An efficient diagnosis algorithm for inconsistent constraint sets

A. FELFERNIG, M. SCHUBERT, AND C. ZEHENTNER
Institute for Software Technology, Graz University of Technology, Graz, Austria
(RECEIVED May 26, 2010; ACCEPTED October 4, 2010)

Abstract
Constraint sets can become inconsistent in different contexts. For example, during a configuration session the set of customer requirements can become inconsistent with the configuration knowledge base. Another example is the engineering phase of a configuration knowledge base where the underlying constraints can become inconsistent with a set of test cases. In such situations we are in the need of techniques that support the identification of minimal sets of faulty constraints that have to be deleted in order to restore consistency. In this paper we introduce a divide and conquer-based diagnosis algorithm (FASTDIAG) that identifies minimal sets of faulty constraints in an overconstrained problem. This algorithm is specifically applicable in scenarios where the efficient identification of leading (preferred) diagnoses is crucial. We compare the performance of FASTDIAG with the conflict-directed calculation of hitting sets and present an in-depth performance analysis that shows the advantages of our approach.

Keywords: Inconsistent Constraint Sets; Interactive Configuration; Model-Based Diagnosis; Preferred Diagnosis

1. INTRODUCTION
Constraint technologies (Tsang, 1993) are applied in different areas such as configuration (Mittal & Frayman, 1989; Fleischanderl et al., 1998; Sinz & Haag, 2007), recommendation (Felfernig et al., 2009), and scheduling (Castillo et al., 2005). There are many scenarios where the underlying constraint sets can become overconstrained. For example, when implementing a configuration knowledge base, constraints can become inconsistent with a set of test cases (Felfernig et al., 2004). Alternatively, when interacting with a configurator application (O’Sullivan et al., 2007; Felfernig et al., 2009), the given set of customer requirements (represented as constraints) can become inconsistent with the configuration knowledge base. In both situations there is a need of an intelligent assistance that actively supports users of a constraint-based application (end users or knowledge engineers). A widespread approach to support users in the identification of minimal sets of faulty constraints is to combine conflict detection (e.g., see Junker, 2004) with a corresponding hitting set algorithm (DeKleer & Williams, 1987; Reiter, 1987; DeKleer et al., 1992). In their original form these algorithms are applied for the calculation of minimal (cardinality) diagnoses that are typically determined with breadth-first search. Further diagnosis algorithms have been developed that follow a best-first search regime where the expansion of the hitting set search tree is guided by failure probabilities of components (DeKleer, 1990). Another example for such an approach is presented in (Felfernig et al., 2009), where similarity metrics are used to guide the (best-first) search for a preferred (plausible) minimal diagnosis (including repairs).

Both simple breadth-first search and best-first search diagnosis approaches are predominantly relying on the calculation of conflict sets (Junker, 2004). In this context, the determination of a minimal diagnosis of cardinality \( n \) requires the identification of at least \( n \) minimal conflict sets. In this paper we introduce a diagnosis algorithm (FASTDIAG) that allows to determine one minimal diagnosis at a time with the same computational effort related to the calculation of one conflict set at a time. The algorithm supports the identification of preferred diagnoses given predefined preferences regarding a set of decision alternatives. FASTDIAG is boosting the applicability of diagnosis methods in scenarios such as online configuration and reconfiguration (Felfernig et al., 2004), recommendation of products and services (Felfernig et al., 2009), and (more generally) in scenarios where the efficient calculation of preferred (leading) diagnoses is crucial (DeKleer, 1990). FASTDIAG is not restricted to constraint-based systems but it
is also applicable, for example, in the context of SAT solving (Marques-Silva & Sakallah, 1996) and description logics reasoning (Friedrich & Shchekotykhin, 2005).

The remainder of this paper is organized as follows. In Section 2 we introduce a simple example configuration task from the automotive domain. In Section 3 we discuss the basic hitting set-based approach to the calculation of diagnoses. In Section 4 we introduce an algorithm (FASTDIAG) for calculating preferred diagnoses for a given overconstrained problem. In Section 5 we present a detailed evaluation of FASTDIAG, which clearly outperforms standard hitting set-based algorithms in the calculation of the topmost-n preferred diagnoses. With Section 6 we provide an overview of related work in the field. The paper is concluded with Section 7.

2. EXAMPLE DOMAIN: CAR CONFIGURATION

Car configuration will serve as a working example throughout this paper. Because we exploit configuration problems for the discussion of our diagnosis algorithm, we first introduce a formal definition of a configuration task. This definition is based on Felfernig et al. (2004) but is given in the context of a constraint satisfaction problem (CSP; Tsang, 1993).

Definition 1 (configuration task). A configuration task can be defined as a CSP \((V, D, C)\). Here, \(V = \{v_1, v_2, \ldots, v_6\}\) represents a set of finite domain variables; \(D = \{\text{dom}(v_1), \text{dom}(v_2), \ldots, \text{dom}(v_6)\}\) represents a set of variable domains \(\text{dom}(v_i)\), where \(\text{dom}(v_i)\) represents the domain of variable \(v_i\); \(C = C_{\text{KB}} \cup C_{\text{R}}\), where \(C_{\text{KB}} = \{c_1, c_2, \ldots, c_g\}\) is a set of domain specific constraints (the configuration knowledge base) that restrict the possible combinations of values assigned to the variables in \(V\); and \(C_{\text{R}} = \{c_{g+1}, c_{g+2}, \ldots, c_l\}\) is a set of customer requirements that is also represented as constraints.

