# Exploiting Structural Metrics in FMEA-Based Abductive Diagnosis

**Roxane Koitz**[*1] and **Franz Wotawa**[1]
[1]Graz University of Technology, Graz, Austria
e-mail: {rkoitz,wotawa}@ist.tugraz.at

## Abstract

Abductive model-based diagnosis provides an intuitive approach to fault identification by reasoning on a description of the system to be diagnosed. Nevertheless, its computational complexity hinders a vast adoption and thus motivates further evaluation of efficient methods. In this paper, we investigate the structural metrics inherent to models and diagnosis problems generated on the basis of Failure Mode Effect Analysis (FMEA). Proceeding on the metrics developed, we investigate their potential as classification features to identify the most suitable diagnosis algorithm for a particular diagnosis problem. Evaluated on artificial and practical samples, our approach shows that the classifier trained on the described metrics is able to indicate the most efficient method in case of a specific diagnosis scenario.

## 1 Introduction

Growing complexity of technical systems complicates an effective as well as efficient fault identification, causing automated diagnosis to be of increasing interest from a theoretical as well as applied point of view. An extensive body of research has concerned itself with model-based diagnosis [Reiter, 1987; de Kleer and Williams, 1987] which reasons on causes for observed anomalies using a description of the system. Depending on the representation two main variations can be distinguished, namely consistency-based and abductive model-based diagnosis. While the former operates on formalizations of the intended behavior, the latter exploits a model of how failures manifest themselves within the system [Console *et al.*, 1991]. By relying on the concept of entailment, abductive reasoning provides consistent root causes for failure indicators in an intuitive way.

Abduction is not merely relevant in the field of diagnostics, but has been applied to diverse fields such as planning [Poole and Kanazawa, 1994] or ontology debugging [Lambrix *et al.*, 2013]. Various approaches to compute abductive explanations have been developed over the last decades, such as set covering [Patil *et al.*, 1982; Guan and Jiang, 2013], abductive logic programming [Kakas *et al.*, 1992], proof-tree completion [McIlraith, 1998], or consequence finding [Marquis, 2000]. Within this paper we address two methods: abductive model-based diagnosis and the parsimonious set covering theory. In the latter a simple diagnosis problem comprises a set of causes, manifestations, and a causal associative network connecting these disjoint sets. A diagnosis is defined as the disorders which cover, i.e. explain, a given set of symptoms. The parsimonious set covering approach has been formalized and later extended to include Bayesian probabilities [Peng and Reggia, 1990]. Several refinements to the basic theory have been proposed such as the improvement of models with additional knowledge or the inclusion of more complex covering relations [Baumeister and Seipel, 2002].

Abduction is a hard problem with an exponential number of solutions in the worst case. Even though certain model representations are tractable [Nordh and Zanuttini, 2008], computing solutions for instances of reasonable size and complexity remains a challenge. Therefore, in this paper we investigate the algorithm selection problem [Rice, 1976] for abductive diagnosis. Algorithm selection addresses the issue of choosing the best performing method for a particular problem instance and advocates for the importance of structural properties of the problem space to determine the preferred approach [Smith-Miles, 2009]. The basic building blocks within this framework are a portfolio of algorithms to choose from, empirical performance data of the algorithms on representative problems and a set of features, which are used to get a notion of the difficulty of a problem instance [Hutter *et al.*, 2006]. Leyton-Brown et al. [2003] describe their portfolio approach to algorithm selection, where they train a classifier for each algorithm within their portfolio to forecast each approach's computation time on the instance and execute the one predicted as most efficient. Machine learning has been identified as a feasible approach to use as a prediction tool. Algorithm selection has been applied for example to SAT [Xu *et al.*, 2008], graph coloring problems [Musliu and Schwengerer, 2013] or tree-decomposition [Morak *et al.*, 2012].

Our problem space is restricted to propositional Horn clause models generated from failure assessments available in practice. Failure Mode Effect Analysis (FMEA) is a widespread fault evaluation tool and captures the way component failures affect the system, thus provides an ideal knowledge source for abductive models. The formal descriptions of these models present certain structural traits, which are used as features in the algorithm selection process. Based on these model attributes and a set of experiments, a machine learning classifier was trained to decide on the most run time efficient abductive reasoning algorithm

[*]Authors are listed in alphabetical order.

for a distinct diagnosis problem. We embedded the selection process within a meta-algorithm, which generates the structural metrics for a given diagnosis scenario, categorizes it on the previously trained classifier and computes the diagnoses using the "best" algorithm according to the prediction.

This paper is organized as follows. In Section 2 we provide background information on logic-based abduction and the set covering approach as our algorithmic methods to abduction. Afterwards, we describe the mapping from an FMEA to a model which can be used in abductive reasoning and the structural metrics inherent to this representation. Section 4 describes the meta-approach and its evaluation on two sample sets. Finally, we conclude and give an outlook on future work.

## 2 Abductive Diagnosis

Within this section we present two abductive diagnosis methods, namely abductive model-based diagnosis building on propositional Horn clauses and the simple set covering approach. We show that these two formalizations in their restricted form can be interchanged.

### 2.1 Model-Based Diagnosis

In abductive model-based diagnosis a system description holds information on how failures affect system variables. On basis of the knowledge and given an observable anomaly, the task is to search for a set of causes, which together with the model logically entail the observation. Furthermore, the explanations have to be consistent with the underlying theory. Since abductive inference is an intractable problem, research has focused on logical subsets which allow to compute explanations in polynomial time [Eiter and Gottlob, 1995]. Based on these findings we focus on a propositional Horn clause formalization [Friedrich *et al.*, 1990] .

