Outlier Detection Using Default Logic

Fabrizio Angiulli[1], Rachel Ben-Eliyahu - Zohary[2], and Luigi Palopoli[3]

[1] ICAR-CNR c/o DEIS
Univ. della Calabria
87030 Rende (CS), Italy
angiulli@icar.cnr.it
[2] Comm. Systems Engineering Dept.
Ben-Gurion Univ. of the Negev
Beer-Sheva 84105, Israel
rachel@bgumail.bgu.ac.il
[3] DIMET
Univ. di Reggio Calabria
Loc. Feo di Vito
89100 Reggio Calabria, Italy
palopoli@ing.unirc.it

Abstract. Default logic is used to describe regular behavior and normal properties. We suggest to exploit the framework of default logic for detecting *outliers* - individuals who behave in an unexpected way or feature abnormal properties. The ability to locate outliers can help to maintain knowledgebase integrity and to single out irregular individuals. We first formally define the notion of an *outlier* and an *outlier witness*. We then show that finding outliers is quite complex. Indeed, we show that several versions of the outlier detection problem lie over the second level of the polynomial hierarchy. For example, the question of establishing if at least one outlier can be detected in a given propositional default theory is Σ_3^P-complete. Although outlier detection involves heavy computation, the queries involved can frequently be executed off-line, thus somewhat alleviating the difficulty of the problem. In addition, we show that outlier detection can be done in polynomial time for both the class of acyclic normal unary defaults and the class of acyclic dual normal unary defaults.

1 Introduction

Default logics were developed as a tool for reasoning with incomplete knowledge. By using default rules, we can describe how things work in general and then make some assumptions about individuals and draw conclusions about their properties and behavior.

In this paper, we suggest a somewhat different usage of default logics. The basic idea is as follows. Since default rules are used for describing regular behavior, we can exploit them for detecting individuals or elements who *do not* behave normally according to the default theory at hand. We call such entities *outliers*. An outlier is an element that shows some properties that are contrary to those that can be logically justified.

Outlier detection can be useful in several application contexts, e.g., to single out exceptional behaving individuals or system components. Note that according to our approach, exceptions are not explicitly listed in the theory as “abnormals,” as is often done in logical-based abduction [12, 2, 3]. Rather, their “abnormality” is singled out exactly because some of the properties characterizing them do not have a justification within the theory at hand. For example, suppose that it usually takes about two seconds to download a one-megabyte file from some server. Then, one day, the system is slower - instead four seconds are needed to perform the same task. While four seconds may indicate a good performance it is helpful to find the source of the delay. Another example might be that someone’s car brakes are making a strange noise. Although they seem to be functioning properly, this is not normal behavior and the car should be serviced. In this case, the car brakes are outliers and the noise is their witness.

Outlier detection can also be used for examining database integrity. If an abnormal property is discovered in a database, the source who reported this observation would have to be double-checked.

Detecting abnormal properties, that is, detecting outliers, can also lead to an update of default rules. Suppose we have the rule that birds fly, and we observe a bird, say Tweety, that does not fly. We report this occurence of an outlier in the theory to the knowledge engineer The engineer investigates the case, finds out that Tweety is, for example, a penguin, and updates the knowledgebase with the default “penguins do not fly.”

In this paper, we formally state the ideas briefly sketched above within the context of Reiter’s default logic. For simplicity, we concentrate on the propositional fragment of default logic although the generalization of such ideas to the realm of first-order defaults also worth exploring. So, whenever we use a default theory with variables, as in some of the following examples, we relate to it as an abbreviation of its grounded version.

The rest of the paper is organized as follows. In Section 2, we give preliminary definitions as well as a formal definition of the concept of an outlier. In Section 3, we describe the complexity of finding outliers in propositional default logic. Section 4 analyzes the complexity of detecting outliers in disjunction-free propositional default logics, and section 5 describes some tractable cases. Related work is discussed in Section 6. Conclusions are given in Section 7.

Because of space limitations, throughout the paper proofs of results are sketched or omitted. Full proofs can be found in [1].

2 Definitions

In this section we provide preliminary definitions for concepts we will be using throughout the paper.

2.1 Preliminaries

The following definitions will be assumed. Let T be a propositional theory. Then T^* denotes its logical closure. If S is a set of literals, then $\neg S$ denotes the set of all literals that are the negation of some literal in S.