A simplified example of a configuration task in the automotive domain is the following. In this example, \(type\) represents the car type, \(pdc\) is the park distance control functionality, \(fuel\) represents the fuel consumption per 100 km, a \(skibag\) allows ski stowage inside the car, and \(4\text{-wheel}\) represents the corresponding actuation type. These variables describe the potential set of requirements that can be specified by the user (customer). The possible combinations of these requirements are defined by a set of constraints that are denoted as \(C_{\text{KB}}\), which is defined as \(C_{\text{KB}} = \{c_1, c_2, c_3, c_4\}\) in our example. Furthermore, we assume the set of customer requirements \(C_{\text{R}} = \{c_5, c_6, c_7\}\).

- \(V = \{\text{type}, \text{pdc}, \text{fuel}, \text{skibag}, 4\text{-wheel}\}\)
- \(D = \{\text{dom}(\text{type}) = \{\text{city}, \text{limo}, \text{combi}, \text{xdrive}\}, \text{dom}(\text{pdc}) = \{\text{yes}, \text{no}\}, \text{dom}(\text{fuel}) = \{4l, 6l, 10l\}, \text{dom}(\text{skibag}) = \{\text{yes}, \text{no}\}, \text{dom}(4\text{-wheel}) = \{\text{yes}, \text{no}\}\}\)
- \(C_{\text{KB}} = \{c_1: 4\text{-wheel} = \text{yes} \Rightarrow \text{type} = \text{xdrive}; c_2: \text{skibag} = \text{yes} \Rightarrow \text{type} = \text{city}; c_3: \text{fuel} = 4l \Rightarrow \text{type} = \text{city}; c_4: \text{fuel} = 6l \Rightarrow \text{type} \neq \text{xdrive}\}\)
- \(C_{\text{R}} = \{c_5: \text{type} = \text{combi}, c_6: \text{fuel} = 4l, c_7: 4\text{-wheel} = \text{yes}\}\)

On the basis of this configuration task definition, we can now introduce the definition of a concrete configuration (solution for a configuration task).

Definition 2 (configuration). A configuration for a given configuration task \((V, D, C)\) is an instantiation \(I = \{v_1 = \text{ins}_1, v_2 = \text{ins}_2, \ldots, v_n = \text{ins}_n\}\), where \(\text{ins}_k \in \text{dom}(v_k)\).

A configuration is consistent if the assignments in \(I\) are consistent with the \(c_i \in C\). Furthermore, a configuration is complete if all variables in \(V\) are instantiated. Finally, a configuration is valid if it is consistent and complete.

3. DIAGNOSING OVERCONSTRAINED PROBLEMS

For the configuration task introduced in Section 2 we are not able to find a solution, for example, a \(\text{combi}\)-type car does not support a fuel consumption of 4l per 100 km. Consequently, we want to identify minimal sets of constraints \(c_i \in C_{\text{R}}\) which have to be deleted in order to be able to identify a solution (restore the consistency). In the example of Section 2 the set of constraints \(C_{\text{R}} = \{c_5, c_6, c_7\}\) is inconsistent with the constraints \(C_{\text{KB}} = \{c_1, c_2, c_3, c_4\}\), that is, no solution can be found for the underlying configuration task. A standard approach to determine a minimal set of constraints that have to be deleted from an overconstrained problem is to resolve all minimal conflicts contained in the constraint set. The determination of such constraints is based on a conflict detection algorithm (e.g., see Junker, 2004), the derivation of the corresponding diagnoses is based on the calculation of hitting sets (Reiter, 1987). Because both the notion of a (minimal) conflict and the notion of a (minimal) diagnosis will be used in the following sections, we provide the corresponding definitions here.

Definition 3 (conflict set). A conflict set is a set \(C_{\text{S}} \subseteq C_{\text{R}}\) such that \(C_{\text{KB}} \cup C_{\text{S}}\) is inconsistent. \(C_{\text{S}}\) is a minimal if there does not exist a conflict set \(C_{\text{S'}}\) with \(C_{\text{S'}} \subset C_{\text{S}}\).

In our working example we can identify three minimal conflict sets, which are \(C_{\text{S}} = \{c_5, c_6\}\), \(C_{\text{S}} = \{c_5, c_7\}\), and \(C_{\text{S}} = \{c_6, c_7\}\).

\(C_{\text{S}}\), \(C_{\text{S}}\), and \(C_{\text{S}}\) are conflict sets because \(C_{\text{S}} \subseteq C_{\text{KB}} \lor C_{\text{S}} \subseteq C_{\text{KB}} \lor C_{\text{S}} \subseteq C_{\text{KB}}\) is inconsistent. The minimality property is fulfilled because a conflict set \(C_{\text{S}}\) does not exist with \(C_{\text{S}} \subset C_{\text{S}}\) or \(C_{\text{S}} \subset C_{\text{S}}\) or \(C_{\text{S}} \subset C_{\text{S}}\). The standard approach to resolve the given conflicts is the construction of a corresponding hitting set directed acyclic graph (HSDAG; Reiter, 1987) where the resolution of all minimal conflict sets automatically corresponds to the identification of a minimal diagnosis. A minimal diagnosis in our application context is a minimal set of customer requirements contained in the set of car features \(C_{\text{R}}\) that has to be deleted from \(C_{\text{R}}\).
in order to make the remaining constraints consistent with \( C_{KB} \). Because we are dealing with the diagnosis of customer requirements, we introduce the definition of a customer requirements diagnosis problem (Definition 4). This definition is based on the definition given in Felfernig et al. (2004).

