Legal Knowledge and Information Systems
G. Sileno et al. (Eds.)

doi:10.3233/FAIA230942

Inferring New Classifications in Legal Case-Based Reasoning

Cecilia DI FLORIO [b], Xinghan LIU [a], Emiliano LORINI [a], Antonino ROTOLO [b] and Giovanni SARTOR [b]

[a] *IRIT-CNRS, University of Toulouse, France*
[b] *Alma Human AI, University of Bologna, Italy*

Abstract. This article continues the research initiated in [1,2], which established a connection between Boolean classifiers and legal case-based reasoning. We relax the assumption that case bases are such that all situations have been decided in favour of the defendant or the plaintiff and we introduce an inductive strategy for assigning plausible outcomes to undecided cases. Using counterfactual reasoning, we propose a method to determine whether, at each step of the induction, a feature is a factor, i.e., it consistently favours a single outcome, or is irrelevant, i.e., it is does not favour any outcome, or is ambiguous, i.e., it favours opposite outcomes.

Keywords. Case-based reasoning, Modal Logic for classifiers, Explainable AI, Defeasible reasoning

1. Introduction

Case-based reasoning (CBR) has played an important role in AI & law research. While different models have been adopted, factor-based representations have been most popular, where, factors are features of cases that support possible outcomes. This line of research started with HYPO [3,4], in whose framework a factor-based representation enables various patterns of analogical reasoning: citing a precedent, distinguishing it, and reasoning a fortiori. John Horty has studied the logical properties of legal CBR, developing a method for determining whether a certain decision is consistent or inconsistent with a case base [5]. Applying the idea of a fortiori reasoning, he has argued that any decision in a new case would be inconsistent with a precedent, if that decision is different from the precedent's even though the new case has a more (or equally) inclusive set of facts favouring the precedent's conclusion and a less (or equally) inclusive set of factors against that conclusion. Further developments are the so-called *reason model*, where the reason is distinguished in a case, namely, it is the specified set of elements that the judge considers to provide sufficient support to the case decision, outweighing the factors to the contrary. Scalable factors (also called dimension) have also been studied.

Recent research, such as in [6,7,2], has argued that some of the usual assumptions in legal CBR can be relaxed or weakened in order to deal with more realistic scenarios. In particular, in the context of the reason model:

- **Inconsistent case bases -** The concept of consistency assumes that the initial background case base is consistent, which is not a realistic assumption; Horty

[8] outlined a broader interpretation of the reason model's constraint concept, extending it to also encompassing inconsistent case bases; recent investigations have been accordingly developed (such as [6,9]);
- **Incomplete knowledge -** The idea of complete knowledge assumes that any case base is such that all situations have been decided in favour of the defendant or the plaintiff and thus, for example, that all features in the base have a direction (i.e., if they are pro and con factors). This assumption can also be unrealistic [10,2].

In this paper we follow the approach of [2] and address the second point. In particular, we work on inference mechanisms with incomplete knowledge for the identification of factors, among the features that are present in a case base, and the determination of their direction [11]. Our approach, which is based on the logical framework described in [12,13], adopts the view that a legal case-based reasoner is nothing but a binary classifier [1,2]. We argue that, with incomplete-knowledge, the identification of which features are factors and of their direction, can be achieved through counterfactual reasoning.

In observing the behaviour of classifiers, a fundamental qualification of any feature p amounts to checking if p can discriminate in favor of an outcome: this happens if there exists a case such that the feature's absence in that case would have lead to the opposite decision. Given that basic idea, a feature can

1. be a *factor* if it discriminates in favour of an outcome and never discriminates in favour of the opposite outcome;
2. be *irrelevant*, if it never discriminates in favour of any outcome;
3. be *ambiguous* if it discriminates for an outcome and also discriminates for the opposite outcome.

Taking into account the above taxonomy, we can devise a broad class of classifiers to infer new cases in the context of incomplete knowledge. In particular,

- On the basis of a fortiori principle we develop an *inductive procedure for inferring classifications for cases.*
- The inductive procedure is defined in a static way, without changing the starting model, and in a dynamic way, by updating the model, as introduced in [12,13].
- We show how the process described in terms of static inference can be equivalently described in terms of dynamic model updating.

