal{F}_\mathcal{O}}(A)\big)\,, \\ pc(A) &= P(\mathcal{H}) \cdot P_{\mathcal{H},\mathcal{F}_\mathcal{O}}\big(A{:}val_{\mathcal{F}_\mathcal{O}}(A)\big)\,, \end{aligned}$$

and otherwise we set $wc(A) = sc(A) = pc(A) = 1$, i.e., 1 is the default value. The quality of a model will be computed by the following weight functions $\| \,.\,,\, . \,\|$ and $\| \,.\, \|$:

$$\begin{aligned} \|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\| &= \textstyle\sum_{A \in a(\mathcal{F}_\mathcal{H}\ \mathcal{O})} wc(A) \cdot pc(A) \cdot sc(A)\,, \\ \|\mathcal{F}\| &= \textstyle\sum_{A \in a(\mathcal{F})} wc(A)\,. \end{aligned} \tag{8}$$

It is easy to see that these formulas are generalizations for the equations given above, when in each equation only one knowledge component is considered.

Independence Assumptions

Nevertheless, when combining different kinds of components we have to keep the following assumptions in mind:

(1) *Independence assumptions between similarities and weights:*
In our formalism the weight $w(A)$ for a parameter A is independent of the values $v \in dom(A)$. Commonly, we call this a *global weight* for a parameter. For some real life applications, however, working with *partial weights* that depend on individual parameter values or diagnoses is more appropriate. E.g., for the diagnostic process the finding "temperature:high" should have a much higher weight than the finding "temperature:normal".
In our reasoning process, if a finding $A{:}v_1$ is predicted, but a very unsimilar finding $A{:}v_2$ is observed, then the observation will still have a high contribution to the overall precision/quality of the related diagnosis. Predicted findings with more similar observations, but a lower weight can only contribute in a smaller way. As a trade-off between knowledge acquisition costs and diagnostic precision, we decided to choose the manual acquisition of global weights, but provide learning methods for partial weights, if an appropriate number of samples is available (cf. [9]).

(2) *Independence assumptions between probabilities and similarities:*
Corresponding to Equation 8, we propose a proportional contribution relative to the similarity and the probability. E.g., for a given probability for a predicted

finding $A{:}v_1$ we assume, that the contribution C of the observation $A{:}v_2$ is proportional to its similarity:

$$C_{\mathcal{H},\mathcal{F}_\mathcal{O}}(A{:}v_2) = P_{\mathcal{H},\mathcal{F}_\mathcal{O}}(A{:}v_1) \cdot sim(v_1, v_2).$$

This assumption creates a massive contradiction, if a diagnosis covers two different values $v_1, v_2 \in dom(A)$ of the same parameter A, which are very unsimilar but have similar probabilities (e.g., a disease is observed either for very young or for very old people). We solved this conflict by permitting multiple findings $A{:}v_i,\ 1 \leq i \leq k$, for the same parameter A to be covered by the same diagnosis (see below)

3.5. *Covering Multiple Parameter Values*

As explained before, in some cases a diagnosis needs to cover multiple findings, which are assigned from the same parameter. E.g., for simple set-covering models "high temp" or "very high temp" may be observed, when "flue" occurs. Since, the modeler does not want to omit one (alternative) observation we introduced XOR set-covering relations[11].
An XOR set-covering relation $r = D \rightarrow_{XOR} \{F_1, \ldots, F_n\}$ defines a disjunctive covering relation, where at most one finding $F_1, \ldots, F_n$ can be observed, if D is present.
XOR set-covering relations have to be applied carefully and for defining a relation $r = D \rightarrow_{XOR} \{F_1, \ldots, F_n\}$ the following assumption is made:

Probability assumption: If there are probabilities defined for each $F_i \in \{F_1, \ldots, F_n\}$, then it holds that:

$$\sum_{1 \leq i \leq n} P(D \mapsto F_i | D) \leq 1$$

For a more detailed definition and interpretation of extended set-covering relations we refer to [11].

3.6. *Characterization of the Best Matching Hypothesis*

Contrary to the set-covering theory by Reggia et al.[3] we are not primarily interested in a hypothesis, which can explain *all* observed findings.
In general, hypotheses are preferred, that predict the largest subset of observed findings, but do not predict too many unobserved findings. In the best case, the set of predicted findings equals the set of observed findings. Nevertheless, an additional diagnosis added to an existing hypothesis can worsen the quality measure, even if it predicts actually unexplained observations.
For a set-covering model with weight knowledge a hill-climbing search will focus on parameters with high weights. In a second step, the measure will try to find explaining diagnoses for the remaining findings with lower weights.

For competing diagnoses with comparable quality measures we have to consider the minimality criterion, i.e. the hypothesis with fewer diagnoses will be preferred. For set-covering models with uncertainty knowledge, apriori probabilities of diagnoses will support this criterion, since usually the combined apriori probability of smaller hypotheses will be higher than of a larger hypothesis.

