On Explanations to Observations

Leo Bertossi[1] and John Doe[2]

[1] Carleton University, bertossi@scs.carleton.ca
[2] Π-Institute

Abstract. In this work we consider the problem of characterizing model-based diagnosis from ...

1 Abductive and Diagnosis Explanations

Consider a propositional language $L(P)$, where P is the set of all propositional variables or atoms. Consider also a subset $A \subseteq P$ of abductible atoms. This set is determined a priori. The explanations to observations will be given as formulas in the sublanguage $L(A)$, while the observations can be represented as formulas of $L(P)$, that is, of the richer language.[3] A theory T of the domain under consideration, will be present. T will be a consistent set of formulas of $L(P)$.

Definition 1. An hypothesis for O with respect to T is a valuation $v : A \to \{0,1\}$ that can be extended to a valuation $v' : P \to \{0,1\}$ that satisfies $T \cup O$, and ..... $\Box$

Remark 1. Notice from Definition 1 that for every valuation $v : A \to \{0,1\}$ there is a unique formula $\varphi_v \in L(A)$ that is a conjunction of literals and that completely determines the valuation (in the sense that any valuation on A that satisfies φ_v coincides with v). This formula can be defined as: $\bigwedge_{p \in A} p^v$, where p^v is defined by p if $v(p) = 1$, and $\neg p$ if $v(p) = 0$. $\Box$

In $Con_A(T \cup O)$, we consider all logical consequences, in the weaker language of explanations $L(A)$, of the domain theory together with the observation. This set constrains all the possible worlds, in the sense that we will only accept those worlds (hypothesis) that satisfy $Con_A(T \cup O)$:

Proposition 1. Let $v : A \to \{0,1\}$. The following are equivalent:

(a) v is an hypothesis for O wrt T.
(b) $v \models Con_A(T \cup O)$. $\Box$

Remark 2. We can see that v is an hypothesis for O wrt T iff $\varphi_v \models Con_A(T \cup O)$.

[3] Note that in the literature, an observation is usually defined as a formula of $L(P \setminus A)$, i.e. without any abductible atoms (see [2])

An abductive explanation is usually defined (see Definition 1 above) as a formula that, together with the domain theory, implies the observation. We have obtained a valuation that satisfies all logical consequences, in the language $L(A)$, of the observation (together with the domain theory), i.e. we have a formula φ_v that implies all logical consequences of $T \cup O$. According to this, an hypothesis v can be considered in some sense as a form of abductive explanation.

Example 1. (see [6, example 2.2]) Among the sentences in the domain theory we have:

$$\begin{aligned} \mathit{and\_gate} \wedge \neg \mathit{abnormal}_1 &\to q, \qquad (1) \\ \mathit{xor\_gate} \wedge \neg \mathit{abnormal}_2 &\to s, \\ \mathit{or\_gate} \wedge \neg \mathit{abnormal}_3 &\to u. \end{aligned}$$

That is, there are several types of abnormality of system components. Here the abductible atoms are $\mathit{abnormal}_1$ (as in (1)), $\mathit{abnormal}_2$, $\mathit{abnormal}_3$. In front of an observation O for the system, we look for a diagnosis, more precisely, a minimal hypothesis for the failure detected through observation O. Such a diagnosis, or minimal hypothesis, could be $\mathit{abnormal}_1 \wedge \neg \mathit{abnormal}_2 \wedge \neg \mathit{abnormal}_3$. In this case, $\{\mathit{abnormal}_1\}$ would be a minimal set of abnormalities. □

Every hypothesis φ of the form $\bigwedge_{p \in A} \pm p$ determines a world. In that regard, it is a strong explanation for O wrt T.

Proposition 2. If $\varphi = \bigwedge_{\Delta} p \wedge \bigwedge_{A \smallsetminus \Delta} \neg p$ is a minimal hypothesis for O wrt T, then

$$T \cup O \cup \bigwedge_{A \smallsetminus \Delta} \neg p \models \bigwedge_{\Delta} p. \qquad (2)$$

Proof: If φ is minimal hypothesis, then Δ is a minimal subset of A with $T \cup O \bigwedge_{A - \Delta} \neg p$ consistent (see Proposition 1). Then, for every $p_0 \in \Delta$: $T \cup O \cup \bigwedge_{(A \smallsetminus \Delta) \cup \{p_0\}} \neg p$ is inconsistent.