**Definition 1 (Knowledge base (KB))** *A knowledge base (KB) is a tuple (A,Hyp,Th) where A denotes the set of propositional variables, Hyp $\subseteq$ A the set of hypotheses, and Th the set of Horn clause sentences over A.*

A knowledge base provides the underlying model for our reasoning, where the set of hypotheses comprises the propositional variables constituting a cause or explanation. The remaining propositional variables. i.e. $\{A \setminus Hyp\}$, are effects or symptoms and the theory determines the relations between hypotheses and effects.
*Example 1:* The following set of hypotheses, propositional variables, and Horn clauses form an abductive *KB*:

$$Hyp = \{h_1, h_2, h_3\}, A = \{h_1, h_2, h_3, o_1, o_2, o_3\},$$

$$Th = \left\{ \begin{array}{c} h_1 \rightarrow o_1, h_2 \rightarrow o_1, h_2 \rightarrow o_2, \\ h_3 \rightarrow o_2, h_3 \rightarrow o_3 \end{array} \right\}$$

An abduction problem considers a knowledge base *KB* and a set of observations, i.e. current symptoms, for which explanations should be computed.

**Definition 2** ***(Propositional Horn Clause Abduction Problem (PHCAP))*** *Given a knowledge base (A,Hyp,Th) and a set of observations Obs $\subseteq$ A then the tuple (A,Hyp,Th,Obs) forms a Propositional Horn Clause Abduction Problem (PHCAP).*

**Definition 3 (Diagnosis; Solution of a PHCAP)** *Given a PHCAP (A,Hyp,Th,Obs). A set $\Delta \subseteq$ Hyp is a solution if and only if $\Delta \cup$ Th $\models$ Obs and $\Delta \cup$ Th $\not\models \bot$. A solution $\Delta$ is parsimonious or minimal if and only if no set $\Delta' \subset \Delta$ is a solution.*

A solution to the PHCAP is an abductive diagnosis, as it provides hypotheses consistently explaining the occurrence of an observation.
*Example 1 (Cont.):* Let us assume we can observe $o_1$ and $o_3$, i.e. $Obs = \{o_1, o_3\}$. The solution to the PHCAP, i.e. the abductive diagnoses, are $\Delta_1 = \{h_1, h_3\}$ and $\Delta_2 = \{h_2, h_3\}$.

An Assumption-Based Truth Maintenance Systems (ATMS) [de Kleer, 1986a; 1986b] is able to compute abductive explanations. Based on a directed graph, where propositional variables are nodes and the edges are determined by the theory, each node owns a label which stores the hypotheses implying said node. Whenever a new clause is added to the theory, the ATMS updates its nodes' labels and further maintains a consistent state. By adding a rule of the type $o_1 \wedge \ldots \wedge o_n \rightarrow explain$ containing all elements of $Obs$ on the left hand side and a new variable not previously contained in $A$, e.g. $explain$, on the right hand side, the ATMS computes the abductive solution as the label of $explain$.

### 2.2 Set Covering

Peng and Reggia [1990] developed the parsimonious set covering theory as a formal approach to abductive diagnosis relying on an associative network of causal connections between disorders and manifestations. A simple diagnosis problem is defined as a 4-tuple $P =< D, M, C, M^+ >$, where $D$ is the set of disorders, $M$ denotes the set of manifestations, $C$ describes the relations in the causation network and $M^+$ comprises the current set of observations. Considering the definitions of the previous subsection $D$ refers to $Hyp$. Further, the manifestations $M$, describe the remaining propositions not included within the hypotheses which are in fact the effects. The causal relations $C$ are given by the theory, i.e. there exists a relation between a disorder $d_i$ and a manifestation $m_j$ whenever there is a clause $d_i \rightarrow m_j$ contained within the theory. Since $M^+$ provides the distinguished subset of symptoms observed it corresponds to $Obs$. As the mapping between a PHCAP and a set covering problem is straightforward, we will use the wording as defined in the previous subsection, e.g. $Hyp$ refers to the set of hypotheses, causes, disorders, etc. For clarity we add one additional set missing from the logic-based framework, i.e. $M$, which we define as $\{A \setminus Hyp\}$. The simple set covering model is equivalent to logic-based abduction with a theory restricted to definite Horn clauses [McIlraith, 1998].

*Example 1 (Cont.):* Considering our example from before, the diagnosis problem can be reduced to a set covering problem:

$$Hyp = \{h_1, h_2, h_3\}, M = \{o_1, o_2, o_3\}, Obs = \{o_1, o_3\}$$

$$Th = \left\{ \begin{array}{c} < h_1, o_1 >, < h_2, o_1 >, < h_2, o_2 >, \\ < h_3, o_2 >, < h_3, o_3 > \end{array} \right\}$$

In order to define a solution to a diagnosis problem within this framework, we define for every hypothesis the set *effects*$(h_i) = \{m_j | < h_i, m_j > \in Th\}$, i.e. the set of objects directly caused by $h_i$, and respectively for each effect $m_j$, the set *causes*$(m_j) = \{h_i | < h_i, m_j > \in Th\}$, i.e. the set of objects which can directly cause $m_j$ [Peng and Reggia, 1990]. Thus, for any subset of disorders $Hyp_I$, we can determine the objects directly caused by it as

$$\mathit{effects}(Hyp_I) = \bigcup_{h_i \in Hyp_I} \mathit{effects}(h_i)$$

Along similar lines, we can observe that

$$\mathit{causes}(M_J) = \bigcup_{m_j \in M_J} \mathit{causes}(m_j)$$

for a set of manifestations $M_J \subseteq M$.

*Example 1 (Cont.):* For example $\mathit{causes}(o_1) = \{h_1, h_2\}$ and $\mathit{effects}(\{h_1, h_2\}) = \{o_1, o_2\}$.