3.7. *An Example with Different Components of Knowledge*

We will illustrate the computation of the quality measures for static set-covering models by an example. The set-covering model is given in Figure 2, and we can observe the findings

$$\mathcal{F}_{\mathcal{O}} = \{\, A_1{:}a, A_2{:}b, A_3{:}c', A_4{:}d \,\},$$

we assume that $sim(c, c') = 0.7$ and $P(D_1) = P(D_2) = 0.5$. It holds that the apriori probabilities for any hypothesis $P(\{D_1\}) = P(\{D_2\}) = P(\{D_1, D_2\}) = 0.25$.

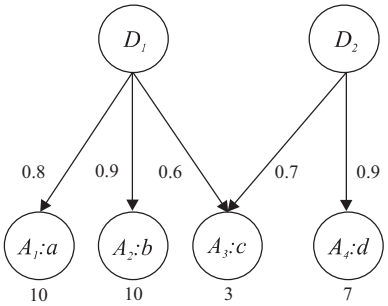

Fig. 2. Covering model for the example. Probabilities for the covering relations are shown at the corresponding edges, weights are depicted under the findings.

We will consider the three hypotheses

$$\mathcal{H}_1 = \{D_1\},\ \mathcal{H}_2 = \{D_2\}\,,\ \text{and}\ \ \mathcal{H}_3 = \{D_1, D_2\},$$

and we assume the completeness value $cv(D_1) = cv(D_2) = 0.5$. Then all $\mathcal{H}$ are sufficiently complete and are considered as possible explanations for $\mathcal{F}_{\mathcal{O}}$.
To measure the quality of the hypotheses we have to calculate $\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\|$ and $\|\mathcal{F}_{\mathcal{O}}\|$. We see that $\|\mathcal{F}_{\mathcal{O}}\|$ is constant for all hypotheses:

$$\|\mathcal{F}_{\mathcal{O}}\| = \sum_{A \in a(\mathcal{F}_{\mathcal{O}})} w(A) = 10 + 10 + 3 + 7 = 30$$

(1) For the hypothesis $\mathcal{H} = \mathcal{H}_1 = \{D_1\}$ we obtain

$$\begin{aligned} \mathcal{F}^{+}_{\mathcal{H},\mathcal{O}} &= \{\, A_1{:}a, A_2{:}b \,\}, \\ \mathcal{F}^{\neg\mathcal{H}}_{\mathcal{O}} &= \{\, A_4{:}d \,\}. \end{aligned}$$

Now the binary weight function is

$$\begin{aligned}\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| &= P(\mathcal{H}) \cdot \textstyle\sum_{A \in \{A_1, A_2, A_3\}} w(A) \cdot P_{\mathcal{H}, \mathcal{F}_{\mathcal{O}}}(A{:}v) \cdot \\ &\qquad \cdot sim\big(val_{\mathcal{F}_{\mathcal{H}}}(A), val_{Fo}(A)\big) = \\ &= P_{\mathcal{H}} \cdot \big(\underbrace{10 \cdot 0.8 \cdot 1.0}_{A_1} + \underbrace{10 \cdot 0.9 \cdot 1.0}_{A_2} + \underbrace{3 \cdot 0.6 \cdot 0.7}_{A_3}\big) = P(\mathcal{H}) \cdot 18.26\end{aligned}$$

Therefore $\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = P(\mathcal{H}) \cdot 0.61 = 0.15$.

(2) For the hypothesis $\mathcal{H} = \mathcal{H}_2 = \{D_2\}$ we obtain

$$\begin{aligned}\mathcal{F}^{+}_{\mathcal{H},\mathcal{O}} &= \{\, A_4{:}d \,\}, \\ \mathcal{F}^{-^{\mathcal{H}}}_{\mathcal{O}} &= \{\, A_1{:}a, A_2{:}b \,\}.\end{aligned}$$

Now the binary weight function is

$$\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = P(\mathcal{H}) \cdot \big(\underbrace{3 \cdot 0.7 \cdot 0.7}_{A_3} + \underbrace{7 \cdot 0.9 \cdot 1.0}_{A_4}\big) = P(\mathcal{H}) \cdot 7.77$$

Therefore $\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = P(\mathcal{H}) \cdot 0.26 = 0.07$.

(3) For the hypothesis $\mathcal{H} = \mathcal{H}_3 = \{D_1, D_2\}$ we obtain

$$\begin{aligned}\mathcal{F}^{+}_{\mathcal{H},\mathcal{O}} &= \{\, A_1{:}a, A_2{:}b, A_4{:}d \,\}, \\ \mathcal{F}^{-^{\mathcal{H}}}_{\mathcal{O}} &= \{\}.\end{aligned}$$

Since the parameter A_3 is covered by both diagnoses we cannot simply read off the probability for $F = A_3{:}c$ but have to compute the probability for the composite causation event $P(\mathcal{H} \mapsto F | \mathcal{H})$, as defined in Equation 7.

