

Linkping Electronic Articles in
Computer and Information Science
Vol. ?(2002): nr ?

Incremental Development of Diagnostic Set–Covering Models with Therapy Effects

Joachim Baumeister
Dietmar Seipel
Frank Puppe

Department of Computer Science
University of Würzburg
Würzburg, Germany

Linkping University Electronic Press
Linköping, Sweden

<http://www.ep.liu.se/ea/cis/2002/00?/>

*Published on , 2002 by
Linköping University Electronic Press
581 83 Linköping, Sweden*

**Linköping Electronic Articles in
Computer and Information Science**
*ISSN 1401-9841
Series editor: Erik Sandewall*

*©2002 Joachim Baumeister
Typeset by the author using \LaTeX
Formatted using étendu style*

Recommended citation:

*<Author>. <Title>. Linköping Electronic Articles in
Computer and Information Science, Vol. ?(2002): nr ?.
<http://www.ep.liu.se/ea/cis/2002/00?/>. , 2002.*

This URL will also contain a link to the author's home page.

*The publishers will keep this article on-line on the Internet
(or its possible replacement network in the future)
for a period of 25 years from the date of publication,
barring exceptional circumstances as described separately.*

*The on-line availability of the article implies
a permanent permission for anyone to read the article on-line,
to print out single copies of it, and to use it unchanged
for any non-commercial research and educational purpose,
including making copies for classroom use.
This permission can not be revoked by subsequent
transfers of copyright. All other uses of the article are
conditional on the consent of the copyright owner.*

*The publication of the article on the date stated above
included also the production of a limited number of copies
on paper, which were archived in Swedish university libraries
like all other written works published in Sweden.
The publisher has taken technical and administrative measures
to assure that the on-line version of the article will be
permanently accessible using the URL stated above,
unchanged, and permanently equal to the archived printed copies
at least until the expiration of the publication period.*

*For additional information about the Linköping University
Electronic Press and its procedures for publication and for
assurance of document integrity, please refer to
its WWW home page: <http://www.ep.liu.se/>
or by conventional mail to the address stated above.*

Abstract

We will consider set-covering models for diagnostic tasks handling uncertain knowledge. Because these models have to be build manually by domain experts, we will show how this effort can be reduced by an incremental development of the set-covering models. Thus simple models can be enhanced by similarities, weights and uncertainty to increase the quality of the knowledge and the resulting system. We will also present mechanisms for generating set-covering models when higher level knowledge about causation effects is available. Finally, we will motivate how our approach can be used for implementing diagnosis systems including therapy effects.

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

1 Introduction

This paper will focus on the problem of developing knowledge based diagnosis systems. The work was motivated by the experience we have made with experts, who tried to build complex models from scratch [11] but did not have enough samples for machine learning algorithms. Even simplifications of complex 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 diagnosis 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 diagnosis 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 covering relations between the elements of these two sets. There exists a 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. To do this 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 [5], 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 and Nau [10]. In [9] they introduce numeric probabilities to describe covering relations in more detail and discuss a transformation to the Theorem of Bayes. Similarly, Long [3] uses probabilistic covering models related to Bayesian networks, but allows the use of forward-loops and conditional links. The system MOLE [1] 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. [4] 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 [7] 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 evaluation 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 above, although it has been proven 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 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 will give a short survey of further extensions we want to work on in the future.

2 A Framework for Diagnosis in Set-Covering Models

Covering models describe relations like:

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

We call these relations *covering relations*; each single covering relation is denoted by $D \rightarrow (A = v)$. 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 . Figure 1 exemplifies a covering model for the diagnoses D_1 and D_2 . Edges indicate explicit covering relations. For example the net states, that diagnosis D_1 forces the observation of the parameters A_1 , A_2 and A_3 . In this way the three covering relations $D_1 \rightarrow (A_1 = a)$, $D_1 \rightarrow (A_2 = b)$ and $D_1 \rightarrow (A_3 = c)$ hold.

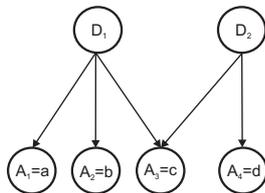


Figure 1: Covering net for the diagnoses D_1 and D_2 and the parameters A_1, \dots, A_4 .

The *basic algorithm* for *set-covering diagnosis* is very simple: Given a set of observed findings, it uses a simple 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 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.