Default logic was introduced by Reiter [13]. A *propositional default theory* Δ is a pair (D, W) consisting of a set W of propositional formulas and a set D of default rules. A *default rule* δ has the form $\frac{\phi:\psi}{\chi}$ (or, equivalently, $\phi : \psi/\chi$), where ϕ, ψ and χ are propositional formulas, called, respectively, *prerequisite*, *justification*, and *consequent* of δ. The prerequisite could be omitted, though justification and consequent are required. If $\psi = \chi$, the default rule is called *normal*. The informal meaning of a default rule δ is the following: if ϕ is known, and if it is consistent to assume ψ, then we conclude χ. An *extension* is a maximal set of conclusions that can be drawn from a theory. An extension E of a propositional default theory $\Delta = (D, W)$ can be finitely characterized through the set D_E of *generating defaults* for E w.r.t. Δ, i.e., the set $D_E = \{\phi : \psi/\chi \in D \mid \phi \in E \wedge \neg\psi \notin E\}$. Indeed, $E = (W \cup \{\chi \mid \phi : \psi/\chi \in D_E\})^*$.

Let Δ be a default theory and l a literal. Then $\Delta \models l$ means that l belongs to every extension of Δ. Similarly, for a set of literals S, $\Delta \models S$ means that every literal $l \in S$ belongs to every extension of Δ. A default theory is *coherent* if it has at least one extension.

We review some basic definitions about complexity theory, particularly, the polynomial hierarchy. The reader is referred to [6] for more on complexity theory. The classes Σ_k^P and Π_k^P are defined as follows: $\Sigma_0^P = \Pi_0^P =$ P and for all $k \geq 1$, $\Sigma_k^P = \text{NP}^{\Sigma_{k-1}^P}$, and $\Pi_k^P = \text{co-}\Sigma_k^P$. Σ_k^P models computability by a nondeterministic polynomial-time algorithm which may use an oracle, loosely speaking a subprogram that can be run with no computational cost, for solving a problem in Σ_{k-1}^P. The class D_k^P, $k \geq 1$, is defined as the class of problems that consists of the conjunction of two independent problems from Σ_k^P and Π_k^P, respectively. Note that for all $k \geq 1$, $\Sigma_k^P \subseteq D_k^P \subseteq \Sigma_{k+1}^P$. A problem A is *complete* for the class $\mathcal{C}$ iff A belongs to $\mathcal{C}$ and every problem in $\mathcal{C}$ is reducible to A by polynomial-time transformations. A well known Σ_k^P-complete problem is to decide the validity of a formula $QBE_{k,\exists}$, that is, a formula of the form $\exists X_1 \forall X_2 \ldots QX_k f(X_1, \ldots, X_k)$, where Q is $\exists$ if k is odd and is $\forall$ if k is even, $X_1, \ldots, X_k$ are disjoint set of variables, and $f(X_1, \ldots, X_k)$ is a propositional formula in $X_1, \ldots, X_k$. Analogously, the validity of a formula $QBE_{k,\forall}$, that is a formula of the form $\forall X_1 \exists X_2 \ldots QX_k f(X_1, \ldots, X_k)$, where Q is $\forall$ if k is odd and is $\exists$ if k is even, is complete for Π_k^P. Deciding the conjunction $\Phi \wedge \Psi$, where Φ is a $QBE_{k,\exists}$ formula and Ψ is a $QBE_{k,\forall}$ formula, is complete for D_k^P.

2.2 Defining outliers

Next we formalize the notion of an outlier in default logic. In order to motivate the definition and make it easy to understand, we first look at an example.

Example 1. Consider the following default theory which represents the knowledge that birds fly and penguins are birds that do not fly, and the observations that Tweety and Pini are birds and Tweety does not fly.

$$D = \left\{ \frac{Bird(x) : Fly(x)}{Fly(x)}, \frac{Penguin(x) : Bird(x)}{Bird(x)}, \frac{Penguin(x) : \neg Fly(x)}{\neg Fly(x)} \right\}$$
$$W = \{Bird(Tweety), Penguin(Pini), \neg Fly(Tweety)\}$$

This theory has two extensions. One extension is the logical closure of $W \cup \{Bird(Pini), \neg Fly(Pini)\}$ and the other is the logical closure of $W \cup \{Bird(Pini), Fly(Pini)\}$. If we look carefully at the extensions, we note that Tweety not flying is quite strange, since we know that birds fly and Tweety is a bird. Therefore, there is no apparent justification for the fact that Tweety does not fly (other than the fact $\neg Fly(Tweety)$ belonging to W). Had we been told that Tweety is a penguin, we could have explained why Tweety does not fly. But, as the theory stands now, we are not able to explain why Tweety does not fly, and, thus, Tweety is an exception. Moreover, if we are trying to nail down what makes Tweety an exception, we notice that if we would have dropped the observation $\neg Fly(Tweety)$ from W, we would have concluded the exact opposite, namely, that Tweety does fly. Thus, $\neg Fly(Tweety)$ "induces" such an exceptionality (we will call *witness* a literal like $\neg Fly(Tweety)$). Furthermore, if we drop from W both $\neg Fly(Tweety)$ and $Bird(Tweety)$, we are no longer able to conclude that Tweety flies. This implies that $Fly(Tweety)$ is a consequence of the fact that Tweety is a bird, and thus $Bird(Tweety)$ is the property of Tweety that behaves exceptionally (or the *outlier*).