The methodology adopts an inference mode in which *we tolerate the presence of ambiguous features, and we provide a way of inferring the classification of cases for which there are precedents that support opposite decisions.*

The paper is structured as follows. Section 2 recalls the logical background for binary classifiers of [12,13,1]. Section 3 recalls the formal account of incomplete knowledge of [2] and applies in Section 4 to characterise the notions of factor, irrelevant and ambiguous feature. Section 5 defines the mechanism for inferring classifications. Some conclusions end the paper.

2. Background: Logic of Binary Classifiers

In this section we briefly recall language and semantics of binary-input classifier logic BCL first appeared in [12,13] and already introduced to model legal CBR [1,2].

We denote a finite set of atomic propositions by Atm, which is a disjoint union $Atm_0 \cup Dec$, where the former stands for the set of input variables of the classifier and $Dec = \{\mathtt{t}(c) : c \in Val = \{0,1,?\}\}$ for the set of all three possible output values of the classifier. In addition, let $Val = \{1,0,?\}$ where elements stand for *plaintiff wins*, *defendant wins* and *indeterminacy* respectively. Hence $\mathtt{t}(c)$ reads as "the actual decision/outcome (of the judge/classifier) takes value c". For $c \in \{0,1\}$, the "opposite" $\overline{c}$ is noted for the value $1-c$. The modal language $\mathscr{L}(Atm)$ of BCL is therefore defined as:

$$\varphi \quad ::= \quad p \mid \mathtt{t}(c) \mid \neg\varphi \mid \varphi \wedge \varphi \mid [X]\varphi,$$

where p ranges over Atm_0, $\mathtt{t}(c)$ ranges over Dec, and $X \subseteq Atm_0$.[1] Modal operator $\langle X \rangle$ is the dual of $[X]$ and is defined as usual: $\langle X \rangle \varphi =_{def} \neg[X]\neg\varphi$. Their meanings will be revealed after Definition 2.

The language $\mathscr{L}(Atm)$ is interpreted relative to classifier models defined as follows.

Definition 1 (Classifier model)**.** *A classifier model (CM) is a pair $C = (S, f)$ where:*

- $S \subseteq 2^{Atm_0}$ *is a set of states (or fact situations), and*
- $f : S \longrightarrow Val$ *is a decision (or classification) function.*

The class of classifier models is noted **CM**.

A pointed classifier model is a pair (C,s) with $C = (S,f)$ a classifier model and $s \in S$. Formulas in $\mathscr{L}(Atm)$ are interpreted relative to a pointed classifier model, as follows.

Definition 2 (Satisfaction relation)**.** *Let (C,s) be a pointed classifier model with $C = (S,f)$ and $s \in S$. Then:*

- $(C,s) \models p \Longleftrightarrow p \in s;$
- $(C,s) \models \mathtt{t}(c) \Longleftrightarrow f(s) = c;$
- *Standard valuation conditions for the Boolean connectives;*
- $(C,s) \models [X]\varphi \Longleftrightarrow \forall s' \in S:$ *if* $(s \cap X) = (s' \cap X)$ *then* $(C,s') \models \varphi.$

A formula φ of $\mathscr{L}(Atm)$ is said to be satisfiable relative to the class **CM** if there exists a pointed classifier model (C,s) with $C \in \mathbf{CM}$ such that $(C,s) \models \varphi$. It is said to be valid if $\neg\varphi$ is not satisfiable relative to **CM** and noted as $\models_{\mathbf{CM}} \varphi$.

We can think of a pointed model (C,s) as a pair (s,c) in f with $f(s) = c$. The formula $[X]\varphi$ is true at a state s if φ is true at all states that are modulo-X equivalent to state s. It has the *selectis paribus* (SP) (selected things being equal) interpretation "features in X being equal, necessarily φ holds (under possible perturbation on the other features)".[2] Notice when $X = \emptyset$, $[\emptyset]$ is the S5 universal modality since every state is modulo-$\emptyset$ equivalent to all states, viz. $(C,s) \models [\emptyset]\varphi \iff \forall s' \in S, (C,s') \models \varphi$.

[1] Notice p and $\mathtt{t}(c)$ have different statuses regarding negation: $\neg p$ means that the input variable p takes value 0, but $\neg\mathtt{t}(c)$ merely means the output does not take value c: we do not know which value it takes, since the output is trinary.

[2] $[Atm_0 \setminus X]\varphi$ has the standard *ceteris paribus* (CP) interpretation "features other than X being equal, necessarily φ holds (under possible perturbation of the features in X)".