The basic sets for this task are the following: $\Omega_{\mathcal{D}}$ is the set of all possible diagnoses included in the model, and $\Omega_{\mathcal{A}}$ is the set of all possible parameters (attributes). To each parameter $A \in \Omega_{\mathcal{A}}$ a range $dom(A)$ of values is assigned, and $\Omega_{\mathcal{V}} = \cup_{A \in \Omega_{\mathcal{A}}} dom(A)$ is the set of all possible values for the parameters. Further we assume $\mathcal{F}_{\Omega_{\mathcal{A}}}$ to be the (universal) set of findings ($A = v$), where $A \in \Omega_{\mathcal{A}}$ is a parameter and $v \in dom(A)$ is a possible value. We will call $\mathcal{F}_{\mathcal{O}} \subseteq \mathcal{F}_{\Omega_{\mathcal{A}}}$ the set of *observed findings*. As well, we will call a set of diagnoses $\mathcal{H} \subseteq \Omega_{\mathcal{D}}$ a *hypothesis*, e.g., $\mathcal{H} = \{D_1, D_2\}$. $\mathcal{F}_{\mathcal{H}} \subseteq \mathcal{F}_{\Omega_{\mathcal{A}}}$ denotes the set of all findings ($A = v$), which are covered by the diagnoses $D \in \mathcal{H}$, i.e., such that $D \rightarrow (A = v)$ holds. We will call $\mathcal{F}_{\mathcal{H}}$ the set of *predicted findings*. For a given set \mathcal{F} of findings, let $\alpha(\mathcal{F})$ be the set of all parameters A for which findings ($A = v$) $\in \mathcal{F}$ exist. For all $A \in \alpha(\mathcal{F})$ we define $\mathcal{F}(A) = v$ to be the value assigned to the parameter A and we assume that there exists only one finding ($A = v$) for each parameter A . We will call $\alpha(\mathcal{F}_{\mathcal{H}})$ and $\alpha(\mathcal{F}_{\mathcal{O}})$ the sets of *predicted* and *observed parameters*, respectively. E.g., for $\mathcal{F}_{\mathcal{O}} = \{(A_1 = a), (A_2 = b), (A_3 = e), (A_4 = d)\}$, we get $\alpha(\mathcal{F}_{\mathcal{O}}) = \{A_1, A_2, A_3, A_4\}$. These definitions lead us to a couple of sets used by the reasoning algorithm.

1. The set $\alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \alpha(\mathcal{F}_{\mathcal{O}}) \cap \alpha(\mathcal{F}_{\mathcal{H}})$ contains all parameters, which are observed and predicted. Corresponding to $\alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ we define the set $\mathcal{F}_{\mathcal{H}, \mathcal{O}} = \{(A = v) \in \mathcal{F}_{\mathcal{H}} | A \in \alpha(\mathcal{F}_{\mathcal{O}})\}$ which contains all predicted findings ($A = v$) such that A is observed (i.e., $\alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) \subseteq \alpha(\mathcal{F}_{\mathcal{H}, \mathcal{O}})$).
2. The set $\alpha^+(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) \subseteq \alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ of *positively observed parameters* contains all observed parameters which are observed with the value that is predicted by the hypothesis.
3. The set $\alpha^-(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) \subseteq \alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ of *negatively observed parameters* contains all observed parameters which are observed with another value than predicted by the hypothesis.
4. The set $\alpha^u(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \alpha(\mathcal{F}_{\mathcal{O}}) \setminus \alpha(\mathcal{F}_{\mathcal{H}})$ of *unspecific observed parameters* contains all observed parameters which are not predicted by the given hypothesis. These parameters cannot be explained by the given hypothesis.

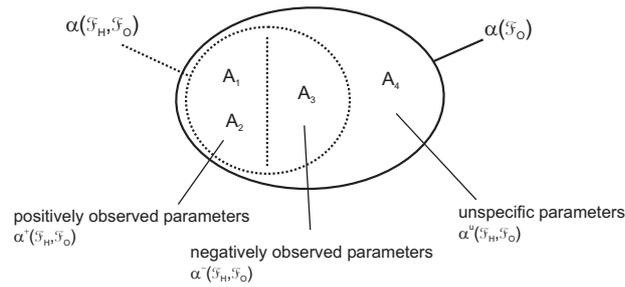


Figure 2: The subsets $\alpha^+(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$, $\alpha^-(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ and $\alpha^u(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ of the observation set $\alpha(\mathcal{F}_{\mathcal{O}})$