From the above example, one could be induced to define an outlier as an individual, i.e., a constant, in our case *Tweety*, that possesses an exceptional property, denoted by a literal having the individual as one of its arguments, in our case $Bird(Tweety)$. However, for a conceptual viewpoint, it is much more general and flexible to single out a property of an individual which is exceptional, rather than simply the individual. That assumed, we also note that within the propositional context we deal with here, we do not explicitly have individuals distinct from their properties and, therefore, the choice is immaterial.

Based on the example and considerations mentioned above, we can define the concept of an outlier as follows.

Definition 1. Let $\Delta = (D, W)$ be a propositional default theory such that W is consistent and $l \in W$ is a literal. If there exists a set of literals $S \subseteq W$ such that:

1. $(D, W_S) \models \neg S$, and
2. $(D, W_{S,l}) \not\models \neg S$.

where $W_S = W \setminus S$ and $W_{S,l} = W_S \setminus \{l\}$, then we say that l is an *outlier* in Δ and S is an *outlier witness set* for l in Δ.

According to this definition, a literal l is an outlier if and only if there is an exceptional property, denoted by a set of literals S, holding in *every* extension of the theory.

The exceptional property is the outlier witness for l. Thus, according to this definition, in the default theory of Example 1 above we should conclude that $Bird(Tweety)$ denotes an outlier and $\{\neg Fly(Tweety)\}$ is its witness. Note that we have defined an outlier witness to be a set, not necessarily a single literal since in some theories taking a single literal does not suffice to form a witness for a given outlier being that all witnesses of such an outlier have a cardinality strictly larger than one.

Example 2. Consider the default theory $\Delta = (D, W)$, where the set of default rules D conveys the following information about weather and traffic in a small town in southern California:

1. $\frac{July \wedge Weekend : \neg Traffic\_Jam \wedge \neg Rain}{\neg Traffic\_Jam \wedge \neg Rain}$ - that is, normally during a July weekend there are no traffic jams nor any rain.
2. $\frac{January : Rain}{Rain}, \frac{January : \neg Rain}{\neg Rain}$ - in January it sometimes rains and sometimes it doesn't rain.
3. $\frac{Weekend \wedge Traffic\_Jam : Accident \vee Rain}{Accident \vee Rain}$ - If there is a traffic jam in the weekend then normally it must be raining or there would have been an accident.

Suppose also that $W = \{July, Weekend, Traffic\_Jam, Rain\}$. Then, the set $S = \{Traffic\_Jam, Rain\}$ is an outlier witness for both $Weekend$ and $July$. Moreover, S is a *minimal* outlier witness set for either $Weekend$ or $July$, since deleting one of the members from S will render S not being a witness set.

Here is another example.

Example 3. Consider the following default theory Δ:

$$D = \left\{ \frac{Income(x) \wedge Adult(x) : Works(x)}{Works(x)}, \frac{FlyingS(x) : InterestTakeOff(x)}{InterestTakeOff(x)}, \frac{FlyingS(x) : InterestNavigate(x)}{InterestNavigate(x)} \right\}$$

$$W = \{Income(Johnny), Adult(Johnny), \neg Works(Johnny), FlyingS(Johnny), \neg InterestTakeOff(Johnny)\}$$

This theory claims that normally adults who have a monthly income work, and students who take flying lessons are interested in learning how to take off and navigate. The observations are that Johnny is an adult who has a monthly income, but he does not work. He is also a student in a flying school but he is not interested in learning how to take-off. Based on the events of September 11, 2001, we'd like our system to conclude that Johnny is the argument of two outliers. Indeed, the reader can verify that the following facts are true:

1. $(D, W_{\neg Works(Johnny)}) \models Works(Johnny)$,
2. $(D, W_{\neg InterestTakeOff(Johnny)}) \models InterestTakeOff(Johnny)$,
3. $(D, W_{\neg Works(Johnny), Adult(Johnny)}) \not\models Works(Johnny)$, and
4. $(D, W_{\neg InterestTakeOff(Johnny), FlyingS(Johnny)}) \not\models InterestTakeOff(Johnny)$

Hence, both $\neg Works(Johnny)$ and $\neg InterestTakeOff(Johnny)$ are outlier witnesses, while $Adult(Johnny)$ and $FlyingS(Johnny)$ are outliers. Note that $Income(Johnny)$ is also an outlier, with the witness $\neg Works(Johnny)$.