$$P(\mathcal{H} \mapsto F | \mathcal{H}) = \underbrace{0.6 \cdot 0.3}_{\mathcal{H}'=\{D_1\}} + \underbrace{0.4 \cdot 0.7}_{\mathcal{H}'=\{D_2\}} + \underbrace{0.6 \cdot 0.7}_{\mathcal{H}'=\{D_1,D_2\}} = 0.88$$

$$\begin{aligned}\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| &= P(\mathcal{H}) \cdot \big(\underbrace{10 \cdot 0.9 \cdot 1.0}_{A_1} + \underbrace{10 \cdot 0.8 \cdot 1.0}_{A_2} + \\ &\qquad + \underbrace{3 \cdot 0.88 \cdot 0.7}_{A_3} + \underbrace{7 \cdot 0.9 \cdot 1.0}_{A_4}\big) = \\ &= P(\mathcal{H}) \cdot 25.15\end{aligned}$$

Therefore $\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = P(\mathcal{H}) \cdot 0.84 = 0.21$.

It is easy to see that Hypothesis 3, $\mathcal{H}_3 = \{\, D_1, D_2 \,\}$, supplies the best covering for the given observation $\mathcal{F}_{\mathcal{O}} = \{\, A_1{:}a, A_2{:}b, A_3{:}c', A_4{:}d \,\}$.

4. Dynamic Set-Covering Models

The previous sections considered set-covering relations as static representations of diagnostic models. Now we will introduce a concept of cause-effect relations between diagnoses and parameters. This approach will allow for integrating therapies and their effects into set-covering models.

Up to now we can only represent set-covering relations for a single severity of a diagnosis. Moreover, it is impossible to model *therapy effects* on parameter values,

since therapy effects weaken or strengthen existing parameter values. This behavior cannot be modelled with the static set-covering models described so far.
For this reason we introduce *dynamic set-covering relations* $D \rightarrow A$ with *severities* as an extension of the existing covering models. A diagnosis now can not only be present or absent, but it can have a severity $s \in \Omega_{sev}$, where

$$\Omega_{sev} = \{0, s_1, \ldots, s_n, 1\} \subseteq [0,1]$$

Severity 0 means that the diagnosis is absent, the severity s_i denote the graded presence of the diagnosis, and severity 1 means a fully present diagnosis. A diagnosis D with severity s will be abbreviated by $D{:}s$. In the following we will treat therapies like diagnoses, since they also can hold different severities (e.g. dosages) and have an effect on parameter values. These effects are described by dynamic set-covering relations.

Definition 10. A *dynamic set-covering relation* between a diagnosis $D \in \Omega_{\mathcal{D}}$ and a parameter $A \in \Omega_{\mathcal{A}}$ is denoted by $D \rightarrow A$.

For each parameter A we assume an ordered, discrete range

$$dom(A) = \{\, v_0, v_1, v_2, \ldots\}$$

with $v_0 \leq v_1 \leq v_2 \leq \ldots$ The smallest value $v_0 \in dom(A)$ is the default value; if there are no changes to A, then the initial finding is $A{:}v_0$. The changes that can be caused by a dynamic set-covering relation $D \rightarrow A$ are determined using an effect function ϵ and a membership function μ: Firstly, the severities s that are possible for D are translated into symbolic effects $\epsilon(D, s, A)$. Secondly, the symbolic effects $\epsilon(D, s, A)$ are translated into probability distributions $\mu\big(D, \epsilon(D, s, A), A\big)$. For every integer $n \in$ such a distribution specifies a probability p_n: if the diagnosis D is present with the severity s, then the value of A is changed by n units with the probability p_n; the value v_i of A is changed to v_j, where $j = \max(0, i + n)$, i.e., usually $j = i + n$, and only if n is negative and $i + n \leq 0$, then $j = 0$.

Definition 11. Dynamic set-covering model.

(1) An *effect function*

$$\epsilon : \Omega_{\mathcal{D}} \times \Omega_{sev} \times \Omega_{\mathcal{A}} \rightarrow \mathcal{E}\,,$$

assigns a symbolic effect $\epsilon(D, s, A) \in \mathcal{E}$ to a given diagnosis $D \in \Omega_{\mathcal{D}}$, a severity $s \in \Omega_{sev}$, and a parameter $A \in \Omega_{\mathcal{A}}$.

(2) A *membership function*

$$\mu : \Omega_{\mathcal{D}} \times \mathcal{E} \times \Omega_{\mathcal{A}} \rightarrow \quad \times [0,1]$$

assigns a probability distribution to a given diagnosis $D \in \Omega_{\mathcal{D}}$, an effect $e \in \mathcal{E}$, and a parameter $A \in \Omega_{\mathcal{A}}$.

(3) A *dynamic set-covering model* is defined as a triple $(\mathcal{R}, \epsilon, \mu)$, where $\mathcal{R}$ is a set of dynamic set-covering relations, ϵ is an effect function, and μ is a membership function.