Figure 2 gives a graphical representation of these sets for the covering relations stated in Figure 1 and the assumption that we observe $\mathcal{F}_{\mathcal{O}} = \{(A_1 = a), (A_2 = b), (A_3 = e), (A_4 = d)\}$ with hypothesis $\mathcal{H} = \{D_1\}$. Thus, we get $\mathcal{F}_{\mathcal{H}} = \{(A_1 = a), (A_2 = b), (A_3 = c)\}$ and $\alpha^+(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \{A_1, A_2\}$. Parameters in $\alpha^-(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) = \{A_3\}$ are predicted by \mathcal{H} , but they are observed with

another value as predicted by \mathcal{H} . This property distinguishes it from the set $\alpha^u(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) = \{A_4\}$, which contains parameters that are not predicted by \mathcal{H} . It always holds that

$$\alpha(\mathcal{F}_\mathcal{O}) = \alpha^+(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) \cup \alpha^-(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) \cup \alpha^u(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}).$$

Quality Measures. The quality measures are based on a binary difference function $\|\cdot, \cdot\| : 2^{\mathcal{F}_{\Omega, A}} \times 2^{\mathcal{F}_{\Omega, A}} \rightarrow \mathbb{R}$, and on an unary weight function $\|\cdot\| : 2^{\mathcal{F}_{\Omega, A}} \rightarrow \mathbb{R}$, for sets of findings:

1. Given $\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O} \subseteq \mathcal{F}_{\Omega, A}$, then $\|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\|$ measures the difference between the predicted and the observed findings.
 - In the simplest case, $\|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\| = |\mathcal{F}_\mathcal{H} \cap \mathcal{F}_\mathcal{O}|$ is the number of findings, for which the predicted value is also observed.
 - 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.
2. Given $\mathcal{F} \subseteq \mathcal{F}_{\Omega, A}$, then $\|\mathcal{F}\|$ is the weight of the set \mathcal{F} of findings. In the simplest case, $\|\mathcal{F}\| = |\alpha(\mathcal{F})|$ 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 2.1 (Quality Measures)

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 predicted findings:

$$\pi(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) = \frac{\|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\|}{\|\mathcal{F}_\mathcal{H}, \mathcal{O}\|} \quad (1)$$

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

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

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

$$\varrho(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) = \pi(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) \cdot \kappa(\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}) = \frac{\|\mathcal{F}_\mathcal{H}, \mathcal{F}_\mathcal{O}\|}{\|\mathcal{F}_\mathcal{O}\|}. \quad (3)$$

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*: Given a *completeness value* $cv(D) \in [0, 1]$ for each diagnosis D , which has to be specified by the expert, we call a hypothesis \mathcal{H} sufficiently complete, if for each diagnosis $D \in \mathcal{H}$ the ratio between the number of observed parameters covered by D and the number of all predicted parameters of D exceeds $cv(D)$:

$$\frac{|\alpha(\mathcal{F}_{\{D\}}, \mathcal{F}_O)|}{|\alpha(\mathcal{F}_{\{D\}})|} \geq cv(D) \quad (4)$$

The introduction of the term sufficiently complete hypothesis is motivated by the fact, that a covering model for a diagnosis will contain more findings than the diagnosis will cause in an average case. Nevertheless in most cases the observation of a percentage of the modeled findings will legitimate the validation of this diagnosis. To emphasize this percentage the modeler has to specify a completeness value $cv(D)$. Unless this factor is reached by the observation set in the current case, the diagnosis may not be considered as a valid subset of any hypothesis.

Extensions. The rest of the paper will concentrate on extensions of the simple covering model presented in Figure 1. Starting with simple covering relations we can apply additional information to improve the diagnostic quality. Each addition forms a supplementary component of the 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 " $pcP \rightarrow (\text{Pain} = \text{very high})$ " and we can observe ($\text{Pain} = \text{high}$). In the simple covering model this observation would be negatively observed and be treated exactly like ($\text{Pain} = \text{normal}$), 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 we have the relations " $pcP \rightarrow (\text{morning stiffness} = \text{true})$ " and " $pcP \rightarrow (\text{rheumatic blood parameter} = \text{high})$ " the high blood parameters may be more valuable to the diagnostic process than the parameter *morning stiffness*. Because of this we will add weights to parameters to emphasize their diagnostic power in comparison to other parameters. 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 " $pcP \rightarrow (\text{Pain} = \text{very high})$ ", which will be introduced as a single knowledge component. Beyond that, one can build generic models by defining causal relationships between severities of diagnoses and parameter values. For example if the disease *pcP* has the severities *normal*, *weak*, *strong*, we can easily model a 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 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 functions $\| \cdot \|$ and $\| \cdot, \cdot \|$ 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 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 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: $\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = |\alpha^+(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})|$ and $\|\mathcal{F}\| = |\mathcal{F}|$. 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 define *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. In cluster analysis problems this function is also called *distance function* (c.f. [2]). We obtain the quality for models with similarities, when we define the functions $\|\cdot, \cdot\|$ and $\|\cdot\|$ as follows:

$$\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = \sum_{A \in \alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})} sim(\mathcal{F}_{\mathcal{H}}(A), \mathcal{F}_{\mathcal{O}}(A)), \quad \|\mathcal{F}\| = |\mathcal{F}|. \quad (5)$$