**Definition 4 (Cover)** *A set $Hyp_I \subseteq Hyp$ is said to cover $M_J \subseteq M$ if $M_J \subseteq \mathit{effects}(Hyp_I)$ and there exists no $Hyp'_I \subset Hyp_I$ with $M_J \subseteq \mathit{effects}(Hyp'_I)$.*

A cover relation exists between a disorder and a manifestation whenever the latter is causally inferred from the former and is a minimal subset. While minimality is not a necessary condition for a cover in the original definition, we enforce this parsimonious criteria as we are only interested in minimal diagnoses [Peng and Reggia, 1986].

**Definition 5 (Set Cover Diagnosis)** *Given a diagnosis problem $P$. A set $\Delta \subseteq Hyp$ is said to be a diagnosis iff $\Delta$ covers Obs.*

*Example 1 (Cont.):* In case we have the same observations as before, i.e. $o_1$ and $o_3$, we can obtain the set covering diagnoses by determining the disorder sets $Hyp_I \subset Hyp$ where $\mathit{effects}(Hyp_I)$ cover *Obs*, which are $\mathit{effects}(\{h_1, h_3\}) = \{o_1, o_2, o_3\}$ and $\mathit{effects}(\{h_2, h_3\}) = \{o_1, o_2, o_3\}$. Hence, the diagnoses are $\Delta_1 = \{h_1, h_3\}$ and $\Delta_2 = \{h_2, h_3\}$.

As it has been shown previously, set covering is equivalent to the hitting set problem [Karp, 1972]. Since a cover states that a certain disorder causally infers a manifestation, we can utilize the set $\mathit{causes}(m_j)$ as previously defined as a similar cover indicator. For each manifestation the set $\mathit{causes}(m_j)$ contains the information on all disorders causing $m_j$. By computing the hitting set of $\mathit{causes}(m_j)$ we derive a disjunction of all disorders included, i.e. each cause constitutes a possible solution. In case we obtain a set of observable manifestations $m_1, \ldots, m_n \in \mathit{Obs}$, the hitting sets of all $\mathit{causes}(m_j) \in \mathit{Obs}$ comprises the diagnoses. This is apparent, as to account for all current manifestations one disorder causing each manifestation has to be present within a single solution. Again we focus on parsimonious solutions, therefore we are solely interested in subset minimal hitting sets (MHS).

**Definition 6 (Abductive Hitting Set Diagnosis)** *Given a diagnosis problem $P$. A set $\Delta \subseteq Hyp$ is said to be a minimal diagnosis iff $\Delta$ is a MHS of S, where $\forall m_j \in \mathit{Obs} : \mathit{causes}(m_j) \in S$.*

*Example 1 (Cont.):* The *causes* sets for the current manifestations are $\mathit{causes}(o_1) = \{h_1, h_2\}$ and $\mathit{causes}(o_3) = \{h_3\}$, thus $S = \{\{h_1, h_2\}, \{h_3\}\}$. The MHSs of $S$ correspond to $\Delta_1$ and $\Delta_2$.

While Peng and Reggia [1990] define a minimum diagnosis as the solution with the smallest cardinality, we choose to use subset minimality as our parsimonious criteria, which is referred to as an irredundant cover within their formalism. Note that one difference exists in the semantic view of a PHCAP and a set covering diagnosis problem; in the PHCAP we assume that a clause represents a necessary causations between a hypothesis and an effect, whereas the association relation only characterizes a causation with probability $0 < p < 1$.

## 3 Structural Analysis

Since developing suitable system models for model-based diagnosis is a tedious task, there are approaches taking advantage of knowledge available to automatically generate system descriptions [Sterling *et al.*, 2014]. A recent approach utilizes the information captured within FMEAs as the basis of their abduction models [Wotawa, 2014]. FMEA is of increasing interest as a systematic assessment of reliability on a component level [Catelani *et al.*, 2010]. It encapsulates possible faults as well as the way they reveal themselves in different system variables [Hawkins and Woollons, 1998]. Therefore, it provides information on causal relations between failures and their symptoms, which in turn can be used as a knowledge base in an abductive diagnosis context [Wotawa, 2014]. The mapping proposed assumes that intermediate effects are not essential to the analysis and in general we consider a closed-world assumption.

In this section, we first describe the mapping from an FMEA to a definite propositional Horn clause theory suitable for abductive reasoning and then present the structural metrics utilized as attributes for classification in our meta-approach.

### 3.1 FMEA-Based Model Development

An FMEA can be defined as a set of tuples consisting of components $COMP$, fault modes $MODES$ and manifestations, which are a subset of the propositional variables $PROPS$ [Wotawa, 2014]. Table 1 illustrates an exemplary FMEA taken from the domain of industrial wind turbines, featuring three component-fault mode pairs and their indicators.

| **Component** | **Fault Mode** | **Effect** |
|---|---|---|
| Fan | Corrosion | P_turbine |
| Fan | TMF | P_turbine, T_cabinet |
| IGBT | HCF | T_cabinet, T_nacelle |

Table 1: *Example 2*: FMEA taken from the wind turbine domain.

**Definition 7** *An FMEA is a set of tuples $(C, M, E)$ where $C \in COMP$ is a component, $M \in MODES$ is a fault mode, and $E \subseteq PROPS$ is a set of effects.*

Since each component-fault mode pair $(C, M)$ constitutes a possible cause for the corresponding symptoms, a proposition $mode(C, M)$ is generated for each pair and included within the set $Hyp$. These hypotheses as well as the effects form the set of propositional variables $A$.

$$Hyp =_{def} \bigcup_{(C,M,E) \in FMEA} \{mode(C, M)\}.$$

$$A =_{def} \bigcup_{(C,M,E) \in FMEA} E \cup \{mode(C, M)\}.$$

A mapping function $\mathfrak{M} : 2^{FMEA} \mapsto HC$ constructs from each tuple of the FMEA a definite propositional Horn clause, i.e. for each record of the FMEA a rule is generated where the component-fault mode pair represents the antecedent while the corresponding effect denotes the consequence.