Then, $T \cup O \cup \bigwedge_{A \smallsetminus \Delta} \neg p \models p_0$. This holds for every $p_0 \in \Delta$. In consequence, $T \cup O \cup \bigwedge_{A \smallsetminus \Delta} \neg p \models \bigwedge_{\Delta} p$, and we obtain (2). □

1.1 The extreme case

Here we develop ....

In Figure 1 we can see that ...

The nice figure can also be in PDF, and the file can be run with PDFLatex.

2 The Case of Causal Theories

In this section we will consider mainly causal theories given by a set of definite propositional clauses.

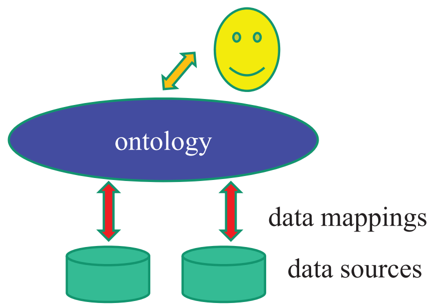

Fig. 1. This is a nice picture.

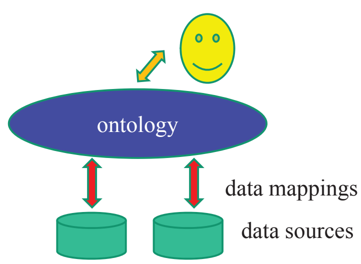

Fig. 2. The same pic. now in a PDF source

Let $T \subseteq L(P)$ be a set of definite propositional clauses. $A \subseteq P$ is the set of atoms not appearing in any head of any clause in T. This is the set of abductible atoms. Notice that if there are atoms as facts in T, i.e. definite clauses with empty body, then these atoms are not abductible. If T is clausal, then there are no facts in it.

$$Circum(T \wedge O; A; P \smallsetminus A) \tag{3}$$

In (3) we have the propositional circumscription of the abductible atoms in $T \wedge O$ with variable non-abductible atoms (see Appendix A).

References

[1] L. Bertossi. On Propositional Circumscription. DRAFT

[2] L. Console, D. Theseider-Dupre and P. Torasso. A Completion Semantics for Object-level Abduction. *Proc. AAAI Symposium on Automated Abduction*, Stanford, AAAI Press, 1990, pp. 72-76.

[3] L. Console, D. Theseider-Dupre and P. Torasso. On the Relationship Between Abduction and Deduction. *J. Logic and Computation*, 1991, 1:661-690.

[4] V. Lifschitz Computing Circumscription. *Proc. IJCAI'85*, Morgan Kaufmann Pub., 1985, pp. 121-127.

[5] V. Lifschitz Circumscription. *Handbook of Logic in Artificial Intelligence and Logic Programming*, Oxford University Press, 1994.

[6] R. Reiter. A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 1987, 32:57-95.

[7] R. Reiter and A.K. Mackworth. A Logical Framework for Depiction and Image Interpretation. *Artificial Intelligence*, 1989/90, 41:125-155.

A Propositional Circumscription

Fixed a propositional language $L(P)$ and sets of atoms $C, V \subseteq P$, we can compare two truth valuations $v, v' : P \leftarrow \{0,1\}$: $v <^{C;V} v'$ iff

i) $\{p \in C \mid v(p) = 1\} \subset \{p \in C \mid v'(p) = 1\}$ and
ii) $v(p) = v'(p)$ for all $p \in P \smallsetminus (C \cup V)$.

Now, given a propositional formula $\varphi \in L(P)$, we denote by $Circum(\varphi; C; V)$ the class of $<^{C;V}$-minimal models of φ, and call it "the circumscription of the atoms in C in φ with variable atoms in V (and fixed atoms in $P \smallsetminus (C \cup V)$)".

This class can be axiomatized by a propositional formula (see [1, 5]).