**Definition 4 (CR diagnosis problem).** A customer requirements diagnosis (CR diagnosis) problem is defined as a tuple \((C_{KB}, C_R)\), where \( C_R \) is the set of given customer requirements and \( C_{KB} \) represents the constraints part of the configuration knowledge base.

The definition of a CR diagnosis that corresponds to a given CR diagnosis problem is the following (see Definition 5).

**Definition 5 (CR diagnosis).** A CR diagnosis for a CR diagnosis problem \((C_{KB}, C_R)\) is a set \( \Delta \subseteq C_R \), such that \( C_{KB} \cup (C_R - \Delta) \) is consistent. \( \Delta \) is minimal if there does not exist a diagnosis \( \Delta' \subseteq \Delta \), such that \( C_{KB} \cup (C_R - \Delta') \) is consistent.

The HSDAG algorithm for determining minimal diagnoses is discussed in detail in Reiter (1987). The concept of this algorithm will be explained on the basis of our working example. It relies on a conflict detection algorithm that is responsible for detecting minimal conflicts in a given set of constraints (in our case, in the given customer requirements). One conflict detection algorithm is QuickXplain (Junker, 2004), which is based on an efficient divide and conquer search strategy. For the purposes of our working example let us assume that the first minimal conflict set determined by QuickXplain is the set \( CS_1 = \{c_5, c_6\} \). Because of the minimality property, we are able to resolve each conflict by simply deleting one element from the set, for example, in the case of \( CS_1 \) we have to either delete \( c_5 \) or \( c_6 \). Each variant to resolve a conflict set is represented by a specific path in the corresponding HSDAG. The HSDAG for our working example is depicted in Figure 1.

The deletion of \( c_5 \) from \( CS_1 \) triggers the calculation of another conflict set \( CS_2 = \{c_6, c_7\} \) because \( C_R - \{c_5\} \cup C_{KB} \) is inconsistent. If we decide to delete \( c_6 \) from \( CS_1 \), \( C_R - \{c_6\} \cup C_{KB} \) remains inconsistent, which means that QuickXplain returns another minimal conflict set, which is \( CS_2 = \{c_5, c_7\} \).

The original HSDAG algorithm (Reiter, 1987) follows a strict breadth-first search regime. Following this strategy, the next node to be expanded in our working example is the minimal conflict set \( CS_3 \), which has been returned by QuickXplain for \( C_R \) – \( \{c_6\} \cup C_{KB} \). In this context, the first option to resolve \( CS_3 \) is to delete \( c_6 \). This option is a valid one and \( \Delta_1 = \{c_5, c_6\} \) is the resulting minimal diagnosis. The second option for resolving \( CS_3 \) is to delete the constraint \( c_7 \). In this case, we have identified the next minimal diagnosis \( \Delta_2 = \{c_5, c_7\} \) because \( C_R - \{c_5, c_7\} \cup C_{KB} \) is consistent. This way we are able to identify all minimal sets of constraints \( \Delta_i \) that, if deleted from \( C_R \), help to restore the consistency with \( C_{KB} \).

If we want to calculate the complete set of diagnoses for our working example, we still have to resolve the conflict set \( CS_2 \). The first option to resolve \( CS_2 \) is to delete \( c_5 \), because \( \{c_5, c_6\} \) has already been identified as a minimal diagnosis, we can close this node in the HSDAG. The second option to resolve \( CS_2 \) is to delete \( c_7 \). In this case we have determined the third minimal diagnosis, which is \( \Delta_3 = \{c_6, c_7\} \).

In our working example we are able to enumerate all possible diagnoses that help to restore consistency. However, the calculation of all minimal diagnoses is expensive and thus in many cases not practicable for interactive settings. Because users are often interested in a reduced subset of all the potential diagnoses, alternative algorithms are needed that are capable of identifying preferred diagnoses (Reiter, 1987; DeKleer, 1990; Felfernig et al., 2009). Such approaches have already been developed (DeKleer, 1990; Felfernig et al., 2009); however, they are still based on the resolution of conflict sets, which is computationally expensive (see Section 5). Our idea presented in this paper is a diagnosis algorithm that helps to determine preferred diagnoses without the need of calculating the corresponding conflict sets. The basic properties of FASTDiag will be discussed in Section 4.