2.3 Defining outlier detection problems

In order to state the computational complexity of detecting outliers, in the rest of the work we refer to the following problems (also referred to as *queries*) defined for an input default theory $\Delta = (D, W)$:

$Q0$: Given Δ, does there exist an outlier in Δ ?
$Q1$: Given Δ and a literal $l \in W$, is there any outlier witness for l in Δ ?
$Q2$: Given Δ and a set of literals $S \subseteq W$, is S a witness for any outlier l in Δ ?
$Q3$: Given Δ, a set of literals $S \subseteq W$, and a literal $l \in W$, is S a witness for l in Δ ?

3 General complexity results

In this section we analyze the complexity associated with detecting outliers. First, we give some preliminary definitions involving notation.

Let L be a set of literals such that $\ell \in L$ implies that $\neg\ell \notin L$. Then we denote by $\mathcal{T}_L$ the truth assignment on the set of letters occurring in L such that, for each positive literal $p \in L$, $\mathcal{T}_L(p) = \mathbf{true}$, and for each negative literal $\neg p \in L$, $\mathcal{T}_L(p) = \mathbf{false}$.

Let T be a truth assignment on the set $x_1, \ldots, x_n$ of letters. Then we denote by $Lit(T)$ the set of literals $\{\ell_1, \ldots, \ell_n\}$, such that ℓ_i is x_i if $T(x_i) = \mathbf{true}$ and is $\neg x_i$ if $T(x_i) = \mathbf{false}$, for $i = 1, \ldots, n$.

Theorem 1. *$Q0$ is Σ_3^P-complete.*

Proof. (Membership) Given a a theory $\Delta = (D, W)$, we must show that there exists a literal l in W and a subset $S = \{s_1, \ldots, s_n\}$ of W such that $(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n$ (query q') and $(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge s_n$ (query q''). Query q' is Π_2^P-complete, while query q'' is Σ_2^P-complete [8, 14]. Thus, we can build a polynomial-time nondeterministic Turing machine with a Σ_2^P oracle, solving query $Q0$ as follows: the machine guesses both the literal l and the set S and then solves queries q' and q'' using two calls to the oracle.

(Hardness) Let $\Phi = \exists X \forall Y \exists Z f(X, Y, Z)$ be a quantified boolean formula, where $X = x_1, \ldots, x_n, Y = y_1, \ldots, y_m$, and Z are disjoint set of variables. We associate with Φ the default theory $\Delta(\Phi) = (D(\Phi), W(\Phi))$, where $W(\Phi)$ is the set $\{l, s_1, \overline{s}_1, \ldots, s_n, \overline{s}_n\}$ consisting of new letters distinct from those occurring in Φ, and $D(\Phi) = D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$:

$$
\begin{aligned}
D_1 &= \Big\{ \delta_{1,i} = \tfrac{:\neg s_i \wedge x_i \wedge e_i}{x_i \wedge e_i}, \\
&\qquad \overline{\delta}_{1,i} = \tfrac{:\neg \overline{s}_i \wedge \neg x_i \wedge e_i}{\neg x_i \wedge e_i} \mid i = 1, \ldots, n \Big\} \\
D_2 &= \Big\{ \delta_{2,i} = \tfrac{:\neg s_i \wedge \neg \overline{s}_i \wedge \neg\alpha \wedge \beta}{\beta} \mid i = 1, \ldots, n \Big\} \cup \\
&\quad \cup \Big\{ \delta_2 = \tfrac{\beta : \alpha}{\alpha} \Big\} \\
D_3 &= \Big\{ \delta_{3,j} = \tfrac{:y_j}{y_j}, \overline{\delta}_{3,j} = \tfrac{:\neg y_j}{\neg y_j} \mid j = 1, \ldots, m \Big\} \\
D_4 &= \Big\{ \delta_4 = \tfrac{l \wedge e_1 \wedge \ldots \wedge e_n : f(X,Y,Z) \wedge g}{g} \Big\} \\
D_5 &= \Big\{ \delta_{5,i} = \tfrac{g : \neg s_i}{\neg s_i}, \overline{\delta}_{5,i} = \tfrac{g : \neg \overline{s}_i}{\neg \overline{s}_i} \mid i = 1, \ldots, n \Big\}
\end{aligned}
$$

where also α, β, g, $e_1, \ldots, e_n$ are new variables distinct from those occurring in Φ. Clearly, $W(\Phi)$ is consistent and $\Delta(\Phi)$ can be built in polynomial time. We next show that Φ is valid iff there exists an outlier in $\Delta(\Phi)$.