3. From Classifiers Models to Case Bases

As we discussed in Section 1, classifiers should allow *incomplete-knowledge*: not all factual situations are associated with a decision in favour of the defendant or the plaintiff. Under this assumption, according to [2] two conditions must be guaranteed. First, a fact situation, besides 0 and 1, can also be classified as ?, where ? means absence of decision. That is, we consider classifier models whose classification function is of the form:

$$f : S \longrightarrow Val \text{ with } Val = \{0, 1, ?\}. \tag{1}$$

Secondly, we have to suppose that every possible situation is taken into account by the classifier. Namely,

$$S = 2^{Atm_0}. \tag{2}$$

Then, $\mathbf{CM}^{inc} = \{C \in \mathbf{CM} : C \text{ satisfies } (1), (2)\}$, is the class of possibly *incomplete-knowledge classifier models*. Satisfiability and validity wrt. $\mathbf{CM}^{inc}$ are defined as usual. An incomplete-knowledge classifier model $C \in \mathbf{CM}^{inc}$ models a *possibly incomplete-knowledge case base* defined as $CB_C = \{k \mid k = (s, f(s)) \text{ with } s \in S\}$.

Example 1. *Let* $C = (S, f)$, *with* $S = 2^{Atm_0}$, $Atm_0 = \{p_1, p_2, p_3\}$ *and* $f : 2^{Atm_0} \longrightarrow \{0, 1, ?\}$, *s.t.* $f(s) = 0$ *iff* $s \in \{\{p_1, p_2, p_3\}, \emptyset\}$, $f(s) = ?$ *iff* $s \in \{\{p_1, p_3\}, \{p_2, p_3\}\}$, $f(s) = 1$ *otherwise. The incomplete-knowledge case base is* $CB_C = \{k_i \mid k_i = (s_i, f(s_i)), i = 1, ..., 8\}$ *detailed in the following table.*

	s_i	$f(s_i)$
k_1 :	$\{p_1, p_2, p_3\}$	0
k_2 :	$\{p_1, p_2\}$	1
k_3 :	$\{p_1, p_3\}$	?
k_4 :	$\{p_2, p_3\}$	?
k_5 :	$\{p_1\}$	1
k_6 :	$\{p_2\}$	1
k_7 :	$\{p_3\}$	1
k_8 :	$\emptyset$	0

Observing the table, we note that p_1 *represents the only difference between the factual situation of case* k_5, *classified as* 1, *and that of case* k_8, *classified as* 0. *Intuitively we can state that* p_1 *is discriminating in favour of* 1 *and that it counterfactually explains the* 1*-decision settled for* k_5. *Furthermore, there is no pair of cases such that* p_1 *is discriminating in favour of* 0. *Hence, we claim that* p_1 *is a factor in the direction of* 1.

4. Factors, Irrelevant and Ambiguous Features

As we argued above, we can distinguish three notions: the notion of *factor* (unidirectional feature), *ambiguous* (multi-directional) feature and *irrelevant* (no-directional) feature. The distinction has been formally characterised in [2] using the idea of *strong counterfactual explanation* for binary classifiers [12].

We start by recalling the following notion of similarity between states.

Definition 3 (Similarity between states)**.** *Let $C = (S, f)$ be a classifier model, $s, s' \in S$. The degree of similarity between s and s' in S relative to the set of features Atm_0, noted $sim_C(s,s',Atm_0)$, is defined as follows:*

$$sim_C(s,s',Atm_0) = |\{p \in Atm_0 : (C,s) \models p \textit{ iff } (C,s') \models p\}|.$$

The dual notion of distance can be defined. In accordance with [14], it is nothing but the Hamming distance, which counts the cardinality difference of features' values.

The following definition introduces the notion of counterfactual in Lewis' style of the form $\varphi \Rightarrow \psi$ whose reading is "if it were φ, (wrt. features in Atm_0) it would be ψ."

Definition 4 (Counterfactual conditional)**.** *Let $C = (S, f)$ be a classifier model, $s \in S$. Then, $(C,s) \models \varphi \Rightarrow \psi$ if and only if $closest_C(s,\varphi,Atm_0) \subseteq ||\psi||_C$, where*

$$closest_C(s,\varphi,Atm_0) = \underset{s' \in ||\varphi||_C}{\arg\max}\, sim_C(s,s',Atm_0),$$

and for every $\varphi \in \mathscr{L}(Atm)$: $||\varphi||_C = \{s \in S : (C,s) \models \varphi\}$.