**4. Calculating Preferred Diagnoses with FASTDiag**

**4.1. Preferred diagnoses**

Users typically prefer to keep the important requirements and to change or delete (if needed) the less important ones (Junker, 2004). The major goal of (model-based) diagnosis tasks is to identify the preferred (leading) diagnoses, which are not necessarily minimal cardinality ones (DeKleer, 1990). For the characterization of a preferred diagnosis we will rely on the definition of a total ordering of the given set of constraints in \( C \) (respectively \( C_R \)). Such a total ordering can be achieved, for example, by directly asking the customer regarding the preferences, by applying multiattribute utility theory (Winterfeldt & Edwards, 1986; Ardissono et al., 2003) where the determined interest dimensions correspond with the attributes of \( C_R \) or by applying the rankings determined by conjoint analysis (Belanger, 2005). The following definition of a lexicographical ordering (Definition 6) is based on total orderings for constraints that has been applied in (Junker, 2004) for the determination of preferred conflict sets.

**Definition 6 (total lexicographical ordering).** Given a total order \( < \) on \( C \), we enumerate the constraints in \( C \) in
increasing $<$ order $c_1, \ldots, c_n$, starting with the least important constraints (i.e., $c_i < c_j \Rightarrow i < j$). We compare two subsets $X$ and $Y$ of $C$ lexicographically:

$$X \succ_{\text{lex}} Y \iff \exists k: c_k \in Y - X \quad \text{and} \quad X \cap \{c_{k+1}, \ldots, c_n\} = Y \cap \{c_{k+1}, \ldots, c_n\}.$$ 

Based on this definition of a lexicographical ordering, we can now introduce the definition of a preferred diagnosis.

**Definition 7 (preferred diagnosis).** A minimal diagnosis $\Delta$ for a given CR diagnosis problem $(C_R, C_{KB})$ is a preferred diagnosis for $(C_R, C_{KB})$ if and only if another minimal diagnosis $\Delta'$ with $\Delta' \succ_{\text{lex}} \Delta$.

In our working example we assumed the lexicographical ordering $(c_5 < c_6 < c_7)$, that is, the most important customer requirement is $c_7$ (the 4-wheel functionality). If we assume that $X = \{c_5, c_7\}$ and $Y = \{c_6, c_7\}$ then $Y - X = \{c_6\}$ and $X \cap \{c_7\} = Y \cap \{c_7\}$. Intuitively, $\{c_5, c_7\}$ is a preferred diagnosis compared to $\{c_6, c_7\}$ because both diagnoses include $c_7$ but $c_5$ is less important than $c_6$. If we change the ordering to $(c_7 < c_5 < c_6)$, FASTDIAG would then determine $\{c_5, c_7\}$ as the preferred minimal diagnosis.

### 4.2. FASTDIAG approach

For the following discussions we introduce the set $AC$, which is initially set to $C_{KB} \cup C_R$ [the union of customer requirements ($C_{KB}$) and the configuration knowledge base ($C_{KB}$)] and subsequently changed when the algorithm runs. The basic idea of the FASTDIAG algorithm (Algorithm 1) is the following. In our example, the set of customer requirements $C_R = \{c_5, c_6, c_7\}$ includes at least one minimal diagnosis because $C_{KB}$ is consistent and $C_{KB} \cup C_R$ is inconsistent. In the extreme case $C_R$ itself represents the minimal diagnosis, which then means that all constraints in $C_R$ are part of the diagnosis, that is, each $c_i \in C_R$ represents a singleton conflict. In our case $C_R$ obviously does not represent a minimal diagnosis—the set of diagnoses in our working example is $\{\Delta_1 = \{c_5, c_6\}, \Delta_2 = \{c_5, c_7\}, \Delta_3 = \{c_6, c_7\}\}$ (see Section 3). The next step in Algorithm 1 is to divide the set of customer requirements $C_R = \{c_5, c_6, c_7\}$ into the two sets $C_1 = \{c_5\}$ and $C_2 = \{c_6, c_7\}$ and to check whether $AC - C_1$ is already consistent. If this is the case, we can omit the set $C_2$ because at least one minimal diagnosis can already be identified in $C_1$. In our case, $AC - C_1$ is inconsistent, which means that we have to consider further elements from $C_2$. Therefore, $C_2 = \{c_6, c_7\}$ is divided into the sets $\{c_6\}$ and $\{c_7\}$. In the next step we can check whether $AC - (C_1 \cup \{c_6\})$ is consistent—this is the case that means that we do not have to further take into account $\{c_7\}$ for determining the diagnosis. Because $\{c_5\}$ does not include a diagnosis but $\{c_5\} \cup \{c_6\}$ includes a diagnosis, we can deduce that $\{c_5\}$ must be part of the diagnosis. The final step is to check whether $AC - \{c_6\}$ leads to a diagnosis without including $\{c_5\}$. We see that $AC - \{c_6\}$ is inconsistent, that is, $\Delta = \{c_5, c_6\}$ is a minimal diagnosis for the CR diagnosis problem $(C_R = \{c_5, c_6, c_7\}, C_{KB} = \{c_1, c_2, c_3, c_4\})$. An execution trace of the FASTDIAG algorithm in the context of our working example is shown in Figure 2.