**Definition 8** *Given an FMEA, the function $\mathfrak{M}$ is defined as follows:*

$$\mathfrak{M}(FMEA) =_{def} \bigcup_{t \in FMEA} \mathfrak{M}(t)$$

*where*

$$\mathfrak{M}(C, M, E) =_{def} \{mode(C, M) \rightarrow e \,|e \in E\}\,.$$

*Example 2 (Cont.):* For the example given in Table 1 we derive the following set of hypotheses, variables and Horn clauses:

$$Hyp = \left\{ \begin{array}{c} mode(Fan, Corrosion), \\ mode(Fan, TMF), mode(IGBT, HCF) \end{array} \right\}$$

$$A = \left\{ \begin{array}{c} mode(Fan, Corrosion), T\_cabinet, \\ P\_turbine, \ldots \end{array} \right\}$$

$$Th = \left\{ \begin{array}{c} mode(Fan, Corrosion) \rightarrow P\_turbine, \\ mode(Fan, TMF) \rightarrow P\_turbine, \\ mode(Fan, TMF) \rightarrow T\_cabinet, \\ mode(IGBT, HCF) \rightarrow T\_cabinet, \\ mode(IGBT, HCF) \rightarrow T\_nacelle \end{array} \right\}$$

### 3.2 Structural Metrics

Analyzing the structure of the theory, we can observe that the models constructed on basis of FMEAs with $\mathfrak{M}$ are characterized by bijunctive definite Horn clauses. Hence, we can easily represent the theory as an acyclic directed graph (DAG) with a forward structure from causes to effects. This structure is in fact equivalent to the problems in the simple set covering theory where hypotheses and manifestations are disjoint sets. The attentive reader might have observed that *Example 1* and *Example 2* are equivalent problems, with $mode(Fan, Corrosion) \equiv h_1, mode(Fan, TMF) \equiv h_2, mode(IGBT, HCF) \equiv h_3, P\_turbine \equiv o_1, T\_cabinet \equiv o_2, T\_nacelle \equiv o_3$. Thus, from this point forward we utilize the formalization featured in the first example due to readability.

Based on the theory, we can easily represent the model as a hypergraph $H = (V, E)$, where $V$ is the set of vertices and $E$ denotes the set of hyperedges $e$ with $e \subseteq V$. Concerning our models the node set comprises all propositional variables, while hyperedges are determined by the theory; for each clause there exists a hyperedge containing the propositional variables of the clause, i.e. $\forall a \in A \rightarrow a \in V$ and $\forall c \in Th \rightarrow \bigcup_{l \in c} |l| \in E$ where $||$ is a function mapping literals to the underlying propositions ignoring negations, i.e., $|\neg p| = p$ and $|p| = p$ for all $p \in A$.

Figure 1 shows the hypergraph built for *Example 1*. Following this representation we can assign a label to each vertex within a hyperedge $E$, such that:

$$label(v) = \begin{cases} \{v\} & \text{if } v \in Hyp \\ \bigcup\limits_{x \in E \wedge x \neq v} label(x) & \text{otherwise} \end{cases}$$

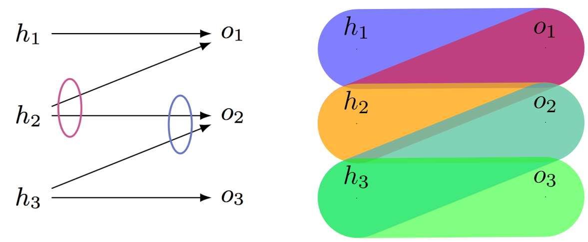



Figure 1: DAG and hypergraph representation of *Example 1*. The DAG shows the metrics effect overlap (left oval) and hypotheses covering (right oval) for pairs of nodes.

Obviously, in case we examine the vertices representing manifestations, the labels correspond to the *causes*-sets, since both contain the hypotheses inducing the corresponding manifestation. Hence, we can utilize the label of the current observations in computing the abductive diagnoses by means of hitting sets. Note, that with this representation we can handle intermediate effects, i.e. manifestations leading to additional symptoms, which are not represented in the simple set covering theory. In this paper, however, we focus on the simple structure as inherent to models created from FMEAs.

On top of both graph representations of the models, we can extract certain metrics relating to the underlying structure. An intuitive measure is the number of causes in the model, since abductive reasoning is exponential in the size of $Hyp$. Furthermore, basic quantities include the number of effects and relations occurring in the model.

**Outdegree and Indegree**

Considering the directed graph determined by the theory we can compute the outdegree of each hypothesis node, i.e. the number of effects caused by said node, as well as the indegree of each effect, i.e. the number of hypotheses implying said manifestation. Using the set covering formalization we can define

$$outdegree(h_i) = |effects(h_i)|$$

and similarly

$$indegree(m_j) = |causes(m_j)|.$$

Collected over the entire model, these measures provide an intuitive metric of the basic magnitude of the theory and the connectedness of the graph.