For simplicity, we work with a finite set

$$\mathcal{E} = \{\ \mathit{N3}, \mathit{N2}, \mathit{N1}, \mathit{0}, \mathit{P1}, \mathit{P2}, \mathit{P3}\ \}$$

of symbols for possible effects. For example the symbol *N1* means a slight decrease of the involved parameter value, whereas *P1* defines an increase, *0* maps no effect to the existing value; the other symbols are a gradually stronger or weaker change of the finding values. In general, if $D \rightarrow A \in \mathcal{R}$ and $\epsilon(D, s, A) = Pk$, then the value v_i of A is increased to v_{i+k}. For negative effects Nk the value is analogously decreased to $v_{max(0,i-k)}$. In Section 4.1 we will describe how multiple effects $D \rightarrow A \in \mathcal{R}$ of different diagnoses (and severities) on the same parameter A can be aggregated.
In many practical cases the value $\mu(D, e, A)$ of a membership function is independent of the arguments D and A. The simplest way to translate the effects into specific findings is to define the *categorical interpretation* of the symbols in $\mathcal{E}$: for all $D, A, k \in \{0, 1, 2, 3\}$,

$$\mu(D, Pk, A) = \{\ (k, 1)\ \},$$
$$\mu(D, Nk, A) = \{\ (-k, 1)\ \}.$$

If $\epsilon(D, s, A) = Pk$ for a given diagnosis D, a severity s and a parameter A, then $D \rightarrow A \in \mathcal{R}$ results in an increase of the parameter value v_i to v_{i+k} with the probability 1. "*Nk*" results in a decrease of the parameter value v_i to v_{i-k} with the probability 1. If a parameter value receives a negative effect, which is greater than its current (positive) value, then the value is set to the default value v_0.
But in general the interpretation of the symbolic effects in $\mathcal{E}$ is *fuzzy*: the effect (e.g., slight decrease of the parameter value) of $D \rightarrow A$ can be uncertain and the change of the value may result in different parameter values with different probabilities. For a *fuzzy interpretation* a membership function is defined. An example is shown in the following:

$$\mu(D, \mathit{P1}, A) = \{\ (0, 0.1), (1, 0.7), (2, 0.2)\ \}.$$

We can see, that the value of parameter A will increase by 1 with 70% probability and it will increase by 2 with 20% probability. The parameter value will not change at all with 10% probability.

4.1. *Aggregation of Multiple Effects*

We will present two alternative ways for aggregating multiple effects on the same parameter A: order-of-magnitude reasoning and accumulative reasoning. Which of the two is appropriate depends on the situation. When bringing the effects into the set-covering relations the modeler has to decide about how the multiple influences on a parameter should be handled, and attach it to the influenced parameter. We

emphasize that these procedures are quite simple but keep the model simple as well and the reasoning step understandable for the user.

Order-of-Magnitude Reasoning

Raiman[12] motivated this procedure by the fact that larger magnitudes of effects might have a significantly larger impact on a finding. E.g., if the diagnosis $D_1 =$ *chronical polyarthritis* (*pcP*) increases the value of the parameter $A =$ *inflammations* by *P3* and another diagnosis D_2 increases the value by *P1*, then D_2 may not have any additional effect on the value of A.
For a *categorical* interpretation of the symbols we only count the largest magnitude when all effects have the same sign. For an *fuzzy* interpretation of the involved effects we have to transform the symbols and calculate the possible events.
For example, consider the set-covering relations $R = \{\, D_1 \rightarrow A,\ D_2 \rightarrow A \,\}$ together with the effect function given by $\epsilon(D_1, 1, A) = P2$ and $\epsilon(D_2, 1, A) = P1$ and the membership function μ:

$$\mu(D_1, P2, A) = \{\ (1, 0.1),\ (2, 0.8),\ (3, 0.1)\ \},$$
$$\mu(D_2, P1, A) = \{\ (0, 0.1),\ (1, 0.8),\ (2, 0.1)\ \}.$$

Using order-of-magnitude reasoning the value of parameter A can be changed by either 1, 2 or 3. For the moment, we only want to consider the probability of the composite causation event, that D_1 with severity 1 and D_2 with severity 1 will cause value of the parameter A to be increased by 2, i.e. for $\mathcal{H} = \{\, D_1{:}1, D_2{:}1 \,\}$ and $F = A{:}v_2$ the composite causation event $P(\mathcal{H} \mapsto F | \mathcal{H})$. For $\mathcal{H} \mapsto F$ we have to consider the following sub-causation events e_i defined by the membership functions for *P2* and *P1*:

$$\begin{aligned}
e_1 &= D_1{:}1 \mapsto A{:}v + 1 \qquad \text{with } P(e_1 \,|\, D_1{:}1) = 0.1,\\
e_2 &= D_1{:}1 \mapsto A{:}v + 2 \qquad \text{with } P(e_2 \,|\, D_1{:}1) = 0.8,\\
e_3 &= D_2{:}1 \mapsto A{:}v + 0 \qquad \text{with } P(e_3 \,|\, D_2{:}1) = 0.1,\\
e_4 &= D_2{:}1 \mapsto A{:}v + 1 \qquad \text{with } P(e_4 \,|\, D_2{:}1) = 0.8,\\
e_5 &= D_2{:}1 \mapsto A{:}v + 2 \qquad \text{with } P(e_5 \,|\, D_2{:}1) = 0.1.
\end{aligned}$$