The idea is that $\varphi \Rightarrow \psi$ holds in a state of a classifier model iff all the closest (i.e., most similar) states to the current one, which make φ true, also make ψ true.[3]

A feature is said to be discriminating in one direction if just removing it from a case classified in that direction suffices for having the opposite classification. Then, the discriminating aspect of features is captured by the notion of strong counterfactual explanation for a decision, which we can express in the language $\mathscr{L}(Atm)$. Indeed, we will say that a formula φ of $\mathscr{L}(Atm)$ *strong* counterfactually explains a decision $c \in \{0,1\}$ for a certain situation s, if not satisfying φ would lead s to be classified as $\overline{c}$.

Definition 5 (Strong counterfactual explanation)**.** *We write $\mathsf{SCfXp}(\varphi,c)$ to mean that φ strong counterfactually explains a decision for $c \in \{0,1\}$ and define it as*

$$\mathsf{SCfXp}(\varphi,c) =_{def} \mathsf{t}(c) \wedge \big(\neg\varphi \Rightarrow \mathsf{t}(\overline{c})\big).$$

Given this notion, we can formally define the notions of (a) *factor* (a feature discriminating only in one direction), (b) *irrelevant feature* (a feature that does not explain any decision), (c) *ambiguous feature* (a feature explaining opposite decisions).[4]

Definition 6 (Factor, irrelevant feature, ambiguous feature [2])**.** *We write*

- $\mathsf{Factor}(p,c)$ *to mean that p is a factor in the direction of $c \in \{0,1\}$ such that*

$$\mathsf{Factor}(p,c) =_{def} \langle\emptyset\rangle\mathsf{SCfXp}(p,c) \wedge [\emptyset]\neg\mathsf{SCfXp}(p,\overline{c});$$

- $\mathsf{Irrelevant}(p)$ *to mean that p is irrelevant such that*

$$\mathsf{Irrelevant}(p) =_{def} \neg\langle\emptyset\rangle\mathsf{SCfXp}(p,1) \wedge \neg\langle\emptyset\rangle\mathsf{SCfXp}(p,0);$$

[3] Formula $\varphi \Rightarrow \psi$ captures the standard notion of conditional logic. One can show that $\Rightarrow$ satisfies all semantic conditions of Lewis' logic of counterfactuals VC [15].

[4] [2] has shown that, in this context, building similarity between cases via Hamming distance coincides with building it via subset inclusion relation (i.e., via shared properties as done in HYPO).

- $\mathsf{Amb}(p)$ *to mean that p is an ambiguous feature such that*

$$\mathsf{Amb}(p) =_{def} \langle\emptyset\rangle\mathsf{SCfXp}(p,1) \wedge \langle\emptyset\rangle\mathsf{SCfXp}(p,0).$$

A classifier model that does not admit ambiguous features is *consistent*.

Definition 7 (Consistency). *A classifier $C = (S, f)$ is* consistent given the current classifications *if and only if for every $s \in S$, $(C,s) \models \mathsf{Cons}$, where*

$$\mathsf{Cons} =_{def} \bigwedge_{p \in Atm_0} \neg\mathsf{Amb}(p).$$

Example 2. *Considering Example 1, we can verify that both p_1 and p_2 are factors in the direction of* 1 *, while p_3 is an ambiguous feature. Hence, C is not consistent.*

In [2] a variant of the a fortiori reasoning by Horty was introduced, taking also into account ambiguity and irrelevance. Intuitively, we expect that if the classifier associates a situation s to an outcome c, then it must assign the same outcome to every situation s' such that: (a) s' includes all factors in the direction of c that are in s (b) s' does *not include* factors in the direction of $\overline{c}$ that are *outside of* s and (c) s' *includes exactly the same ambiguous and irrelevant features that are in s*. In this sense, we can also say that *s supports* a decision as c for each s' as above. Looking at Example 1, we can say that this form of reasoning a fortiori fails and will create conflicting situations.

Example 3. *Consider Example 1. The factual situation in k_1 is classified as* 0 *and it contains more factors in the direction of* 1 *(namely p_2) than the situation in k_3. So, a fortiori, k_3 should be decided as* 0*. Namely, k_1 supports a decision in the direction of* 0 *for k_3. But, we can see that, since p_1 is a factor for* 1*, k_7 supports a decision for* 1 *for k_3.*

5. Inferring Classifications

One interesting research issue is how to infer new factors given a set of features variously qualified. Such inference mechanisms can be very useful, e.g., when (a) new decisions for previously undecided cases are provided or discovered or (b) robust explanatory models are needed to assess the outcome obtained through machine learning and predictive algorithms applied to judicial corpora. In these cases, it could for instance happen that: a) features previously labelled as irrelevant may become ambiguous or factors, b) features previously labelled as factors may become ambiguous in the "updating" of the case base, etc. If so, we identify *pro tanto* irrelevant features and *pro tanto* factors, i.e. features that are irrelevant or factors given the current information.