As a special case, we get the simple covering rate, if we define $sim(v, v') = 0$, if $v \neq v'$, and $sim(v, v') = 1$, if $v = v'$.

3.2 Weighted Findings in Covering Models

The introduction of weights for observed values is another common generalization of the simple covering model. Given a weight function $w : \Omega_{\mathcal{A}} \rightarrow \mathbb{N}_+$, we define

$$\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = \sum_{A \in \alpha^+(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})} w(A), \quad \|\mathcal{F}\| = \sum_{A \in \alpha(\mathcal{F})} w(A). \quad (6)$$

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

3.3 Uncertainty in Covering Relations

An important enhancement of covering models is the possibility to state uncertain covering relations. Thereby the diagnosis forces observations of specified findings not in any case but only with a specified probability. To facilitate such propositions and their integration in the computation of covering models we need some assumptions.

Assumptions and additional knowledge.

1. For each diagnosis $D \in \Omega_{\mathcal{D}}$ the apriori probability P_D is given and the occurrence of a diagnosis is independent from other diagnoses. As a consequence, the apriori probability for a hypothesis $\mathcal{H} = \{D_1, \dots, D_n\}$ is the product of the single probabilities.

2. For each covering relation $D \rightarrow (A = v)$ the conditional probability $P_D(A = v)$ of the occurrence of this relation is given in the model. The occurrence of the covering relation is independent of the occurrence of any other covering relation.

As an implication of the second item, the probability $P_{\mathcal{H}}(A = v)$ is calculated from the probabilities of events causing the occurrence of A and belonging to the combined hypothesis $\mathcal{H} = \{D_1, \dots, D_n\}$. The causing events are defined as subsets of the hypothesis \mathcal{H} reflecting the possible combinations. For example, assuming $\mathcal{H} = \{D_1, D_2\}$ we have to take the events (D_1) , (D_2) and $(D_1 \wedge D_2)$ for causing $(A_3 = c)$ into account. It holds that

$$P_{\mathcal{H}}(A = v) = \sum_{\mathcal{H}' \subseteq \mathcal{H}} \prod_{D \in \mathcal{H}'} \left(P_D(A = v) \cdot \left(1 - \prod_{D' \in \mathcal{H} \setminus \mathcal{H}'} P_{D'}(A = v) \right) \right) \quad (7)$$

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

Probabilistic covering models. We view

$$P_{\mathcal{H}} = \prod_{D \in \mathcal{H}} \left(P_D \cdot \prod_{D' \in \Omega_D \setminus \{D\}} (1 - P_{D'}) \right) \quad (8)$$

as the apriori probability of \mathcal{H} given by the diagnoses included in \mathcal{H} . The quality measure for a probabilistic covering model is defined given the following functions $\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\|$ and $\|\mathcal{F}\|$:

$$\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = \sum_{A \in \alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})} P_{\mathcal{H}} \cdot P_{\mathcal{H}, \mathcal{F}_{\mathcal{O}}}(A = v), \quad \|\mathcal{F}\| = |\mathcal{F}|. \quad (9)$$

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

$$P_{\mathcal{H}, \mathcal{F}_{\mathcal{O}}}(A = v) = \begin{cases} P_{\mathcal{H}}(A = v) & \text{for } A \in \alpha^+(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}), \\ 1 - P_{\mathcal{H}}(A = v) & \text{for } A \in \alpha^-(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}) \end{cases} \quad (10)$$

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 *leaky-diagnosis*. A leaky-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 leaky-diagnosis, which is categorically connected to all available findings. To shrink the emerging number of probabilities, we assume a constant probability for all covering relations between the leaky-diagnosis and a finding. If the model contains weights it is easy to see that the leaky probabilities can be adapted with respect to the weights. Large weights will decrease the leaky probability whereas small weights increase the probability. As a consequence, for every hypothesis we have to consider the leaky-diagnosis to be included in the hypothesis as well. It is easy to understand that with the usage of the leaky-diagnosis there will be no unspecific observed parameters because the leaky-diagnosis holds covering relations to all findings by default. So $\alpha^u(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})$ will be empty for all \mathcal{H} and all $\mathcal{F}_{\mathcal{O}}$.