**Covering and Overlap**

As can be seen from Figure 1 there are several disorders covering the same effect, that is a single observation can be explained by multiple causes. We can easily determine a covering metric for any given pair of hypotheses by building the ratio between common manifestations and the total number of manifestations caused by the hypotheses:

$$covering(h_i, h_j) = \frac{|effects(h_i) \cap effects(h_j)|}{|effects(h_i) \cup effects(h_j)|}$$

In a similar manner, we define the overlap of two effects as their shared causes in relation to all their causes.

$$overlap(m_i, m_j) = \frac{|causes(m_i) \cap causes(m_j)|}{|causes(m_i) \cup causes(m_j)|}$$

**Algorithm 1** MetAB

```
procedure MetAB (A, Hyp, Th, Obs)
    m ← retrieveClassifier()                                  ▷ Retrieves trained model
    φ_offline ← retrieveMetrics(A, Hyp, Th)                   ▷ Retrieves the previously computed model metrics
    φ_online ← computeMetrics(A, Hyp, Th, Obs)                ▷ Computes the instance-based features
    φ ← φ_offline ∪ φ_online
    algorithm ← predict(φ, m)                                 ▷ Forecasts the best performing algorithm for the diagnosis problem
    Δ-Set ← diagnose(algorithm, A, Hyp, Th, Obs)              ▷ Computes diagnoses based on the predicted algorithm for the PHCAP
    return Δ-Set
end procedure
```

In case there is a unique hypothesis explaining an observation, this is referred to as a pathognomonic effect [Peng and Reggia, 1990]. Whenever a pathognomonic symptom is involved, we cannot compute an overlap relation. In Figure 1 the red oval on the left shows how $o_1$ and $o_2$ overlap at $h_2$, while the blue right oval marks the shared observation $o_2$ of $h_2$ and $h_3$. By collecting these measures for any pair of hypotheses or effects, we can compute these values over the entire model.

**Independent Diagnosis Subproblem**

Independent diagnosis subproblems occur whenever the directed graph and hypergraph are not connected, i.e. there exist subproblems within the model which have disjoint hypotheses and effect sets. Note that if all effects are pathognomonic, then each cause-effect relation represents its own diagnosis subproblem and thus the diagnosis model is orthogonal. Imagine the clause $h_2 \rightarrow o_2$ missing from the theory of *Example 1*. In this case we would have two independent diagnosis subproblems, namely one including $h_1, h_2$ and $o_1$ and the other one being $h_3, o_2$ and $o_3$.

**Path Length**

Another measure of connectedness within the model is the path length on the hypergraph. In particular, we measure the length of path between nodes representing hypotheses. Note, that for a model there are possibly several hypergraphs depending on the number of independent diagnosis subproblems, thus we disregard paths between nodes belonging to different diagnosis subproblems.

**Observation Dependent Metrics**

Since not only the topology of the model is of interest, but also the structure of the current diagnosis problem, we measure the overlap among the elements of $Obs$, the indegrees of the corresponding nodes and determine the number of diagnosis subproblems involved, in case several exist.

## 4 Algorithm Selection and Meta-Approach

Since general abductive reasoning belongs to the NP-hard problems, a research area of interest is to discover efficient methods to compute explanations. Thus, we investigate the feasibility to utilize the metrics defined in the previous section to train a classifier able to select the most time efficient approach for a particular diagnosis scenario. First formalized by Rice [1976], algorithm selection aims at identifying the "best performing" approach for a specific problem instance [Kotthoff, 2012]. Thereby, each problem can be described by a set of attributes which together with empirical execution data allows a trained predictor to forecast the most valuable algorithm on an instance.

The general idea of the meta-algorithm based on algorithm selection is straightforward. As mentioned the foundation of model-based diagnosis is a description of the system to be diagnosed. Thus, the majority of the features can be computed offline on the diagnosis models present. Further, within this phase the empirical data on computation times of the various abductive reasoning approaches can be collected and on basis of the metrics and the runtime information a machine learning classifier is trained.

Whenever the diagnosis process is triggered by a detected anomaly, we retrieve the previously trained classifier as well as the already computed metrics ($\phi_{offline}$), determine the remaining instance-based attributes ($\phi_{online}$) of the particular PHCAP and create the feature vector for classification determined by the measures. Based on these features and the machine learning classifier we in turn retrieve a predicted best algorithm for this scenario. Subsequently, we can instantiate the diagnosis engine with the corresponding abduction method as well as diagnosis problem and compute the abductive explanations. Algorithm 1 describes the online portion of the meta-approach, which is executed whenever new diagnoses are to be computed.

To evaluate the feasibility of the meta-technique we conducted experiments on two different data sets which we will explain in detail in the upcoming subsection. For the machine learning part of our meta-algorithm we employed the Waikato Environment for Knowledge Analysis (WEKA) library [Hall *et al.*, 2009], which provides a vast number of classification methods. The abductive reasoning algorithms forming our prediction categories are an ATMS as well as several hitting set algorithms, namely the Binary Hitting Set Tree (BHSTREE) [Lin and Jiang, 2003], HSDAG [Reiter, 1987], and HST [Wotawa, 2001]. To collect training and test data for classification, we exploited a Java implementation of an ATMS as well as BHSTREE[1] and Python implementations [Quaritsch and Pill, 2014][2] of the remaining hitting set methods.

### 4.1 Data

Our data for classification originate from two separate sources; on the one hand a small corpus of FMEAs and on the other hand generated artificial examples. The former comprises publicly available as well as internally used FMEAs recording fault knowledge from diverse domains. We automatically mapped these failure assessments to abductive knowledge bases, which we can use for logic-based as well as set covering abduction. All in all we conducted experiments on twelve FMEAs with various numbers of hypotheses ($4 \leq |Hyp| \leq 32$), effects ($5 \leq |M| \leq 30$) and

[1] http://www.ist.tugraz.at/modremas/index.html

[2] http://modiaforted.ist.tugraz.at/downloads/pymbd.zip

| Actual \ Predicted | HSDAG | BHSTREE | HST | ATMS | Total |
|---|---|---|---|---|---|
| **HSDAG** | 14/20 | 18/0 | 0/0 | 5/0 | 37/20 |
| **BHSTREE** | 1/5 | 220/14 | 0/0 | 16/0 | 267/19 |
| **HST** | 0/2 | 3/1 | 2/1 | 1/0 | 6/4 |
| **ATMS** | 3/0 | 44/0 | 3/0 | 94/0 | 144/0 |
| **Total** | 18/27 | 285/15 | 5/1 | 119/0 | 424/43 |