We don't require here consistency (Definition 7) and adopt a rather "liberal" approach, according to which we tolerate the presence of ambiguous features, and we provide a way of inferring the classification of cases for which there are precedents that support opposite decisions. Notice that more skeptical modes are possible, for example if one infers classifications for cases but stops the inference process for cases involved if some features turn out to be ambiguous.

In adopting this approach, our aim is to infer as much as possible new classifications, without stopping the inference. However, it may happen that not all potential inferences can be drawn in a consistent way, or that some features that were previously classified as factors are no longer so. In order to choose among more inferential options, we need to associate the classifiers with a total pre-order on the powerset of the set of states S. Conceptually, such a solution seems to us reasonable, because the order can be built using, e.g., the ranking of courts involved in the previous classifications, the importance of values promoted [16], and so forth.

Definition 8 (Ordered Classifier)**.** *An ordered classifier model is a triple $C_{\mathsf{or}} = (S, f, \preceq)$ where S is a set of states, f is a classification function and $\preceq$ is a total preorder on $\mathscr{P}(S)$.*

The class of ordered classifier models is noted $\mathbf{CM}^{\mathsf{or}}$. A pointed ordered classifier model is a pair (C_{or}, s) with $C_{\mathsf{or}} = (S, f, \preceq)$ an ordered classifier model and $s \in S$. Formulas in $\mathscr{L}(Atm)$ are interpreted relative to a pointed ordered classifier model as usual (see Definition 2). Satisfiability and validity wrt. $\mathbf{CM}^{\mathsf{or}}$ are defined in the usual way.

We now inductively define the inference mechanism of classifications for cases providing both a static and a dynamic version. In doing so, we retain the a fortiori principle and resort to the preorder only when more incompatible inferences are available.

5.1. Static Inference

Definition 9. *Let $k \geq 0$, $c \in \{0,1\}$. We write $\mathsf{Cl}^{\mathsf{k}}(c)$ to mean that a classification in the direction of c for the current factual situation can be inferred at step k, by preference over conflicting precedents and define it recursively as follows.*

1. $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{0}(c)$ *iff* $(C_{\mathsf{or}}, s) \models \mathsf{t}(c)$;

2 $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{\mathsf{k+1}}(c)$ *iff*
$(C_{\mathsf{or}}, s) \models \neg \mathsf{Cl}^{j \leq k}(\overline{c}) \wedge \bigvee_{T \subseteq 2^{Atm_0}} \mathsf{SupDec}^{k}(c, T) \wedge \neg \bigvee_{U \subseteq 2^{Atm_0}, T \preceq U} \mathsf{SupDec}^{k}(\overline{c}, U)$;

where

- $\mathsf{Cl}^{\mathsf{j} \leq \mathsf{k}}(c)$ *is an abbreviation for* $\bigvee_{j \leq k} \mathsf{Cl}^{\mathsf{j}}(c)$
- *for all* $T \subseteq 2^{Atm_0}$*,* $c \in \{0,1\}$
$(C_{\mathsf{or}}, s) \models \mathsf{SupDec}^{k}(c, T)$ *iff*
$T = \{s_1 \in S \mid (C_{\mathsf{or}}, s_1) \models \mathsf{Cl}^{\mathsf{j} \leq \mathsf{k}}(c), s_1 \setminus s \subseteq F^{k}_{\overline{c}}$ *and* $s \setminus s_1 \subseteq F^{k}_{c}\}$

and $F^{k}_{c} =_{\mathsf{def}} \{p \mid (C_{\mathsf{or}}, s') \models \mathsf{Factor}^{k}(p, c)$ *for all* $s' \in S\}$ *with*

- $\mathsf{Factor}^{k}(p, c) =_{def} \langle \emptyset \rangle \mathsf{SCfXp}^{k}(p, c) \wedge \neg \langle \emptyset \rangle \mathsf{SCfXp}^{k}(p, \overline{c})$ *and*
$\mathsf{SCfXp}^{k}(p, c) =_{def} \mathsf{Cl}^{\mathsf{j} \leq \mathsf{k}}(c) \wedge (\neg p \Rightarrow \mathsf{Cl}^{\mathsf{j} \leq \mathsf{k}}(\overline{c}))$.