Taking a few assumptions into account it can be shown that the probabilistic covering rate form an approximation for the conditional probability of a given hypothesis according to the Theorem of Bayes. Due to the limited space we omit the description of the transformation between the covering relations and the Bayesian probabilities.

3.4 Combining the Different Components of Knowledge

In the section above 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 the model 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 $wc(A) = w(A)$, $sc(A) = \text{sim}(\mathcal{F}_{\mathcal{H}}(A), \mathcal{F}_{\mathcal{O}}(A))$, and $pc(A) = P_{\mathcal{H}} \cdot P_{\mathcal{H}, \mathcal{F}_{\mathcal{O}}}(A = v)$; otherwise, $wc(A) = sc(A) = pc(A) = 1$, i.e., wc , sc , and pc are omitted in the calculation. The quality of a model will be computed by the following functions $\|\cdot, \cdot\|$ and $\|\cdot\|$:

$$\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\| = \sum_{A \in \alpha(\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}})} wc(A) \cdot pc(A) \cdot sc(A), \quad \|\mathcal{F}\| = \sum_{A \in \mathcal{F}} wc(A) \quad (11)$$

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

3.5 An Example

For clarifying the equations defined for static covering models we will illustrate the computation of the covering rate by an example. The covering model is given in Figure 3, 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 $\text{sim}(c, c') = 0.7$ and $P_{D_1} = P_{D_2} = 0.5$. It holds that $P_{\{D_1\}} = P_{\{D_2\}} = P_{\{D_1, D_2\}} = 0.25$.

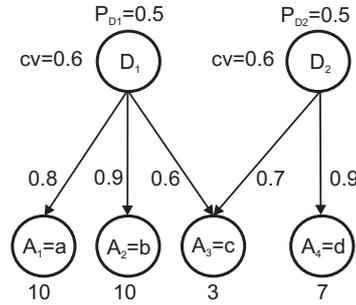


Figure 3: Covering model for the example. Probabilities for the covering relations are shown at the corresponding edges, weights are depicted under the findings.

We will consider the three hypotheses $\mathcal{H}_1 = \{D_1\}$, $\mathcal{H}_2 = \{D_2\}$ and $\mathcal{H}_3 = \{D_1, D_2\}$, which are sufficiently complete, since D_1 and D_2 exceed their specified completeness value $cv(D_1) = cv(D_2) = 0.5$.

To measure the *quality* of the hypotheses we have to calculate $\|\mathcal{F}_{\mathcal{H}}, \mathcal{F}_{\mathcal{O}}\|$ and $\|\mathcal{F}_{\mathcal{O}}\|$. $\|\mathcal{F}_{\mathcal{O}}\|$ is constant for all hypotheses:

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

Hypothesis $\mathcal{H}_1 = \{D_1\}$

We get

$$\begin{aligned}\alpha^+(\mathcal{F}_{\mathcal{H}_1}, \mathcal{F}_{\mathcal{O}}) &= \{A_1, A_2\}, \\ \alpha^-(\mathcal{F}_{\mathcal{H}_1}, \mathcal{F}_{\mathcal{O}}) &= \{A_3\}, \\ \alpha^u(\mathcal{F}_{\mathcal{H}_1}, \mathcal{F}_{\mathcal{O}}) &= \{A_4\}\end{aligned}$$

$$\begin{aligned}\|\mathcal{F}_{\mathcal{H}_1}, \mathcal{F}_{\mathcal{O}}\| &= P_{\mathcal{H}_1} \cdot \sum_{A \in \{A_1, A_2, A_3\}} w(A) \cdot P_{\mathcal{H}_1, \mathcal{F}_{\mathcal{O}}}(A = v) \cdot \text{sim}(\mathcal{F}_{\mathcal{H}_1}(A), \mathcal{F}_{\mathcal{O}}(A)) = \\ &= P_{\mathcal{H}_1} \cdot \left(\underbrace{10 \cdot 0.8 \cdot 1.0}_{A_1} + \underbrace{10 \cdot 0.9 \cdot 1.0}_{A_2} + \underbrace{3 \cdot 0.6 \cdot 0.7}_{A_3} \right) = P_{\mathcal{H}_1} \cdot 18.26\end{aligned}$$

$$\text{Therefore } \varrho(\mathcal{F}_{\mathcal{H}_1}, \mathcal{F}_{\mathcal{O}}) = P_{\mathcal{H}_1} \cdot 0.61 = 0.15$$