In the rest of the proof we denote by $\sigma(s_i)$ ($\widehat{\sigma}(x_i)$ resp.) the literal x_i (s_i resp.) and by $\sigma(\overline{s}_i)$ ($\widehat{\sigma}(\neg x_i)$ resp.) the literal $\neg x_i$ ($\overline{s}_i$ resp.), for $i = 1, \ldots, n$. Letting S be a subset of $\{s_1, \overline{s}_1, \ldots, s_n, \overline{s}_n\}$ ($\{x_1, \neg x_1, \ldots, x_n, \neg x_n\}$ resp.), we denote by $\sigma(S)$ ($\widehat{\sigma}(S)$ resp.) the set $\{\sigma(s) \mid s \in S\}$ ($\{\widehat{\sigma}(s) \mid s \in S\}$ resp.).

($\Rightarrow$) Suppose that Φ is valid. Then we can show that l is an outlier in $\Delta(\Phi)$. As Φ is valid, then there exists a truth assignment T_X on the set X of variables such that T_X satisfies $\forall Y \exists Z f(X, Y, Z)$. Let $S = \widehat{\sigma}(Lit(T_X))$. It can be shown that we can associate to each truth assignment T_Y on the set Y of variables, one and only one extension E_Y of $(D(\Phi), W(\Phi)_S)$. In particular, $E_Y \supseteq Lit(T_X) \cup Lit(T_Y)$. As Φ is valid, then $\neg f(X, Y, Z) \notin E_Y$ and $E_Y \models \neg S$. Furthermore, since there is no other extension of $(D(\Phi), W(\Phi)_S)$, then $(D(\Phi), W(\Phi)_S) \models \neg S$.

Consider now the theory $(D(\Phi), W(\Phi)_{S,l})$. We note that the literal l appears in the precondition of rule δ_4, whose conclusion g represents, in turn, the precondition of the rules in the set D_5, rules that allow to conclude $\neg S$, and that l does not appear in the conclusion of any rule of $D(\Phi)$. Thus $(D(\Phi), W(\Phi)_{S,l}) \not\models \neg S$. Hence l is an outlier in $\Delta(\Phi)$.

($\Leftarrow$) Suppose that there exists an outlier in $\Delta(\Phi)$. It can be shown that the outlier is l. Hence, there exists a nonempty set of literals $S \subseteq W(\Phi) - \{l\}$ such that S is an outlier witness for l in $\Delta(\Phi)$. It can then be shown that $S = \{s'_1, \ldots, s'_n\}$, where s'_i is either s_i or $\overline{s}_i$, for $i = 1, \ldots, n$. Now we show that $\mathcal{T}_{\sigma(S)}$ satisfies $\forall Y \exists Z f(X, Y, Z)$, i.e. that Φ is valid. For each set $L = \{\ell_1, \ldots, \ell_m\}$, where ℓ_j is either y_j or $\neg y_j$, for $j = 1, \ldots, m$, there exists one extension E_L of $(D(\Phi), W(\Phi) - S)$ such that $E_L \supseteq L$. We note also that $E_L \supseteq \sigma(S)$. Thus, in order for l to be an outlier in $\Delta(\Phi)$, it must be the case that for each set L, $\neg f(X, Y, Z) \notin E_Y$ i.e., that $\mathcal{T}_{\sigma(S) \cup L}$ satisfies $f(X, Y, Z)$. Hence, we can conclude that Φ is valid.

Theorem 2. *$Q1$ is Σ_3^P-complete.*

Proof. The proof is analogous to that used in Theorem 1.