![Figure 2. FASTDIAG execution trace for the C6 diagnosis problem (C_R = {c5, c6, c7}, C_{KB} = {c1, c2, c3, c4}). Enumerations 1–6 show the order in which the different incarnations of the FD function are activated.](https://example.com/figure2)

**Algorithm 1: FASTDIAG**

```python
1 func FASTDIAG(C \subseteq AC, AC = \{c_1\ldots c_n\}) : diagnosis \Delta
2 if isEmpty(C) or inconsistent(AC - C) return \emptyset
3 else return FD(\emptyset, C, AC);
4 func FD(D, C = \{c_1\ldots c_q\}, AC : diagnosis \Delta
5 if D \not= \emptyset and consistent(AC) return \emptyset;
6 if singleton(AC) return C;
7 k = q/2;
8 C_1 = \{c_1\ldots c_k\}; C_2 = \{c_{k+1}\ldots c_q\};
9 D_1 = FD(C_1, C_2, AC - C_1);
10 D_2 = FD(D_1, C_1, AC - D_1);
11 return(D_1 \cup D_2)
```

### 4.3. Calculating $n > 1$ diagnoses

In order to be able to calculate $n > 1$ diagnoses\(^3\) with FASTDIAG we have to adopt the HSDAG construction introduced in Reiter (1987) by substituting the resolution of conflicts (see Fig. 1) with the deletion of elements $c_i$ from $C_R$ (see Fig. 3). In this case, a path in the HSDAG is closed if no further diagnoses can be identified for this path or the elements of the current path are a superset of an already closed path (containment check). Conform to the HSDAG approach presented in Reiter (1987), we expand the search tree in a breadth-first manner. In our working example, we can delete $\{c_5\}$ (one element of the first diagnosis $\Delta_1 = \{c_5, c_6\}$) from the set $C_R$ of diagnosable elements and restart the algorithm.

\(^3\)Typically a CR diagnosis problem has more than one related diagnosis.
for finding another minimal diagnosis for the CR diagnosis problem ($\{c_6, c_7\}$, $C_{KB}$). Because $AC - \{c_5\}$ is inconsistent, we can conclude that $C_R = \{c_6, c_7\}$ includes another minimal diagnosis ($\Delta_2 = \{c_6, c_7\}$), which is determined by FASTDIAG for the CR diagnosis problem ($C_R - \{c_5\}$, $C_{KB}$). Finally, we have to check whether the CR diagnosis problem ($\{c_5, c_7\}$, $C_{KB}$) leads to another minimal diagnosis. This is the case, that is, we have identified the last minimal diagnosis that is $\Delta_3 = \{c_5, c_7\}$. The calculation of all diagnoses in our working example on the basis of FASTDIAG is depicted in Figure 3.

Note that for a given set of constraints (C) FASTDIAG always calculates the preferred diagnosis in terms of Definition 7. If $\Delta_1$ is the diagnosis returned by FASTDIAG and we delete one element from $\Delta_1$ (e.g., $c_8$), then FASTDIAG returns the preferred diagnosis for the CR diagnosis problem ($\{c_5, c_6, c_7\} - \{c_8\}$, $C_{KB}$), which is $\Delta_2$ in our example case, that is, $\Delta_1 >_{lex} \Delta_2$. Consequently, diagnoses part of one path in the search tree (such as $\Delta_1$ and $\Delta_2$ in Fig. 3) are in a strict preference ordering. However, there is only a partial order between individual diagnoses in the search tree in the sense that a diagnosis at level $k$ is not necessarily preferable to a diagnosis at level $k + 1$.

4.4. FASTDIAG properties

A detailed listing of the basic operations of FASTDIAG is shown in Algorithm 1. First, the algorithm checks whether the constraints in C contain a diagnosis, that is, whether $AC - C$ is consistent—the function assumes that it is activated in the case that $AC$ is inconsistent. If $AC - C$ is inconsistent or $C = \emptyset$, FASTDIAG returns the empty set as result (no solution can be found, line 2 of the algorithm). If at least one diagnosis is contained in the set of constraints C, FASTDIAG activates the FASTDIAG function that is in charge of retrieving a preferred diagnosis (line 3 of the algorithm). FASTDIAG follows a divide and conquer strategy where the recursive function FASTDIAG divides the set of constraints (in our case the elements of $C_R$) into two different subsets ($C_1$ and $C_2$; line 8 of the algorithm) and tries to figure out whether $C_1$ already contains a diagnosis (line 5 of the algorithm). If this is the case, FASTDIAG does not further take into account the constraints in $C_2$. If only one element is remaining in the current set of constraints $C$ and the current set of constraints in $AC$ is still inconsistent, then the element in $C$ is part of a minimal diagnosis (line 6 of the algorithm). FASTDIAG is complete in the sense that if $C$ contains exactly one minimal diagnosis then FD will find it. If there are multiple minimal diagnoses then one of them (the preferred one, see Definition 7) is returned. The recursive function FD is triggered if $AC - C$ is consistent and $C$ consists of at least one constraint. In such a situation a corresponding minimal diagnosis can be identified. If we assume the existence of a minimal diagnosis $\Delta$ that cannot be identified by FASTDIAG, this would mean that there exists at least one constraint $c_a$ in $C$ that is part of the diagnosis but not returned by FD. The only way in which elements can be deleted from $C$ (i.e., not included in a diagnosis) is by the return $\emptyset$ statement in FD and $\emptyset$ is only returned in the case that $AC$ is consistent, which means that the elements of $C_2$ ($C_1$) from the previous FD incarnation are not part of the preferred diagnosis. Consequently, it is not possible to delete elements from $C$, which are part of the diagnosis. FASTDIAG computes only minimal diagnoses in the sense of Definition 5. If we assume the existence of a nonminimal diagnosis $\Delta$ calculated by FASTDIAG, this would mean that there exists at least one constraint $c_a$ with $\Delta - \{c_a\}$ is still a diagnosis. The only situation in which elements of $C$ are added to a diagnosis $\Delta$ is if $C$ itself contains exactly one element. If $C$ contains only one element (let us assume $c_a$) and $AC$ is inconsistent (in the function FASTDIAG) then $c_a$ is the only element that can be deleted from $AC$, that is, $c_a$ must be part of the diagnosis.