Plainly speaking, we can explain the definition 'from the bottom up'. The inference process proceeds as follows. Suppose we have inferred the classifications up to step k ($\mathsf{Cl}^{\mathsf{j} \leq \mathsf{k}}(\cdot)$). Based on these, we can then extract on the basis of strong counterfactual explanation at step k ($\mathsf{SCfXp}^{k}(p, c)$) factors at step k ($\mathsf{Factor}^{k}(p, c)$) in the usual way. Then, we will say that a set T of situations supports a decision in the direction of c ($\mathsf{SupDec}^{k}(c, T)$) for a given situation s, if each situation of T forces a classification for c on the basis of inferred factors, as intuitively introduced at the end of the previous section (i.e. if it includes an equally or more inclusive set of factors for c, and no additional factor for $\overline{c}$ wrt. to s). Finally, we infer a classification as c at step $k+1$ ($\mathsf{Cl}^{\mathsf{k+1}}(c)$), for a considered situation s, if 1) s was not already inferred in the opposite direction; 2) there is a set T of states supporting a decision in the direction of c for s, and there is no set U of situations that support a decision in the direction $\overline{c}$, such that U is preferred to T wrt. the order $\preceq$.

Example 4. *Let us elaborate Example 1 to obtain an ordered case base. In particular, we require that for all $U \subseteq 2^{Atm_0}$, $U \preceq T$ with $T = \{\{p_1, p_2, p_3\}\}$. Recall that the set of factors in the direction of* 1 *at step* 0 *is $F_1^0 = \{p_1, p_2\}$. Note that we have for $s \in \{s_3, s_4\}$ that $(C_{or}, s) \models \neg \mathsf{Cl}^0(1) \wedge \mathsf{SupDec}^0(0, T)$. So, we can infer k_3 and k_4 as* 0*, since T is "stronger" than any other $U \subseteq 2^{Atm_0}$. We infer the classification of all other states again, with the exception of k_7. We cannot infer it as* 0 *because it has already been classified as* 1*; we cannot infer it as* 1 *because T supports a decision for k_3 in the direction of* 0.

	s_i	Cl^0	Cl^1
k_1 :	$\{p_1, p_2, p_3\}$	0	0
k_2 :	$\{p_1, p_2\}$	1	1
k_3 :	$\{p_1, p_3\}$	?	**0**
k_4 :	$\{p_2, p_3\}$	?	**0**
k_5 :	$\{p_1\}$	1	1
k_6 :	$\{p_2\}$	1	1
k_7 :	$\{p_3\}$	1	–
k_8 :	$\emptyset$	0	0

In the table "-" means that no classification as 0 *or* 1 *can be inferred.* [5] *This aspect deserves further attention. The fact that a classification for k_7 cannot be inferred hints that k_7 is involved in a form of ambiguity. Recall p_3 is an ambiguous feature, and is essentially so by virtue of counterfactual reasoning applied to cases k_1 and k_7. We also know that k_1 is stronger than any other case. Accordingly, we would say that k_7 should have been classified in the opposite direction in order to avoid the ambiguity of p_3. This intuition is reflected in the impossibility of inferring the classification of k_7 at step* 1.[6]

We highlight that we make it explicit in the definition that we can infer the classification in a direction c for a case if we have not already inferred for that case the classification in the opposite direction $\overline{c}$. Based on this, we obviously have the following.

Proposition 1. *Let $k \geq 1$. It holds the following validity*

$$\models \mathsf{Cl}^{\mathsf{k}}(c) \rightarrow \bigwedge_{j \geq k} \neg \mathsf{Cl}^{\mathsf{j}}(\overline{c}).$$

Since we cannot infer in a different direction an already inferred case, at most we can have 2^{Atm_0} iterations, inferring one new case per iteration. Namely, the inferential process we have defined gets stabilised:

Proposition 2. *It exists $k \leq 2^{|Atm_0|}$ s.t there is no $s \in 2^{Atm_0}$ s.t $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{\mathsf{k}}(c)$ and $(C_{\mathsf{or}}, s) \not\models \mathsf{Cl}^{\mathsf{j} \leq \mathsf{k}-1}(c)$.*

[5] Recall that, by definition 9, we can infer classifications as 0 or 1. We cannot infer them as ?, which means absence of decision. Instead, the symbol "-" indicates impossibility of inference.