Hypothesis $\mathcal{H}_2 = \{D_2\}$

We get

$$\begin{aligned}\alpha^+(\mathcal{F}_{\mathcal{H}_2}, \mathcal{F}_{\mathcal{O}}) &= \{A_4\}, \\ \alpha^-(\mathcal{F}_{\mathcal{H}_2}, \mathcal{F}_{\mathcal{O}}) &= \{A_3\}, \\ \alpha^u(\mathcal{F}_{\mathcal{H}_2}, \mathcal{F}_{\mathcal{O}}) &= \{A_1, A_2\}\end{aligned}$$

$$\|\mathcal{F}_{\mathcal{H}_2}, \mathcal{F}_{\mathcal{O}}\| = P_{\mathcal{H}_2} \cdot \left(\underbrace{3 \cdot 0.7 \cdot 0.7}_{A_3} + \underbrace{7 \cdot 0.9 \cdot 1.0}_{A_4} \right) = P_{\mathcal{H}_2} \cdot 7.77$$

$$\text{Therefore } \varrho(\mathcal{F}_{\mathcal{H}_2}, \mathcal{F}_{\mathcal{O}}) = P_{\mathcal{H}_2} \cdot 0.26 = 0.06$$

Hypothesis $\mathcal{H}_3 = \{D_1, D_2\}$

We get

$$\begin{aligned}\alpha^+(\mathcal{H}_3, \mathcal{F}_{\mathcal{O}}) &= \{A_1, A_2, A_4\}, \\ \alpha^-(\mathcal{H}_3, \mathcal{F}_{\mathcal{O}}) &= \{A_3\}, \\ \alpha^u(\mathcal{H}_3, \mathcal{F}_{\mathcal{O}}) &= \{\}\end{aligned}$$

Since the parameter A_3 is covered by both diagnoses we cannot simply read off the probability but have compute it by assuming the events (D_1) , (D_2) and $(D_1 \wedge D_2)$ to cause the occurrence of A_3 . Applying Equation 7 we receive:

$$P_{\mathcal{H}_3}(A_3 = c) = \underbrace{0.6 \cdot 0.3}_{\{D_1\}} + \underbrace{0.6 \cdot 0.7}_{\{D_2\}} + \underbrace{0.18 \cdot 0.28}_{\{D_1, D_2\}} = 0.88$$

$$\begin{aligned}\|\mathcal{F}_{\mathcal{H}_3}, \mathcal{F}_{\mathcal{O}}\| &= P_{\mathcal{H}_3} \cdot \left(\underbrace{10 \cdot 0.9 \cdot 1.0}_{A_1} + \underbrace{10 \cdot 0.8 \cdot 1.0}_{A_2} + \underbrace{3 \cdot 0.88 \cdot 0.7}_{A_3} + \underbrace{7 \cdot 0.9 \cdot 1.0}_{A_4} \right) = \\ &= P_{\mathcal{H}_3} \cdot 0.84\end{aligned}$$

$$\text{Therefore } \varrho(\mathcal{F}_{\mathcal{H}_3}, \mathcal{F}_{\mathcal{O}}) = P_{\mathcal{H}_3} \cdot 0.84 = 0.21$$

It is easy to see that the hypothesis $\mathcal{H}_3 = \{D_1, D_2\}$ 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 viewed covering relations as static representations of diagnostic models. Now we will introduce a concept of cause–effect relations that generates covering models dynamically. This approach will allow for integrating therapies and their effects into covering models.

Up to now we can only represent covering relations for a single severity of a diagnosis. This limitation restricts the modeling facilities in several ways:

1. To build a covering model for each diagnosis and each severity implies a huge effort of knowledge acquisition for the expert.
2. It is impossible to model therapy effects on parameter values, since therapy effects weaken or strengthen existing parameter values. This behaviour cannot be modelled with the static covering models described so far.

For this reason we will introduce *severities* and *severity effects* as an extension of the existing covering models. Diagnoses now can not only have the state “true” or “false”, but can hold severities denoted in general by the set $\Omega_{sev} = \{0, sev_1, \dots, sev_v\}$ of all possible severities: 0 means that the diagnosis is absent, sev_v is the strongest occurrence of the diagnosis. In the following we will treat therapies like diagnoses, since they also can hold different severities (dosages) and have an effect on parameter values.