Table 2: Confusion Matrix for the test sets (FMEA / artificial). The rows represent the actual number of instances within the category, while the columns show the predicted outcome.

| | FMEA | AI |
|---|---|---|
| \|Training Set\| | 1696 | 211 |
| \|Test Set \| | 424 | 43 |
| Total Test Time | 5ms | 2 ms |
| Correctly Classified Instances | 330 (77.83 %) | 35 ( 81.39 %) |
| Incorrectly Classified Instances | 94 (22.17 %) | 8 ( 18.6 %) |
| Kappa statistic | 0.5797 | 0.6627 |
| Mean absolute error | 0.1335 | 0.1379 |
| Root mean squared error | 0.2905 | 0.3472 |
| Relative absolute error | 46.7288% | 36.3459 % |
| Root relative squared error | 77.3024 % | 78.0153 % |

Table 3: Classification Statistics

clauses ($12 \leq |Th| \leq 105$). For each experiment we randomly chose the number of observations as well as the manifestations themselves from the effects available within the model. A total of 2120 experiments where conducted on the FMEA sample set.

In case of the artificial portion of our classification data, we produced examples with a varying number of hypotheses ($10 \leq |Hyp| \leq 500$), effects ($4 \leq |M| \leq 13001$) and clauses ($100 \leq |Th| \leq 13500$). Furthermore, we chose the effect overlap randomly as well as the outdegree of the disorders. Due to the implementation of the artificial example generator, we do not observe several independent diagnosis subproblems within these models. We collected the data on 252 experiments with $|Obs|$ ranging from 1 to 25. Clearly, the majority of these models is larger in size than the FMEA examples and thus computationally more expensive.

With each experiment run we collected the following 27 metrics in accordance to the previous section:

1. Logic model specific
   - Number of hypotheses
   - Number of effects
   - Number of causal relations, i.e. clauses in the theory
   - Number of independent diagnosis subproblems
2. DAG
   - Outdegree of hypothesis nodes (maximum, average, standard deviation)
   - Indegree of effect nodes (maximum, average, standard deviation)
   - Covering (maximum, average, standard deviation)
   - Overlap (maximum, average, standard deviation)
3. Hypergraph
   - Path length (maximum, average, standard deviation)[3]
4. Instance specific/Observation dependent
   - Number of observations
   - Indegree current observation nodes (maximum, average, standard deviation)
   - Overlap current observation (maximum, average, standard deviation)
   - Number of independent diagnosis subproblems including current observations

All these metrics build our feature vector for the classification. The variable to be predicted is the algorithm, i.e. ATMS, BHSTREE, HSDAG, or HST which would be the most efficient on the current diagnosis problem.

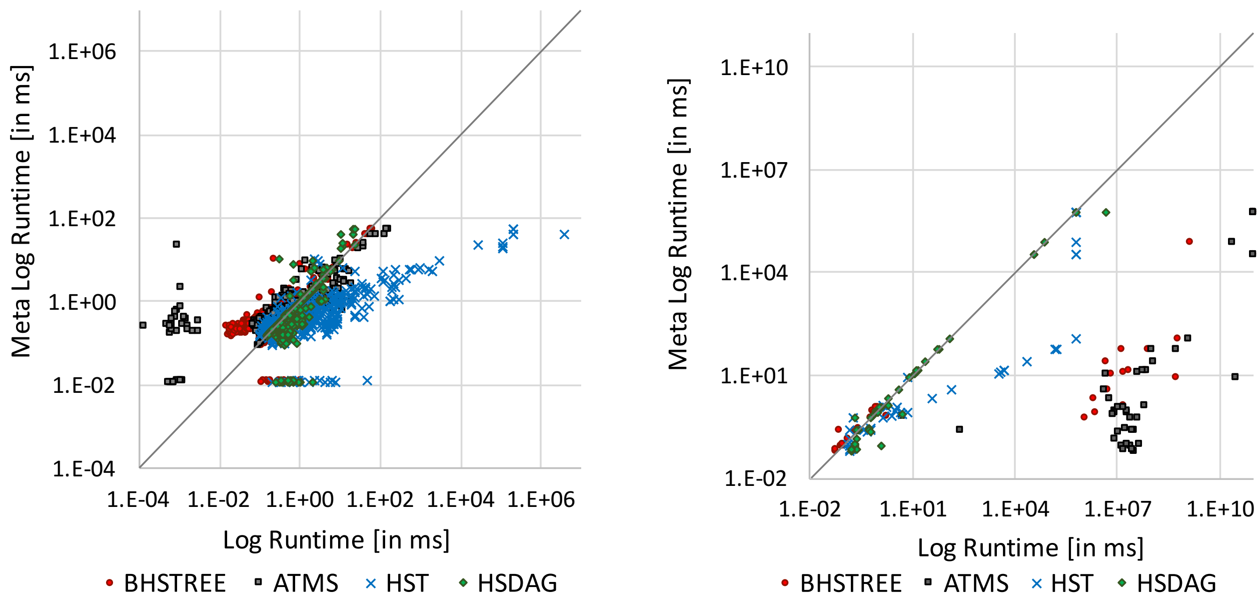
Each experiment data series was split into a training set comprising 80% of the data and a test set of 20%. Dividing the data, we made sure to split it in such a way that the test set comprises models of various sizes. The instances were created with increasing number of hypotheses. Before selecting the classification method, we performed cross validation on several classification algorithms available in WEKA on the training data. Based on the accuracy obtained we decided to use a multilayer perceptron as the classifier for both the FMEA-based models and the artificial examples.