[6] This intuition should make it possible to correct and revise the case base. But this is left for future work.

5.2. Dynamic Update

Let $C_{\mathsf{or}} = (S, f, \preceq)$ be an ordered classifier model. This time, we can update the model iteratively as follows, . Define

- $C^0_{\mathsf{or}} =_{\mathsf{def}} (S, h_0, \preceq)$ with $h_0 =_{\mathsf{def}} f$.
- For $k \geq 0$ update the model as $C^{k+1}_{\mathsf{or}} =_{\mathsf{def}} (S, h_{k+1}, \preceq)$, with

$$h_{k+1}(s) = \begin{cases} 0 & \text{if } (C^k_{\mathsf{or}}, s) \models \mathsf{Cl}^1(0) \\ 1 & \text{if } (C^k_{\mathsf{or}}, s) \models \mathsf{Cl}^1(1) \\ h_k(s) & \text{otherwise.} \end{cases}$$

Note that $\mathsf{Cl}^1(0)$ and $\mathsf{Cl}^1(1)$ are defined in Definition 9. So, we update the model each time with new classifications following the same reasoning applied at each step of static inference. In this sense, it is sufficient to perform "the first inference step", described in Definition 9.

The following proposition shows a form of equivalence between static inference and dynamic updating. More precisely, dynamic inference at a certain step 'cumulates' what has been inferred statically until that step.

Proposition 3. *For all $s \in S$, $k \geq 0$, $c \in \{0,1\}$, $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{j \leq k}(c)$ iff $(C^k_{\mathsf{or}}, s) \models \mathsf{t}(c)$*

Proof. Let $k = 0$. Then $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^0(c)$, $c \in \{0,1\}$ iff $(C, s) \models \mathsf{t}(c)$ iff $(C^0_{\mathsf{or}}, s) \models \mathsf{t}(c)$.
Suppose that for all $k \geq 1$, $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{j \leq k}(c)$, iff $(C^k_{\mathsf{or}}, s) \models \mathsf{t}(c)$.
Suppose now that $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{j \leq k+1}(c)$. So, we know that $(C_{\mathsf{or}}, s) \models \neg \mathsf{Cl}^{i \leq k}(\overline{c}) \wedge \bigvee_{T \subseteq 2^{Atm_0}} \mathsf{SupDec}^k(c, T) \wedge \neg \bigvee_{U \subseteq 2^{Atm_0}, T \preceq U} \mathsf{Supdec}^k(\overline{c}, U)$. By induction hypothesis we can verify that $(C^k_{\mathsf{or}}, s) \models \neg \mathsf{t}(\overline{c}) \wedge \bigvee_{T \subseteq 2^{Atm_0}} \mathsf{SupDec}^0(c, T) \wedge \neg \bigvee_{U \subseteq 2^{Atm_0}, T \preceq U} \mathsf{Supdec}^0(\overline{c}, U)$. Hence we have $(C^k_{\mathsf{or}}, s) \models \mathsf{Cl}^1(c)$. So $(C^{k+1}_{\mathsf{or}}, s) \models \mathsf{t}(c)$. To prove that $(C^{k+1}_{\mathsf{or}}, s) \models \mathsf{t}(c)$ implies $(C_{\mathsf{or}}, s) \models \mathsf{Cl}^{k+1}(c)$, proceed in "reverse order".

□

Example 5. *Consider C_{or} of Example 4. Recall that statically we infer a classification as 1 for k_7 at step 0 but not at step 1. Namely $(C_{or}, s_7) \models \mathsf{Cl}^0(1) \wedge \neg(\mathsf{Cl}^1(1) \vee \mathsf{Cl}^1(0))$. But then, by definition, $h_1(s_7) = 1$ and so, in dynamic updating at step 1, we have $(C^1_{\mathsf{or}}, s_7) \models \mathsf{t}(1)$ (i.e. s_7 is still classified as 1 at first step). This reflects the cumulativity nature of dynamic updating. Moreover, this shows that static inference, differently from dynamic updating, allows the cases 'causing ambiguity' (e.g. k_7 here) to be highlighted.*

6. Conclusion

Following the extensive AI & Law literature springing from the study of HYPO and CATO, in the last decade a significant effort has been put in investigating axioms as well as formal properties of factor-based case-based reasoning, and in providing the logical foundations for such a type of reasoning (see, among others, [8,17,18,19,20,21,6,1,11, 22]). Also due to development of explainable AI (XAI) [23,24], the quest for logical foundations of factor-based CBR has been recently focused, e.g., on formal models of argumentative explanation [21] or on logics for classifier systems [1].