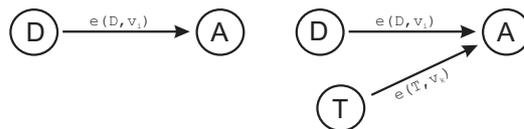


Figure 4: A single (only diagnosis D as a cause) and a binary effect (diagnosis D and therapy T as causes) on parameter A

For implementing these capabilities we want to give an intuitive definition for an effect function e . It will be used to express effects from diagnoses/therapies to findings. An important convention is that we only can define effect functions on findings with a scalar range of values. So the values contained in the range need to have an explicit order. For simplicity we define an abstract set $\mathcal{E} = \{-, -, -, -, -, 0, +, ++, +++\}$ of symbols for possible effects. For example the symbol “-” means a slight decrease of the involved finding, whereas “+” defines an increase, “0” maps no effect to the existing value; the other symbols are gradually stronger or weaker movements of the findings.

As shown in Figure 4 effect functions are denoted at the edges between the corresponding nodes. With the implementation of effects for covering models the following problems arise:

1. Interpretation: In principle the symbolic effects in \mathcal{E} are not clearly defined. Nevertheless, we need a precise value to compute the quality measures. But the effect (e.g., slight decrease of the parameter value) can be uncertain and the movement of the value may result in different parameter values with different probabilities.
2. Accounting: If a parameter receives more than one effect, then the question arises, how these multiple effects should be accounted against each other to compute the resulting parameter value.

When proposing solutions for the problems above we have to keep in mind that the representation has to stay as simple as possible to provide an easy knowledge acquisition for the expert.

4.1 Interpretation

The simplest way to translate the effects into concrete findings is to define a categorical transformation between the symbols of set \mathcal{E} to an integer. For example we can declare the symbol “+” to stand for an increase of the finding by +1 unit. Then a parameter A starting with value v_j and receiving an single effect “+” would get the new value v_{j+1} . In the covering model this will result in the finding ($A = v_{j+1}$).

For an uncertain interpretation of the effect we will introduce a membership function $\mu : \mathcal{E} \rightarrow \mathcal{T}$ for each symbol in \mathcal{E} . \mathcal{T} is defined as a set of tuples $[mov, p(mov)]$, where mov is an integer indicating how much the value of the parameter will be moved (increased/decreased) and $p(mov)$ the probability of this movement.

$$\mu(+) = \{ (0, 0.1), (+1, 0.8), (+2, 0.1) \} \quad (12)$$

In Equation 12 we can see an example for the symbol “+”, which denotes, that we will have no movement at all with 10% probability, a shift of the parameter value by +1 with 80% probability and an increase by +2 with 10% probability.

4.2 Accounting of Multiple Effects

The accounting of multiple effects on a parameter is not obvious and we will present two options which are appropriate for different kinds of situations. When bringing the effects into the covering relations the modeler has to decide about how the multiple influences on a node should be handled and denote it at the influenced node. We emphasize that these procedures are quite simple but keep the model simple as well and the reasoning step understandable for the user. The two options are:

Order-of-magnitude reasoning. In [8] this procedure is motivated by the fact that larger magnitudes of effects might have a significantly larger impact on a finding. E.g., if the diagnosis *pcP* increases the value of parameter *inflammations* by “+++” as well as another diagnosis increases *inflammation* by “+”, then the second diagnosis may not have any additional significance on the value of the parameter.

For a *certain* interpretation of the symbols we only count the largest magnitude when all effects have the same sign. For an *uncertain* interpretation of the involved effects we have to transform the symbols and calculate the possible events. E.g., w.r.t. the binary effect in Figure 4 we assume that “ $e(D, A) = ++$ ” and “ $e(T, A) = +$ ”. The symbols “++” and “+” are defined by the following equations:

$$\begin{aligned} \mu(++) &= \{ (+1, 0.1), (+2, 0.8), (+3, 0.1) \}, \\ \mu(+) &= \{ (0, 0.1), (+1, 0.8), (+2, 0.1) \}. \end{aligned} \quad (13)$$

It is easy to see that the possible movements of A caused by these two effects are defined by $Mov = \{+1, +2, +3\}$. For the moment, we only want to compute the probability that A is increased by +2. To simplify the calculations we abbreviate the following events with e_i :

$$\begin{aligned} e_1 &= [e(D, A) = (+2, 0.8)], & e_3 &= [e(T, A) = (+1, 0.8)], \\ e_2 &= [e(T, A) = (\pm 0, 0.1)], & e_4 &= [e(T, A) = (+2, 0.1)]. \end{aligned}$$

A is shifted by $+2$ because of the events $e_1 \wedge e_2$ or $e_1 \wedge e_3$ or $e_1 \wedge e_4$. For the overall probability of $e_{mov=+2}$ we compute

$$p(e_{mov=+2}) = e_1e_2 + e_1e_3 + e_1e_4 = 0.08 + 0.64 + 0.08 = 0.8.$$

The calculation of the probabilities for the events $e_{mov=+1}$ and $e_{mov=+3}$ is analogous. A problem arises, when we have two effects with different signs but equal magnitude (e.g., $+$ and $-$). In this case a general solution is not possible and the modeler has to define a reasonable heuristic.