The causation event $\mathcal{H} \mapsto F$ occurs if $(e_2 \wedge e_3) \vee (e_2 \wedge e_4) \vee (e_2 \wedge e_5) \vee (e_1 \wedge e_5)$. We compute the probability for the composite causation event $P(\mathcal{H} \mapsto F | \mathcal{H})$ by

$$\begin{aligned}
P(\mathcal{H} \mapsto F | \mathcal{H}) = {} & P(e_2 | D_1{:}1) \cdot P(e_3 | D_2{:}1) + P(e_2 | D_1{:}1) \cdot P(e_4 | D_2{:}1) +\\
& P(e_2 | D_1{:}1) \cdot P(e_5 | D_2{:}1) + P(e_1 | D_1{:}1) \cdot P(e_5 | D_2{:}1) =\\
= {} & 0.08 + 0.64 + 0.08 + 0.01 = 0.81.
\end{aligned}$$

The calculation of the probabilities for the composite causation events $P(\mathcal{H} \mapsto F | \mathcal{H})$ for $F = A{:}v_1$ and $F = A{:}v_3$ is analogous. A problem arises, when we have two effects symbols with different signs but the same magnitude (e.g., *P1* and *N1*). In this case a general solution is not possible and the modeler has to define a reasonable heuristic.

Accumulative Reasoning

If a parameter value is influenced not only by diagnoses but also by therapies, then an accumulative approach seems to be more appropriate. Here we compute the change on a parameter value by transforming each symbolic effect into its numeric representation. Then we simply obtain the resulting change by the sum of the numeric values. This procedure will work fine for a categorical interpretation, but we have to define this for a fuzzy interpretation more precisely. The idea is that we have to compute the cartesian product to get all possible outcomes for an influenced finding.
For example, consider the set-covering relations $R = \{\, D \to A,\, T \to A \,\}$ together with the effect function given by $\epsilon(D, 1, A) = P2$ and $\epsilon(T, 1, A) = N1$, and the following membership function μ:

$$\mu(D, P2, A) = \{\ (1, 0.1),\ (2, 0.8),\ (3, 0.1)\ \},$$
$$\mu(T, N1, A) = \{\ (-2, 0.1),\ (-1, 0.8),\ (0, 0.1)\ \}.$$

We can see, that the accumulative aggregation of the two symbolic effects can change the value of the parameter A by either -1, 0, 1, 2 or 3. Let us assume the combined causation event $\mathcal{H} \mapsto F$ for $\mathcal{H} = \{\, D{:}1, T{:}1 \,\}$ and $F = A{:}v_0$ ("A does not change its value for D and T"). If we want to compute the probability of the combined causation event $P(\mathcal{H} \mapsto F | \mathcal{H})$, then we simply have to consider the following sub-causation events:

$$\begin{array}{lll} e_1 = D{:}1 \mapsto A{:}v + 1 & \text{with} & P(e_1 \,|\, D{:}1) = 0.1 \\ e_2 = T{:}1 \mapsto A{:}v - 1 & \text{with} & P(e_2 \,|\, T{:}1) = 0.8 \\ e_3 = D{:}1 \mapsto A{:}v + 2 & \text{with} & P(e_3 \,|\, D{:}1) = 0.8 \\ e_4 = T{:}1 \mapsto A{:}v - 2 & \text{with} & P(e_4 \,|\, T{:}1) = 0.1 \end{array}$$

The causation event $\mathcal{H} \mapsto F$ occurs if $(e_1 \wedge e_2) \vee (e_3 \wedge e_4)$. Under the assumption (cf. Section 3.3) that the sub-causation events are independent we can compute

$$\begin{aligned} P(\mathcal{H} \mapsto F | \mathcal{H}) &= P(e_1 | D{:}1) \cdot P(e_2 | T{:}1) + P(e_3 | D{:}1) \cdot P(e_4 | T{:}1) = \\ &= 0.8 \cdot 0.1 + 0.1 \cdot 0.8 = 0.16. \end{aligned}$$

The calculations for the remaining composite causation events $P(\mathcal{H} \mapsto F | \mathcal{H})$ for $\mathcal{H} = \{\, D{:}s_d, T{:}s_t \,\}$ for $s_d, s_t \in [0, 1]$ and $F = A{:}v + x$ and $x \in \{\, -1, 1, 2, 3 \,\}$ are analogous.