5. EVALUATION

5.1. Performance of FASTDIAG

In this section we will compare the performance of FASTDIAG with the performance of the hitting set algorithm (Reiter, 1987) in combination with the QUICKXPLAIN conflict detection algorithm introduced in (Junker, 2004).

The worst case complexity of FASTDIAG in terms of the number of consistency checks needed for calculating one minimal diagnosis is $2d \times \log_2(n/d) + 2d$, where $d$ is the minimal diagnoses set size and $n$ is the number of constraints (in C). The best case complexity is $\log_2(n/d) + 2d$. In the worst case each element of the diagnosis is contained in a different path of the search tree: $\log_2(n/d)$ is the depth of the path, $2d$ represents the branching factor and the number of leaf-node consistency checks. In the best case all ele-
ments of the diagnosis are contained in one path of the search tree.

The worst case complexity of QuickXplain in terms of consistency checks needed for calculating one minimal conflict set is $2k \cdot \log_2(n/k) + 2k$ where $k$ is the minimal conflicts set size and $n$ is again the number of constraints (in C; Junker, 2004). The best case complexity of QuickXplain in terms of the number of consistency checks needed is $\log_2(n/k) + 2k$ (Junker, 2004). Consequently, the number of consistency checks per conflict set (QuickXplain) and the number of consistency checks per diagnosis (FastDiag) fall into a logarithmic complexity class.

Let $n_{cs}$ be the number of minimal conflict sets in a constraint set and $n_{diag}$ be the number of minimal diagnoses, then we need $n_{diag}$ FastDiag calls (see Algorithm 1) plus $n_{cs}$ additional consistency checks and $n_{cs}$ activations of QuickXplain with $n_{diag}$ additional consistency checks for determining all diagnoses.

The results of a performance evaluation of FastDiag are depicted in the Figure 4, Figure 5, Figure 6, and Figure 7. The basis for these evaluations was the bicycle configuration knowledge base taken from the CLib4 (www.itu.dk/research/cla/externals/clib/) configuration benchmarks library (34 variables and about 65 constraints). For this example knowledge base we randomly generated different sets of requirements (of cardinality 5, 7, 10, and 15 requirements) and measured the performance of calculating corresponding diagnosis sets (the first diagnosis, first 5 diagnoses, first 10 diagnoses, and all diagnoses). The run time performance of the different diagnosis algorithms and the needed amount of TP calls is shown in the Figures 4–7. As solver we used the CLib-based decision diagram representation that allows for backtracking-free solution search. The tests have been executed on a standard desktop computer (Intel Core 2 Quad CPU QD9400 2.66-GHz CPU with 2 GB of RAM). Note that we have evaluated the performance of FastDiag with different other benchmark configuration knowledge bases on the CLib Web page with basically the same result. FastDiag shows to be a valuable alternative for determining diagnoses in interactive settings especially for calculating the preferred first-$n$ solutions.

Figure 4 shows a comparison between the hitting set based diagnosis approach (denoted as HSDAG) and the FastDiag algorithm (denoted as FastDiag) in the case that only one diagnosis is calculated. FastDiag clearly outperforms the HSDAG approach independent of the way in which diagnoses are calculated (breadth-first or best-first). Figure 5 shows the performance evaluation for calculating the topmost-5 minimal diagnoses. The result is similar to the one for calculating the first diagnosis, that is, FastDiag outperforms the two HSDAG versions. Our evaluations show that FastDiag is very efficient in calculating preferred minimal diagnoses.
5.2. Empirical evaluation

Based on a computer configuration dataset of the Graz University of Technology \((N = 415\) configurations) we evaluated the three presented approaches with respect to their capability of predicting diagnoses that are acceptable for the user (diagnoses leading to selected configurations). Each entry of the dataset consists of a set of initial user requirements \(C_R\) inconsistent with the configuration knowledge base \(C_{KB}\) and the configuration that had been finally selected by the user. Because the original requirements stored in the dataset are inconsistent with the configuration knowledge base, we could determine those diagnoses that indicated which minimal sets of requirements have to be deleted in order to be able to find a solution.