Accumulative reasoning. If a parameter is influenced not only by diagnoses but also by therapies, then an accumulative approach seems to be appropriate. Here we compute a final effect on the parameter by transforming each symbolic influence into its numeric representation. Then we simply obtain the result by the sum of the numeric values. As this procedure will work fine for categorical effects we have to define this for uncertain effects more precisely. The idea is that we have to compute the cartesian product to get all possible outcomes for an influenced finding.

An example might clarify the procedure: Recalling the binary effect depicted in Figure 4 and assuming that “ $e(D, A) = ++$ ” and “ $e(T, A) = -$ ”, the symbols “ $++$ ” and “ $-$ ” are defined as follows:

$$\begin{aligned} \mu(++) &= \{(+1, 0.1), (+2, 0.8), (+3, 0.1)\}, \\ \mu(-) &= \{(-2, 0.1), (-1, 0.8), (0, 0.1)\} \end{aligned} \quad (14)$$

Computing these two effects we can see that the set of possible movements of parameter A is defined as $Mov = \{-1, 0, +1, +2, +3\}$. If we want to know the probability of the event $e_{mov=0}$ (“ A does not move its value ($mov = 0$)”), we simply have to consider the two events $e_1 = [(T = -1) \wedge (D = 1)]$ and $e_2 = [(T = -2) \wedge (D = 2)]$. Notice that the sub-events of e_1 and e_2 are independent (see assumptions in Section 3.3), and so we compute

$$p(e_{mov=0}) = e_1 + e_2 = (0.8 \cdot 0.1) + (0.1 \cdot 0.8) = 0.16.$$

Thus, the covering relation $r = [D \wedge T] \rightarrow [A = 0]$ has the probability $p(r) = 0.16$. The calculations for the other values contained in Mov are analogous.

We have shown how to integrate cause-effect relations into set-covering relations. W.r.t. the observations it is simple to compute the probabilities for the observations and to compute the quality measures for this generated model as described in Section 3.

5 Summary and Future Work

In this paper we 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 covering net 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. Currently we are implementing a prototype of the problem-solver which will be used by experts in the fields of medicine and geo-ecology.

So far we have only considered models containing one level of covering relations (e.g. $D \rightarrow (A = v)$). In the future we want to consider multiple levels since

we would like to represent covering relations like $D_i \rightarrow D_k \rightarrow (A = v)$. One arising question is the handling of uncertainty in multi level nets. We are planning to prove that an intuitive accounting of probabilities can be seen as an approximation of the inference methods used in belief networks [6].

References

- [1] 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 Publishers, 1988.
- [2] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Mateo, California, 2000.
- [3] William J. Long. Medical diagnosis using a probabilistic causal network. *Applied Artificial Intelligence*, 3:367–383, 1989.
- [4] P. Lucas, A. Tholen, and G. van Oort. An intelligent system for pacemaker reprogramming. *Artificial Intelligence in Medicine*, 17:249–269, 1999.
- [5] 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, Boulder, Colorado, 1982.
- [6] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, California, 1988.
- [7] Frank Puppe. Knowledge reuse among diagnostic problem-solving methods in the Shell-Kit D3. *Int. J. Human-Computer Studies*, 49:627–649, 1998.
- [8] Olivier Raiman. Order of magnitude reasoning. *Artificial Intelligence*, 51:11–38, 1991.
- [9] J.A. Reggia and D.S. Nau. *Abductive Inference Models für Diagnostic Problem-Solving*. Springer, Berlin, 1990.
- [10] J.A. Reggia, D.S. Nau, and P.Y. Wang. Diagnostic expert systems based on a set covering model. *Journal of Man-Machine Studies*, 19:437–460, 1983.
- [11] Andreas Warnke. Dynamic Belief Networks für zeitabhängige Therapieprozesse. Master’s thesis, Wuerzburg University, Department of Computer Science VI, 2001.