4.2. *Generation and Evaluation of Hypotheses*

Given an observation $\mathcal{F}_\mathcal{O} \subseteq \Omega_\mathcal{F}$, let $\mathcal{D} \subseteq \Omega_\mathcal{D}$ be the set of all diagnoses/therapies, which can explain a parameter $A \in a(\mathcal{F}_\mathcal{O})$; we assume that each $A \in a(\mathcal{F}_\mathcal{O})$ can be explained by at least one $D \in \mathcal{D}$. We take the powerset $\mathcal{P}(\mathcal{D})$ of $\mathcal{D}$ as the candidate set of hypotheses for explaining $\mathcal{F}_\mathcal{O}$. For each hypothesis $\mathcal{H} \in \mathcal{P}(\mathcal{D})$ and for each parameter $A \in a(\mathcal{F}_\mathcal{O})$, let

$$\mathcal{H}_A = \{\, D \in \mathcal{H} \,|\, D \text{ covers } A \,\} \subseteq \mathcal{H}.$$

Then, we apply Equation 8 for $F \in \mathcal{F}_{\mathcal{H}}$ and $A = a(F)$:

$$pc(A) = \begin{cases} P(\mathcal{H}_A \mapsto F | \mathcal{H}_A) & \text{if } F \in \mathcal{F}_{\mathcal{O}}, \\ 0 & \text{otherwise.} \end{cases}$$

Using this modified distance function we can compute the quality of each hypothesis $\mathcal{H}$. The hypothesis $\mathcal{H}$ with the maximal quality $\varrho(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ will be considered as the best explanation.

We want to illustrate the previous considerations with an example, that contains all aspects which we have described above. Consider the dynamic set-covering relations in $R = \{ D \rightarrow A,\ T \rightarrow A \}$ and the effect function ϵ:

s	0	0.5	1
$\epsilon(D, s, A)$	*0*	*P2*	*P3*

s	0	0.3	0.7	1
$\epsilon(T, s, A)$	*0*	*N1*	*N2*	*N3*

Observe, that ϵ now depends on its first argument, the diagnosis and the therapy, respectively. For a categorical interpretation and an accumulative aggregation of the specified effect symbols we can compute the following causation events (assuming that neither D nor T is assigned to severity 0):

$\mathcal{H} = \{ D{:}s_d, T{:}s_t \}$	$\epsilon(D, s_d, A) + \epsilon(T, s_t, A)$	*Causation Event*
$\mathcal{H}_1 = \{ D{:}0.5, T{:}0.3 \}$	$P2 + N1 = P1$	$\mathcal{H}_1 \mapsto A{:}1$
$\mathcal{H}_2 = \{ D{:}0.5, T{:}0.7 \}$	$P2 + N2 = P0$	$\mathcal{H}_2 \mapsto A{:}0$
$\mathcal{H}_3 = \{ D{:}0.5, T{:}1 \}$	$P2 + N3 = N1$	$\mathcal{H}_3 \mapsto A{:}0$
$\mathcal{H}_4 = \{ D{:}1, T{:}0.3 \}$	$P3 + N1 = P2$	$\mathcal{H}_4 \mapsto A{:}2$
$\mathcal{H}_5 = \{ D{:}1, T{:}0.7 \}$	$P3 + N2 = P1$	$\mathcal{H}_5 \mapsto A{:}1$
$\mathcal{H}_6 = \{ D{:}1, T{:}1 \}$	$P3 + N3 = P0$	$\mathcal{H}_6 \mapsto A{:}0$

Thus, we can explain the observation $\mathcal{F}_{\mathcal{O}} = \{ A{:}1 \}$ by the hypothesis $\mathcal{H}_1$ or by the hypothesis $\mathcal{H}_5$. Since we applied a categorical transformation, all causation events have an equal likelihood. Therefore both hypotheses have the same quality measure, assuming an equal apriori probability for D and T.

4.3. *Complexity of Set-Covering Models*

It is well-known, that abductive reasoning is intractable in general (cf. Bylander et al.[13]). Due to the hypothesize-and-test strategy, static set-covering models need $\mathcal{O}(2^{|\Omega_{\mathcal{D}}|})$ hypotheses to be generated to find the best explanation for a given observation. This becomes even worse with dynamic set-covering models, since diagnoses will be assigned not only to a binary value (present vs. absent), but now can have a discrete value range of severities.

Although a lot of work has been done to improve hypothesis generation for abductive models (e.g. Console et al.[14], van der Gaag and Wessels[15]), we see only small room for a practical diagnosis application with dynamic set-covering models. Nevertheless, dynamic models are an interesting approach, when building knowledge-based

intelligent tutoring systems. In such a system a virtual patient case (observation set) can be generated from a set-covering model and a specified set of diagnoses and therapies. Using a dynamic set-covering model with fuzzy effects, the case is randomly sampled from the given membership functions. Then, the generated patient case is presented to the user, who has to find diagnoses explaining the presented findings and the given therapies. It is worth noticing, that the generated observation sets (patient cases) can be explained to the tutoring student due to the set-covering knowledge. Hörnlein, Baumeister and Puppe[16] describe the application of set-covering models for intelligent tutoring systems in more detail.
In this way, the system does not need to find an optimal solution (hypothesis) for a given observation, but is used to generate an observation set for a given set of diagnoses and therapies. With fuzzy effect functions it is possible to sample different patient cases for a small set-covering model.

5. Applications Using Set-Covering Models

In this section we discuss two applications which we have developed using the set-covering approach described in this paper. So far, only static set-covering models have been applied. Due to the complexity restrictions we are planning to apply dynamic set-covering models only for developing tutoring systems; in this application domain, severities usually are known beforehand, and they are used for simulating case descriptions, which decreases the computational complexity dramatically but retains the dynamic behaviour of the model.