## 4.2 Evaluation Results

As can be seen in Table 3 the classification based on the metrics reaches a satisfactory success rate on the FMEA-based as well as artificial examples. The confusion matrix in Table 2 shows the number of correctly and wrongly classified instances. Each row depicts the number of instances where each algorithm was in fact the most efficient approach, while the columns represent the machine classifier's forecast, i.e. how many times each abduction method was predicted to have the smallest runtime. From the table we can observe for the FMEA-based samples that the BHSTREE approach computes the most instances the fastest, followed by the ATMS. This contingency table further shows that the neural network overestimates BHSTREE, thus predicts it more often than it should. For the remaining methods we can observe the contrary. Yet, whenever the classifier categorizes the problem incorrectly, the suggested algorithm is the second or third most efficient. In regard to the artificial examples Table 2 shows that HSDAG and BHSTREE outperform the other approaches in the portfolio, i.e. ATMS and HST. In particular, the ATMS could not compute diagnoses efficiently enough on the artificial test set. Similarly to the FMEA-based results, we can observe that HSDAG in this case is favored by the predictor as it is selected on more instances than it is supposed to be selected.

To discover whether our meta-approach provides an efficiency improvement, we compared computation time on

[3]Path length between hypothesis vertices.




(a) FMEA-based samples

(b) Artificial samples

Figure 2: Runtime performance comparison between the meta-approach and the direct abduction algorithms on the test sets.

both test sets for all methods, i.e. our meta-algorithm and each abductive reasoning technique. The runtime for the meta-algorithm is determined by first, the computation of the instance based metrics, second the time it takes to create the feature vector, supply it to the classifier and predicting the best algorithm and third the diagnosis time of the suggested procedure. Figure 2 depicts the performance of the meta-approach in comparison to the algorithms from the portfolio based on the test sets. As can be seen from Figure 2a our meta-algorithm is not able to outperform on average (M=1.55 ms, SD=5.74) all direct diagnosis methods, i.e. BHSTREE (M=1.39 ms, SD=5.67), HSDAG (M=1.12 ms, SD=2.57), ATMS (M=2.71 ms, SD=12.98), HST (M=11256.21 ms, SD=178418.09) for the FMEA-based examples. The reason being that the mean time to collect the metrics of the PHCAP is close to the actual diagnosis time of these problems. From Figure 2a we can see that in some instances, the ATMS performs rather well in comparison to the meta-approach, however, this is compensated by several inefficient runtime records on various instances.

In case of the artificial examples our approach performs well, i.e. on average the meta-algorithm is the most efficient technique. Due to the size of these samples, the computation of the properties only demands a fraction of the actual diagnosis run time. On average our algorithm requires 72227.36 ms and is thus 99.9% faster than the ATMS and BHSTREE, 56.24% faster than the HSDAG and 45.51% faster than HST. A paired t-test ($\alpha = 0.05, t = 2.02$) determined that the meta-approach is significantly more efficient than the ATMS ($p = 0.003$), BHSTREE ($p = 0.006$), and HST ($p = 0.021$) for the artificial samples. As can be seen in Figure 2b in the lower left quadrant where smaller runtimes are depicted, there are instances which can be computed faster with BHSTREE, HSDAG, and HST. In the remaining scatter plot we can see that for computationally more expensive instances the HSDAG and the meta-approach perform equally, due to the meta-approach selecting the HSDAG as the abduction strategy most often.

## 5 Conclusion and Future Work

Computation time of abductive diagnosis depends primarily on the underlying model. Thus, we investigated algorithm selection, as a form to predict the best performing method based on the structural properties of problem instances. We explore metrics inherent to the structure of FMEA-based models, which form the feature vector for a classifier as part of a meta-approach. Evaluated on two test sets, the metrics led to a satisfactory selection of the best algorithm for a particular diagnosis problem. In case of the FMEA-based samples our meta-algorithm was restrained by the time necessary to construct the feature vector and thus could not outperform all abduction methods. Our approach shows its value when operating on larger problem instances, where it performs well and in fact is on average the most efficient in comparison. What we can derive from the evaluation is that even though the sample sets are similar in their topological properties, the best performing algorithms differentiate between the example sources. Thus, we believe that further investigation within the direction of algorithm selection may allow a very efficient general meta-approach.

Despite the satisfactory classification results, there are certain metrics worth investigating such as treewidth based on the decomposition of the hypergraph as well as different attribute combinations to determine the metrics most suited. Regarding the restricted representation class, we plan on expanding our approach to more expressive models.

## Acknowledgments

The work presented in this paper has been supported by the FFG project Applied Model Based Reasoning (AMOR) under grant 842407. We would further like to express our gratitude to our industrial partner, Uptime Engineering GmbH.

## References


[Baumeister and Seipel, 2002] Joachim Baumeister and Dietmar Seipel. Diagnostic reasoning with multilevel set-covering models. Technical report, DTIC Document, 2002.

[Catelani *et al.*, 2010] Marcantonio Catelani, Lorenzo Ciani, and Valentina Luongo. The FMEDA approach to improve the safety assessment according to the IEC61508. *Microelectronics Reliability*, 50:1230–1235, 2010.

[Console *et al.*, 1991] Luca Console, Daniele Theseider Dupre, and Pietro Torasso. On the Relationship Between Abduction and Deduction. *Journal of Logic and Computation*, 1(5):661–690, 1991.

[de Kleer and Williams, 1987] Johan de Kleer and Brian C Williams. Diagnosing Multiple Faults. *Artificial Intelligence*, 32(1):97–130, 1987.

[de Kleer, 1986a] Johan de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.

[de Kleer, 1986b] Johan de Kleer. Problem solving with the ATMS. *Artificial Intelligence*, 28(2):197–224, 1986.