As suggested in [11]—especially in the machine learning perspective—one aspect has remained in the background and needs a specific logical inquiry: the identification of factors, among the features within a case base, and the determination of their direction.

In [2] we proposed a novel approach to address this issue, starting from the intuition, introduced in [1], that a case base can be represented through a binary classifier.This enabled us to identify not only factors but also ambiguous and irrelevant features. In this paper, we have presented an inductive framework which allowed us to extend such a features analysis, to infer undecided cases in a conflict tolerant setting.

References

[1] Liu X, Lorini E, Rotolo A, Sartor G. Modelling and Explaining Legal Case-Based Reasoners Through Classifiers. In: JURIX 2022. IOS Press; 2022. p. 83-92.

[2] Di Florio C, Liu X, Lorini E, Rotolo A, Sartor G. Finding Factors in Legal Case-Based Reasoning. In: Logics for AI and Law 23. College Publications; 2023. p. 175-92.

[3] Ashley KD. Modeling Legal Argument: Reasoning with Cases and Hypotheticals. MIT; 1990.

[4] Rissland EL, Ashley KD. A Case-based System for Trade Secrets Law. In: ICAIL 1987. ACM; 1987. p. 60-6.

[5] Horty JF. The Result Model of Precedent. Legal Theory. 2004;10:19-31.

[6] Canavotto I. Precedential Constraint Derived from Inconsistent Case Bases. In: JURIX 2022. IOS Press; 2022. p. 23 -32.

[7] van Woerkom WK, Grossi D, Prakken H, Verheij B. Hierarchical Precedential Constraint. In: ICAIL '23. ACM; 2023. p. 333—342.

[8] Horty JF. Rules and reasons in the theory of precedent. Legal theory. 2011;17:1-33.

[9] Peters JGT, Bex FJ, Prakken H. Model- and data-agnostic justifications with A Fortiori Case-Based Argumentation. In: Proceedings of ICAIL 2023. ACM; 2023. p. 207-16.

[10] Odekerken D, Bex F, Prakken H. Justification, stability and relevance for case-based reasoning with incomplete focus cases. In: Proceedings of ICAIL 2023. ACM; 2023. p. 177-86.

[11] Bench-Capon TJ, Atkinson K. Precedential constraint: the role of issues. In: ICAIL'21. ACM; 2021. p. 12–21.

[12] Liu X, Lorini E. A logic for binary classifiers and their explanation. In: CLAR 2021. Springer; 2021. p. 302-21.

[13] Liu X, Lorini E. A unified logical framework for explanations in classifier systems. Journal of Logic and Computation. 2023;33(2):485-515.

[14] Dalal M. Investigations into a theory of knowledge base revision: preliminary report. In: Proceedings of the Seventh National Conference on Artificial Intelligence; 1988. p. 475—479.

[15] Lewis DK. Counterfactuals. Harvard University Press; 1973.

[16] Bench-Capon TJM, Sartor G. A model of legal reasoning with cases incorporating theories and values. Artif Intell. 2003;150(1-2):97-143.

[17] Prakken H. A formal analysis of some factor- and precedent-based accounts of precedential constraint. Artificial Intelligence and Law. 2021.

[18] Canavotto I, Horty J. Piecemeal Knowledge Acquisition for Computational Normative Reasoning. In: AIES'22. ACM; 2022. p. 171–180.

[19] Horty JF. Reasoning with dimensions and magnitudes. Artif Intell Law. 2019;27(3):309-45.

[20] Horty JF. Modifying the reason model. Artif Intell Law. 2021;29(2):271-85.

[21] Prakken H, Ratsma R. A top-level model of case-based argumentation for explanation: Formalisation and experiments. Argument Comput. 2022;13(2):159-94.

[22] Amgoud L, Beuselinck V. Towards a Principle-Based Approach for Case-Based Reasoning. In: Scalable Uncertainty Management. Cham: Springer International Publishing; 2022. p. 37-46.

[23] Miller T, Hoffman R, Amir O, Holzinger A, editors. Artificial Intelligence journal: Special issue on Explainable Artificial Intelligence (XAI). vol. 307; 2022.

[24] Atkinson K, Bench-Capon T, Bollegala D. Explanation in AI and law: Past, present and future. Artificial Intelligence. 2020;289:103387.