5.1. *Diagnosing Contamination of Small Water Streams with* Limpact

Recently, we have successfully completed an application and evaluation of a set-covering diagnostic system (Limpact) in a geo-ecological domain. Limpact estimates the pesticide contamination of small lowland streams with agricultural catchment areas. The system considers the abundance data of 39 macroinvertebrate taxa. Limpact can be used via the WWW[17].
A previous version of the knowledge base has been constructed with heuristic diagnostic rules and included about 1000 rules[18,19]; it has been implemented within the diagnostic shell-kit D3[8]. As expected the rule base has become complex.
For this reason, in a second iteration the domain expert represented his knowledge by simple set-covering relations and added weights and exclusion conditions[11] to increase the diagnostic quality. As shown by Neumann and Baumeister[20] the quality of the new knowledge base using set-covering relations is comparable to the quality of the old knowledge base using heuristic rules.
As a remarkable result we significantly reduced the complexity of the knowledge base: Instead of about 1000 rules containing complex conditions, we applied about 700 simple set-covering relations in the new version of Limpact. Therefore the formalized set-covering knowledge is smaller in size and complexity. The expert

Diagnosis: *High Contamination*	
Parameter	Abundance Data
Agabus sp.	< 3
Anabolia nervosa	< 55
Asellus aquaticus	< 65
Baetis sp.	< 13
Baetis vernus	< 39
Ceratopogonidae sp.	< 15
Chaetopteryx villosa	< 81
Chironomidae sp. red	< 203
Chironomidae sp. white	< 18
Platambus maculatus	> 200
Plectrocnemia conspersa	in $[400; 530]$
Ptychopteridae sp.	> 200
Radix ovata	in $[2; 5]$
Simuliidae sp.	in $[2; 6]$

Fig. 3. Excerpt of the set-covering model for diagnosis *High Contamination*.

found the semantics of set-covering relations as easy and quickly to understand as heuristic rules. Figure 5.1 shows an excerpt of the LIMPACT knowledge base with 14 set-covering relations for the diagnosis *High Contamination*. A detailed discussion of the knowledge base and the evaluation is presented in [20].

5.2. *Learning Set-Covering Models from Data in* SonoConsult

In addition to a manual acquisition approach we also considered *semi-automatic* learning methods of set-covering models, where the user can interact with the learning methods by providing additional knowledge. The results of this work were presented by Baumeister et al.[9], which described learning methods for set-covering models as well as for knowledge extensions like weights and similarities. As a special characteristic, these methods are able to process additional background knowledge like hierarchical or ontological information. Atzmueller et al. [21] compared the set-covering approach with a knowledge-intensive partitioning approach. Our evaluation was based on the knowledge-based documentation and consultation system for sonography SONSOCONSULT (an advanced and isolated part of HEPATOCONSULT[22]) being in routine use in the DRK-hospital in Berlin/Köpenick based on the diagnostic shell kit D3[8]. The installation in Berlin currently gathers about 300 cases per month, which are highly structured and have a high diagnostic quality; usually they are able to infer the correct diagnosis. As a remarkable result we showed that the learned set-covering models outperformed a previously evaluated case-based approach which was not able to handle multiple diagnoses contained in the cases.

6. Summary

In this paper we have presented an incremental approach for building set-covering models using qualitative and quantitative knowledge combined in one formalism. Starting with a simple qualitative model one can extend the set-covering model by similarities, weights, uncertainty and severities. For all extensions we provided a formal description and a guidance to integrate it into the set-covering theory. The goal of each extension is the improvement of the diagnostic quality achieved by the model. Furthermore we introduced a novel approach called dynamic set-covering models, for which we considered different ways of the interpretation and the accounting of (multiple) symbolic effects.
Besides these theoretical aspects we described two applications, in which we applied our approach: For the LIMPACT project we have shown, that incremental development of set-covering models significantly reduced the complexity of the knowledge base and its acquisition costs. The results of SONOCONSULT-Project describe how to learn set-covering models and knowledge extensions like similarities or weights from cases. In addition, the set-covering approach was integrated into a case-based reasoning framework.

Currently we are working on an improved method for learning set-covering models from cases, which was presented in [9]. Furthermore, set-covering models show interesting relations with case-based reasoning. Besides a similar usage of knowledge extensions, set-covering models and case-based reasoning provide a semantically equivalent method for hypothesis evaluation. Thus, we are planning to work on a formal framework that combines both approaches.
The development of static set-covering models can be compared to the methods for building Bayesian networks[23]: In both representations, the expert starts defining relationships between diagnoses and findings by a qualitative structure. Then, in a second step the network is extended by quantitative knowledge. In Bayesian networks the expert can only state probabilities describing the dependencies between the nodes, and the semantics of the network is well-defined. In set-covering models the knowledge is more explicit and motivated by the mental models of the experts. Since different knowledge extensions are combined, the semantics of set-covering models has to be defined more carefully (see Section 3.4). Currently, we are working on a formal comparison between both approaches.