[Eiter and Gottlob, 1995] Thomas Eiter and Georg Gottlob. The complexity of logic-based abduction. *Journal of the ACM (JACM)*, 42(1):3–42, 1995.

[Friedrich *et al.*, 1990] Gerhard Friedrich, Georg Gottlob, and Wolfgang Nejdl. Hypothesis classification, abductive diagnosis and therapy. In *Expert Systems in Engineering Principles and Applications*, pages 69–78. Springer, 1990.

[Guan and Jiang, 2013] Qixue Guan and Yueqiu Jiang. The parsimonious covering theory applied for communication equipment malfunction. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th International Conference on*, pages 637–641. IEEE, 2013.

[Hall *et al.*, 2009] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[Hawkins and Woollons, 1998] Peter G. Hawkins and Davis J. Woollons. Failure modes and effects analysis of complex engineering systems using functional models. *Artificial Intelligence in Engineering*, 12:375–397, 1998.

[Hutter *et al.*, 2006] Frank Hutter, Youssef Hamadi, Holger H Hoos, and Kevin Leyton-Brown. Performance prediction and automated tuning of randomized and parametric algorithms. In *Principles and Practice of Constraint Programming-CP 2006*, pages 213–228. Springer, 2006.

[Kakas *et al.*, 1992] Antonis C. Kakas, Robert A. Kowalski, and Francesca Toni. Abductive logic programming. *Journal of logic and computation*, 2(6):719–770, 1992.

[Karp, 1972] Richard M Karp. *Reducibility among combinatorial problems*. Springer, 1972.

[Kotthoff, 2012] Lars Kotthoff. Algorithm selection for combinatorial search problems: A survey. *arXiv preprint arXiv:1210.7959*, 2012.

[Lambrix *et al.*, 2013] Patrick Lambrix, Fang Wei-Kleiner, Zlatan Dragisic, and Valentina Ivanova. Repairing missing is-a structure in ontologies is an abductive reasoning problem. In *WoDOOM*, pages 33–44, 2013.

[Leyton-Brown *et al.*, 2003] Kevin Leyton-Brown, Eugene Nudelman, Galen Andrew, Jim McFadden, and Yoav Shoham. A portfolio approach to algorithm select. In *Proceedings of the 18th international joint conference on Artificial intelligence*, pages 1542–1543. Morgan Kaufmann Publishers Inc., 2003.

[Lin and Jiang, 2003] Li Lin and Yunfei Jiang. The computation of hitting sets: Review and new algorithms. *Information Processing Letters*, 86(4):177–184, 2003.

[Marquis, 2000] Pierre Marquis. Consequence finding algorithms. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 41–145. Springer, 2000.

[McIlraith, 1998] Sheila A McIlraith. Logic-based abductive inference. *Knowledge Systems Laboratory, Technical Report KSL-98-19*, 1998.

[Morak *et al.*, 2012] Michael Morak, Nysret Musliu, Reinhard Pichler, Stefan Rümmele, and Stefan Woltran. Evaluating tree-decomposition based algorithms for answer set programming. In *Learning and Intelligent Optimization*, pages 130–144. Springer, 2012.

[Musliu and Schwengerer, 2013] Nysret Musliu and Martin Schwengerer. Algorithm selection for the graph coloring problem. In *Learning and Intelligent Optimization*, pages 389–403. Springer, 2013.

[Nordh and Zanuttini, 2008] Gustav Nordh and Bruno Zanuttini. What makes propositional abduction tractable. *Artificial Intelligence*, 172(10):1245–1284, 2008.

[Patil *et al.*, 1982] Ramesh S Patil, Peter Szolovits, and William B Schwartz. Modeling knowledge of the patient in acid-base and electrolyte disorders. *Artificial Intelligence in Medicine*, 345348, 1982.

[Peng and Reggia, 1986] Yun Peng and James A Reggia. Plausibility of diagnostic hypotheses: The nature of simplicity. In *AAAI*, volume 86, pages 140–145, 1986.

[Peng and Reggia, 1990] Yun Peng and James A Reggia. *Abductive inference models for diagnostic problem-solving*. Springer, 1990.

[Poole and Kanazawa, 1994] David Poole and Keiji Kanazawa. A decision-theoretic abductive basis for planning. In *AAAI Spr. Symp. on Decision-Theoretic Planning*, 1994.

[Quaritsch and Pill, 2014] Thomas Quaritsch and Ingo Pill. Pymbd: A library of mbd algorithms and a light-weight evaluation platform. *Proceedings of Dx-2014*, 2014.

[Reiter, 1987] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.

[Rice, 1976] John R Rice. The algorithm selection problem. *Advances in computers*, 15:65–118, 1976.

[Smith-Miles, 2009] Kate A Smith-Miles. Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys (CSUR)*, 41(1):6, 2009.

[Sterling *et al.*, 2014] Raymond Sterling, Peter Struss, Jesús Febres, Umbreen Sabir, and Marcus M Keane. From modelica models to fault diagnosis in air handling units. In *10th International Modelica Conference*, 2014.

[Wotawa, 2001] Franz Wotawa. A variant of reiter's hitting-set algorithm. *Information Processing Letters*, 79(1):45–51, 2001.

[Wotawa, 2014] Franz Wotawa. Failure mode and effect analysis for abductive diagnosis. In *Proceedings of the International Workshop on Defeasible and Ampliative Reasoning (DARe-14)*, volume 1212. CEUR Workshop Proceedings, ISSN 1613-0073, 2014. http://ceur-ws.org/Vol-1212/.

[Xu *et al.*, 2008] Lin Xu, Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown. Satzilla: portfolio-based algorithm selection for sat. *Journal of Artificial Intelligence Research*, pages 565–606, 2008.