Theorem 3. *$Q2$ is D_2^P-complete.*

Proof. (Membership) Given a theory $\Delta = (D, W)$ and a subset $S = \{s_1, \ldots, s_n\} \subseteq W$, we should verify that $(D, W_S) \models \neg s_1 \wedge \ldots \wedge \neg s_n$ (statement q') and there exists a literal $l \in W$ such that $(D, W_{S,l}) \not\models \neg s_1 \wedge \ldots \wedge \neg s_n$ (statement q''). Solving q' is in Π_2^P. As for statement q'', it can be decided by a polynomial time nondeterministic Turing machine, with an oracle in NP, that (a) guesses both the literal $l \in W$ and the set $D_E \subseteq D$ of generating defaults of an extension E of $(D, W_{S,l})$ together with an order of these defaults; (b) checks the necessary and sufficient conditions that D_E must satisfy to be a set of generating defaults for E (see [15] for a detailed description of these conditions), by multiple calls to the oracle; and (c) verifies that $\neg s_1 \wedge \ldots \wedge \neg s_n \notin E$ by other calls to the oracle. It can be shown that the total number of calls to the oracle is polynomially bounded. Thus, $Q2$ is the conjunction of two independent problems, one in Π_2^P (q') and the other in Σ_2^P (q''), i.e. it is in D_2^P.

(Hardness) Let $\Delta_1 = (D_1, W_1)$ and $\Delta_2 = (D_2, W_2)$ be two propositional default theories such that W_1 and W_2 are consistent, let s_1, s_2 be two letters, and let q be the statement $(\Delta_1 \models s_1 \wedge \Delta_2 \not\models s_2)$. W.l.o.g, we can assume that Δ_1 and Δ_2 contain different letters, the letter s_1 occurs in D_1 but not in W_1 (and, from the previous condition, not in Δ_2), and the letter s_2 occurs in D_2 but not in W_2 (and hence not in Δ_1). We associate with q the default theory $\Delta(q) = (D(q), W(q))$ defined as follows. Let $D_1 = \{\frac{\alpha_i:\beta_i}{\gamma_i} \mid i = 1, \ldots, n\}$ and let $L_1 = \{\ell_1, \ldots, \ell_m\} \subseteq W_1$ be all the literals belonging to W_1, then $D(q) = \{\frac{s_2 \wedge \alpha_i:\beta_i}{\gamma_i} \mid i = 1, \ldots, n\} \cup \{\delta_j = \frac{:\neg\ell_i \wedge \neg\mu \wedge \nu}{\nu} \mid j = 1, \ldots, m\} \cup \{\delta_0 = \frac{\nu:\mu}{\mu}\} \cup D_2$, and $W(q) = W_1 \cup W_2 \cup \{\neg s_1, s_2\}$, where ν and μ are new letters distinct from those occurring in Δ_1 and Δ_2, and from s_1 and s_2. It can be shown that q is true iff $\{\neg s_1\}$ is a witness for some outlier in $\Delta(q)$. We note that q is the conjunction of a Π_2^P-hard and a Σ_2^P-hard independent problems, thus this proves the hardness part.

Theorem 4. *$Q3$ is D_2^P-complete.*

4 Disjunction-free theories

Disjunction-free theories form a significant subset of propositional default theories because they are equivalent to extended logic programs under stable model semantics [7]. A finite propositional theory $\Delta = (D, W)$ is *disjunction-free* (DF in short), if W is a set of literals, and the precondition, justification and consequence of each default in D is a conjunction of literals. As we see below, outlier detection for DF theories is still quite complex.

Theorem 5. *$Q0$ restricted to disjunction-free theories is Σ_2^P-complete.*

Proof. (Membership) The membership proof is analogous to that of Theorem 1. We note that when disjunction-free theories are considered, q' and q'' are co-NP-complete and NP-complete, respectively.

(Hardness) Let $\Phi = \exists X \forall Y f(X, Y)$ be a quantified boolean formula, where $X = x_1, \ldots, x_n$ and $Y = y_1, \ldots, y_m$ are disjoint set of variables, and $f(X, Y) = C_1 \wedge \ldots \wedge C_r$, with $C_k = t_{k,1} \vee \ldots \vee t_{k,u_k}$, and each $t_{k,1}, \ldots, t_{k,u_k}$ is a literal, for $k = 1, \ldots, r$. We associate to Φ the default theory $\Delta(\Phi) = (D(\Phi), W(\Phi))$, with $W(\Phi)$ the set $\{l, s_1, \overline{s}_1, \ldots, s_n, \overline{s}_n\}$ of new letters distinct from those occurring in Φ, and $D(\Phi) = D_1 \cup D_2 \cup D_3 \cup D_5 \cup D'_1 \cup D'_2$, where D_1, D_2, D_3 and D_5 are the sets of defaults as in Theorem 1 and D'_1 and D'_2 are the following sets of defaults:

$$D'_1 = \left\{ \frac{t_{k,h}:c_k}{c_k} \mid k = 1, \ldots, r;\ h = 1, \ldots, u_k \right\}$$
$$D'_2 = \left\{ \frac{l \wedge e_1 \wedge \ldots \wedge e_n \wedge c_1 \wedge \ldots \wedge c_r : g}{g} \right\}$$

where also $\alpha, \beta, g, e_1, \ldots, e_n, c_1, \ldots, c_r$ are new variables distinct from those occurring in Φ. Clearly, $W(\Phi)$ is consistent and $\Delta(\Phi)$ can be built in polynomial time. The rest of the proof is similar to that of Theorem 1.