References

1. Andreas Warnke. Dynamic Belief Networks for Temporal Therapy Processes (in german). Master's thesis, University Wuerzburg, Department for Computer Science VI, 2001.
2. Ramesh S. Patil, Peter Szolovits, and William B. Schwartz. Modeling Knowledge of the Patient in Acid-Base and Electrolyte Disorders. *In: Szolovits, P. (Ed.). Artificial Intelligence in Medicine, Westview Press*, 1982.
3. James A. Reggia, Dana S. Nau, and Pearl Y. Wang. Diagnostic Expert Systems Based on a Set Covering Model. *Journal of Man-Machine Studies*, 19(5):437–460, 1983.

4. Yun Peng and James A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving*. Springer Verlag, Berlin, 1990.
5. William J. Long. Medical Diagnosis using a Probabilistic Causal Network. *Applied Artificial Intelligence*, 3:367–383, 1989.
6. Larry Eshelman. *Mole: A Knowledge-Acquisition Tool for Cover-and-Differentiate Systems*, pages 37–79. In: Sandra Marcus (ed.): Automating Knowledge Acquisition for Expert Systems. Kluwer Academic Publisher, 1988.
7. Peter Lucas, Astrid Tholen, and Geeske van Oort. An Intelligent System for Pacemaker Reprogramming. *Artificial Intelligence in Medicine*, 17:249–269, 1999.
8. Frank Puppe. Knowledge Reuse among Diagnostic Problem-Solving Methods in the Shell-Kit D3. *International Journal of Human-Computer Studies*, 49:627–649, 1998.
9. Joachim Baumeister, Martin Atzmueller, and Frank Puppe. Inductive Learning for Case-Based Diagnosis with Multiple Faults. In *Proceedings of the 6th European Conference on Case-Based Reasoning, LNAI 2416, Springer Verlag*, pages 28–42, Aberdeen, Scotland, 2002.
10. Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publisher, 2000.
11. Joachim Baumeister and Dietmar Seipel. Diagnostic Reasoning with Multilevel Set–Covering Models. In *Proceedings of the 13th International Workshop on Principles of Diagnosis (DX-02)*, Semmering, Austria, 2002.
12. Olivier Raiman. Order of Magnitude Reasoning. *Artificial Intelligence*, 51:11–38, 1991.
13. Tom Bylander, Dean Allemang, Michael C. Tanner, and John R. Josephson. The Computational Complexity of Abduction. *Artificial Intelligence*, 49(1-3):25–60, 1991.
14. Luca Console, Luigi Portinale, and Daniele Theseider Dupre. Using Compiled Knowledge to Guide and Focus Abductive Diagnosis. *Knowledge and Data Engineering*, 8(5):690–706, 1996.
15. Linda van der Gaag and Maria Wessels. Efficient Multiple-Disorder Diagnosis by Strategic Focusing. Technical Report UU-CS-1994-23, Universiteit Utrecht, NL, 1994.
16. Alexander Hörnlein, Joachim Baumeister, and Frank Puppe. Generating Sequenced Consultations for Therapy Monitoring in Case-Based Tutoring Systems (in german). In F. Puppe et al., editor, *Proceedings zum 7. Workshop der GMDS AG Computergestützte Lehr- und Lernsysteme in der Medizin (CBT-2003), Universität Würzburg*. Shaker Verlag, 2003.
17. The LIMPACT-Project: http://www.limpact.de.
18. Michael Neumann, Joachim Baumeister, Matthias Liess, and Ralf Schulz. An Expert System to Estimate the Pesticide Contamination of Small Streams using Benthic Macroinvertebrates as Bioindicators, Part 2: The Knowledge Base of LIMPACT. *Journal Ecological Indicators, Elsevier Science*, 2(4):391–401, 2003.
19. Michael Neumann, Matthias Liess, and Ralf Schulz. An Expert System to Estimate the Pesticide Contamination of Small Streams using Benthic Macroinvertebrates as Bioindicators, Part 1: The Database of LIMPACT. *Journal Ecological Indicators, Elsevier Science*, 2(4):379–389, 2003.
20. Michael Neumann and Joachim Baumeister. A Rule-Based vs. a Model-Based Implementation of the Knowledge System LIMPACT and its Significance for Maintenance and Discovery of Ecological Knowledge. In *Proceedings of the 3rd Conference of the International Society of Ecological Informatics (ISEI-02)*, Rome, Italy, 2002.
21. Martin Atzmueller, Joachim Baumeister, and Frank Puppe. Evaluation of two Strategies for Case-Based Diagnosis handling Multiple Faults. In *Proceedings of the 2nd Confernence of Professional Knowledge Management (WM2003)*, Luzern, Switzerland, 2003.

22. Hans-Peter Buscher, Ch. Engler, A. Führer, S. Kirschke, and F. Puppe. HepatoConsult: A Knowledge-Based Second Opinion and Documentation System. *Artificial Intelligence in Medicine*, 24(3):205–216, 2002.
23. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publisher, San Mateo, California, 1988.