We evaluated the prediction accuracy of the three diagnosis approaches \((\text{breadth-first HSDAG, FASTDIAG, and best-first HSDAG})\). We first measured the distance between the predicted position of a diagnosis leading to a selected configuration and the expected position of the diagnosis (which is 1). This distance was measured in terms of the root mean square deviation \((\text{RMSD};\) see Formula 1\). The results of this first analysis are the following: an important result is that FASTDIAG has the lowest RMSD value \((0.95)\), best-first HSDAG has a similar prediction quality \((\text{RMSD} = 0.97)\), and breadth-first HSDAG has the worst prediction quality \((\text{RMSD} = 1.64)\).

\[ \text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{predicted position}_i - 1)^2}. \]  

(1)

RMSD is an often used quality estimate but it provides only a limited view on the precision of a \((\text{diagnosis})\) prediction. Therefore, we wanted to analyze the precision of the diagnosis selection strategies discussed in this paper—a measure for the precision of a diagnosis algorithm is depicted in Formula 2. The idea behind this measure is to describe how often a diagnosis that leads to a selected configuration (selected by the user) is among the topmost\(n\) ranked diagnoses. As shown in Table 1, FASTDIAG and best-first HSDAG have highest prediction accuracy in terms of precision, whereas the breadth-first HSDAG approach shows the worst precision.

\[ \text{precision} = \frac{\text{correctly predicted diagnoses}}{\text{predicted diagnoses}}. \]  

(2)

We applied a Mann–Whitney \(U\) test in order to statistically analyze differences between the three diagnosis approaches in terms of ranking behavior. We conducted a pairwise comparison between the diagnosis approaches on the basis of the mentioned Mann–Whitney \(U\) test. We could identify a signif-
6. RELATED WORK

6.1. Knowledge base analysis

The authors of Felfernig et al. (2004) introduce an algorithm for the automated debugging of configuration knowledge bases. The idea is to combine a conflict detection algorithm such as QUICKXPLAIN (Junker, 2004) with the hitting set algorithm in the automated debugging of configuration knowledge bases. The hitting set algorithm (Reiter, 1987) is used in model-based diagnosis (MBD; Reiter, 1987) for the calculation of minimal diagnoses. In this context, conflicts are induced by test cases (examples) that, for example, stem from previous configuration sessions, have been automatically generated, or have been explicitly defined by domain experts. Further applications of MBD in constraint set debugging are introduced in Felfernig et al. (2007) where diagnosis concepts are used to identify minimal sets of faulty transition conditions in state charts and in Felfernig et al. (2008) where MBD is applied for the identification of faulty utility constraint sets in the context of knowledge-based recommendation. In contrast to Felfernig et al. (2004, 2007, 2008), our work provides an algorithm that allows to directly determine diagnoses without the need to determine corresponding conflict sets. FASTDIAG can be applied in knowledge engineering scenarios for calculating preferred diagnoses for faulty knowledge bases given that we are able to determine reasonable ordering for the given set of constraints; this could be achieved, for example, by the application of corresponding complexity metrics (Chen & Suen, 2003).

6.2. Conflict detection

In contrast to the algorithm presented in this paper, calculating diagnoses for inconsistent requirements typically relies on the existence of (minimal) conflict sets. A well-known algorithm with a logarithmic number of consistency checks, depending on the number of constraints in the knowledge base and the cardinality of the minimal conflicts, is QUICKXPLAIN (Junker, 2004). It has made a major contribution to more efficient interactive constraint-based applications. QUICKXPLAIN is based on a divide and conquer strategy. FASTDIAG relies on the same principle of divide and conquer but with a different focus, namely, the determination of minimal diagnoses. QUICKXPLAIN calculates minimal conflict sets based on the assumption of a linear preference ordering among the constraints. Similarly, if we assume a linear preference ordering of the constraints in C, FASTDIAG calculates preferred diagnoses.

6.3. Interactive settings

Note that in the interactive configuration scenario discussed in this paper our goal was to support open configuration that lets the user explore the configuration space where the system proactively points out inconsistent requirements, such a functionality is often provided by commercial configuration environments. O’Sullivan et al. (2007) focus on interactive settings where users of constraint-based applications are confronted with situations where no solution can be found. In this context, O’Sullivan et al. (2007) introduce the concept of minimal exclusion sets that correspond to the concept of minimal diagnoses as defined in Reiter (1987). As mentioned, the major focus of O’Sullivan et al. (2007) are settings where the proposed algorithm supports users in the identification of acceptable exclusion sets. The authors propose an algorithm (representative explanations) that helps to improve the quality of the presented exclusion set (in terms of diversity) and thus increases the probability of finding an acceptable exclusion set for the user. Our diagnosis approach calculates preferred diagnoses in terms of a predefined ordering of the constraint set. Thus, compared to the work of O’Sullivan et al. (2007), we follow a different approach in terms of focusing more on preferences than on the degree of representativeness.

6.4. Diagnosis algorithms

There are a couple of algorithms that help to improve the efficiency of diagnosis determination; they are further developments of the original algorithm introduced by Reiter (1987). These approaches focus on making the construction of hitting sets more efficient. Wotawa (2001) introduces an algorithm that reduces the number of subset checks compared to the original HSDAG approach (Reiter, 1987). Fijany and Vatan (2004) introduce an approach to represent the problem of determining minimal hitting sets as a corresponding integer programming problem. Further approaches to optimize the determination of hitting sets are discussed in Lin and Jiang (2003). All the mentioned approaches rely on (minimal) conflict sets that are the basis for calculating a set of minimal diagnoses, whereas FASTDIAG is a complete and minimal diagnosis algorithm without the need of conflict sets. It is important to mention that especially when calculating the first $n$-diagnoses (for $n>1$, i.e., not a single diagnosis), FASTDIAG can also exploit the mentioned algorithms of Lin and Jiang (2003) and Wotawa (2001) for the calculation of more than one diagnosis, that is, it is not bound to the usage of the original HSDAG algorithm.