Theorem 6. *$Q1$ restricted to disjunction-free theories is Σ_2^P-complete.*

Theorem 7. *$Q2$ and $Q3$ restricted to disjunction-free theories are D^P-complete.*

5 Tractable Cases

In this section, we look for some classes of default theories for which outlier detection is computationally tractable.

Definition 2. A default theory is *normal mixed unary* (NMU in short) iff W is a set of literals and D is a set of defaults of the form $\frac{y:x}{x}$ where y is either missing or a literal and x is a literal.

Definition 3. An NMU default theory is normal unary (NU in short) iff the prerequisite of each default is either missing or positive. An NMU default theory is dual normal unary (DNU in short) iff the prerequisite of each default is either missing or negative.

Thus, NMU, NU, and DNU theories have a quite simple structure. In spite of that, the complexity of detecting outliers from these theories remain often quite high, as demonstrated by the following results (proofs are omitted for the sake of brevity).

Theorem 8. *The following hold over* NMU *default theories:*

– $Q0$ *and* $Q1$ *are* Σ_2^P*-complete.*
– $Q2$ *and* $Q3$ *are* D^P*-complete.*

Theorem 9. *The following hold over* NU *and* DNU *default theories:*

– $Q0$ *and* $Q1$ *are NP-complete.*
– $Q2$ *and* $Q3$ *are in P.*

Thus, restricting our attention to NMU, NU, or DNU theories does not suffice to attain tractability of the most general queries $Q0$ and $Q1$. Some further restriction is needed, which is considered next.

Theorem 10 ([9] [16]). *Suppose Δ is a normal (dual normal) unary default theory. We can decide whether a literal belongs to every extension of Δ in time $O(n^2)$, where n is the length of the theory.*

Definition 4. The *atomic dependency graph* of an NMU default theory Δ is a directed graph whose nodes are all atoms in the language of Δ, and such that there is an arc directed from p to q iff there is a default in Δ in which p or $\neg p$ is a prerequisite and q or $\neg q$ is a consequence.

Definition 5. A normal (dual normal) unary default theory is *acyclic* iff its atomic dependency graph is acyclic.

Theorem 11. *Queries $Q0$, $Q1$, $Q2$ and $Q3$, restricted to the class of acyclic NU or acyclic DNU default theories can be solved in polynomial time in the size of the input theory.*

Proof. It can be shown that for any acyclic NMU default theory $\Delta = (D, W)$ such that W is consistent and for any literal l in W, any minimal outlier witness set for l in Δ is at most 1 in size. Theorem's statement then follows from Theorem 10.

6 Related Work

The research on logical-based abduction [12, 2, 3] is related to outlier detection. In the framework of logic-based abduction, the domain knowledge is described using a logical theory T. A subset X of hypotheses is an abduction explanation to a set of manifestations M if $T \bigcup X$ is a consistent theory that entails M. Abduction resembles outlier detection in that it deals with exceptional situations.

The work most relevant to our study is perhaps the paper by Eiter, Gottlob, and Leone on abduction from default theories [4]. There, the authors have presented a basic model of abduction from default logic and analyzed the complexity of the main abductive reasoning tasks. They presented two modes of abductions: one based on brave reasoning and the other on cautious reasoning. According to these authors, a default abduction problem (DAP) is a tuple $\langle H, M, W, D \rangle$ where H is a set of ground literals called *hypotheses*, M is a set of ground literals called *observations*, and (D, W) is a default theory. Their goal, in general, was to explain some observations from M by using various hypotheses in the context of the default theory (D, W). They suggest the following definition for an explanation:

Definition 6 ([4]). *Let $P = \langle H, M, D, W \rangle$ be a DAP and let $E \subseteq H$. Then, E is a* skeptical explanation *for P iff*

1. $(D, W \bigcup E) \models M$*, and*
2. $(D, W \bigcup E)$ *has a consistent extension.*

There is a close relationship between outliers and skeptical explanations, as the following theorem states. The theorem also holds for ordered semi-normal default theories [5].

Theorem 12. *Let $\Delta = (D, W)$ be a normal default theory, where W is consistent. Let $l \in W$ and $S \subseteq W$. S is an outlier witness set for l iff $\{l\}$ is a minimal skeptical explanation for $\neg S$ in the DAP $P = \langle \{l\}, \neg S, D, W_{S,l} \rangle$*

Hence, we can say that S is an outlier witness for l if $l \in W$, l is a skeptical explanation for S, but still $\neg S$ holds in every extension of the theory.

Despite the close relationship between outlier detection and abduction demonstrated by the above theorem (especially for normal defaults) we believe that there is a significant difference between the two concepts. In abduction, we have to single out a set of manifestations and a set of potential explanations. Outlier detection, on the other hand, has much more to do with knowledge discovery. The task in outlier detection is to *learn* who the exceptionals (the outliers), or the suspects, if you wish, are, and to justify the suspicion (that is, list the outlier witnesses).

It also turns out that reducing outlier detection queries to abduction and vice versa is not straightforward, and therefore, when analyzing the computational complexities involved in answering outlier detection queries we have preferred to use the classical Boolean formula satisfiability problems.

Theory \ Query	$Q0$	$Q1$	$Q2$	$Q3$
Propositional	Σ_3^P-c	Σ_3^P-c	D_2^P-c	D_2^P-c
DF, NMU	Σ_2^P-c	Σ_2^P-c	D^P-c	D^P-c
NU, DNU	NP-c	NP-c	P	P
Acy. NU, Acy. DNU	P	P	P	P

Table 1. Complexity results for outlier detection

7 Conclusion

Suppose you are walking down the street and you see a blind person walking in the opposite direction. You believe he is blind because he is feeling his way with a walking stick. Suddenly, something falls out of his bag, and to your surprise, he finds it immediately without probing about with his fingers, as you would expect for a blind person. This kind of behavior renders the "blind" person suspicious.

The purpose of this paper has been to formally mimic this type of reasoning using default logic. We have formally defined the notion of an outlier and an outlier witness, and analyzed the complexities involved, pointing out some non-trivial tractable subsets. The complexity results are summarized in Table 1, where $\mathcal{C}$-c stands for $\mathcal{C}$-complete. As explained in the introduction, outlier detection can also be used for maintaining database integrity and completeness.

This work can be extended in several ways. First, we can develop the concept of outliers in other frameworks of default databases, like System Z [11] and Circumscription [10]. Second, we can look for intelligent heuristics that will enable us to perform the involved heavy computational task more efficiently. Third, we can study the problem from the perspective of default theories as a "semantic check toolkit" for relational databases.

Acknowledgements. The authors gratefully thank Michael Gelfond for some fruitful discussions on the subject of the paper and Francesco Scarcello for providing insights about some of the computational complexity tools we have used.

References

1. Fabrizio Angiulli, Rachel Ben-Eliyahu-Zohary, and Luigi Palopoli. Outlier detection using default logic, 2003. Full manuscript.
2. Luca Console, Daniele Theseider Dupré, and Pietro Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1(5):661–690, 1991.
3. Thomas Eiter and Georg Gottlob. The complexity of logic-based abduction. *JACM*, 42(1):3–42, 1995.
4. Thomas Eiter, Georg Gottlob, and Nicola Leone. Semantics and complexity of abduction from default theories. *Artificial Intelligence*, 90(1-2):177–223, 1997.
5. David W. Etherington. Formalizing nonmonotonic reasoning systems. *Artificial Intelligence*, 31:41–85, 1987.

6. M. R. Garey and D. S. Johnson. *Computers and intractability; a guide to the theory of NP-completeness*. W.H. Freeman, 1979.
7. Michael Gelfond and Vladimir Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9:365–385, 1991.
8. George Gottlob. Complexity results for nonmonotonic logics. *J. Logic Computat.*, 2(3):397–425, 1992.
9. Henry A. Kautz and Bart Selman. Hard problems for simple default logics. *Artificial Intelligence*, 49:243–279, 1991.
10. John McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.
11. Judea Pearl. System z: a natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *TARK-90*, pages 121–135, Monterey, CA, 1990.
12. David Poole. Normality and faults in logic-based diagnosis. In *IJCAI-89*, pages 1304–1310, 1989.
13. Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.
14. Jonathan Stillman. The complexity of propositional default logics. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 794–800, San Jose, CA, July 1992.
15. A. Zhang and W. Marek. On the classification and existence of structures in default logic. *Fundamenta Informaticae*, 13(4):485–499, 1990.
16. Rachel Ben-Eliyahu Zohary. Yet some more complexity results for default logic. *Artificial Intelligence*, 139(1):1–20, 2002.