Table 1. Precision of FastDiag versus HSDAG based approaches

<table>
<thead>
<tr>
<th>Top-n Diagnoses</th>
<th>Breadth-First (HSDAG)</th>
<th>FastDiag</th>
<th>Best-First (HSDAG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 1$</td>
<td>0.51</td>
<td>0.70</td>
<td>0.74</td>
</tr>
<tr>
<td>$n = 2$</td>
<td>0.75</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td>$n = 3$</td>
<td>0.87</td>
<td>0.97</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Note: HSDAG, hitting set directed acyclic graph.
Lin and Jiang (2002) introduce an approach to determine hitting sets on the basis of genetic algorithms; a similar approach to the determination of diagnoses is presented in Feldman et al. (2008) who introduce a stochastic fault diagnosis algorithm, which is based on greedy stochastic search. Such approaches show to significantly improve search performance; however, there is no general guarantee of completeness and diagnosis minimality. Finally, there exist a couple of algorithms that are improving the algorithmic performance of diagnosis calculation due to additional knowledge about the structural properties of the diagnosis problem. For example, Jannach and Liegl (2006) show the determination of (minimal) diagnoses for the case of conjunctive queries on database tables (the set of diagnoses can be precompiled by executing the individual parts of the query on the given data set). Siddiqi and Huang (2007) show one approach to exploit structural properties of system descriptions to improve the overall performance of diagnosis determination; in this case cones are areas in a gate with a certain structure and a certain probability of including a diagnosis, and the search process focuses on exactly those areas. FASTDIAG does not exploit specific properties of the underlying constraint set; however, taking into account such properties can further improve the performance of the algorithm. Corresponding evaluations are within the scope of future work.

6.5. Personalized diagnosis

Many of the existing diagnosis approaches do not take into account the need for personalizing the set of diagnoses to be presented to a user. Identifying diagnoses of interest in an efficient manner is a clear surplus regarding the acceptance of the underlying application, for example, users of a configurator application are not necessarily interested in minimal cardinality diagnoses (Reiter, 1987) but rather in those that correspond to their current preferences. A first step toward the application of personalization concepts in the context of knowledge-based recommendation is presented in Felfernig et al. (2009). The authors introduce an approach that calculates leading diagnoses on the basis of similarity measures used for determining n-nearest neighbors. A general approach to the identification of preferred diagnoses is introduced in DeKleer (1990), where probability estimates are used to determine the leading diagnoses with the overall goal to minimize the number of measurements needed for identifying a malfunctioning device. Basic principles of determining diagnoses in knowledge-based recommendation scenarios are discussed in Jannach and Liegl (2006). Furthermore, Froehlich et al. (1994) introduce a logical characterization of preferences that are expressed as preference relations on single diagnoses and modal logical formulas on groups of diagnoses. In contrast to our work, Froehlich et al. (1994) do not provide an algorithm to efficiently calculate preferred diagnoses. We see our work as a major contribution in this context because FASTDIAG helps to identify leading diagnoses more efficiently. Further empirical studies in different application contexts are within the major focus of our future work.

7. CONCLUSION

In this paper we have introduced a new diagnosis algorithm (FASTDIAG), which allows the efficient calculation of one diagnosis at a time with logarithmic complexity in terms of the number of consistency checks. Thus, the computational complexity for the calculation of one minimal diagnosis is equal to the calculation of one minimal conflict set in hitting set-based diagnosis approaches. The algorithm is especially applicable in settings where the number of conflict sets is equal to or larger than the number of diagnoses, or in settings where preferred (leading) diagnoses are needed. Issues for future work are the determination of repair actions for diagnoses, the further development of FASTDIAG for supporting anytime diagnosis tasks, and the conduction of further empirical studies in different configurator application domains.

REFERENCES


---

Alexander Felfernig is a Professor of applied software engineering at Graz University of Technology. Alexander is also Cofounder and Director of ConfigWorks, a company focused on the development of knowledge-based recommendation technologies. Prof. Felfernig’s research focuses on intelligent methods and algorithms supporting the development and maintenance of complex knowledge bases. Furthermore, he is interested in the application of AI techniques in the software engineering context, for example, the application of decision and recommendation technologies to make software requirements engineering processes more effective. In 2009, Dr. Felfernig received the Heinz–Zemanek Award from the Austrian Computer Society for his research.

Monika Schubert is a PhD student in the group of Applied Software Engineering at Graz University of Technology. She received her MS in software engineering and economy from Graz University of Technology. Her research focuses on knowledge-based systems, intelligent product configuration, model-based diagnosis, and product recommendation. Monika is also interested in user interaction with complex knowledge bases.

Christoph Zehentner is a PhD student in the Applied Software Engineering Group at Graz University of Technology. He received his MS in software engineering and economy from Graz University of Technology. His research focuses on testing and debugging of knowledge-based systems as well as recommendations in group decision environments. Christoph is also interested in multiple-agent system coordination and decision making.