KLM-Style Defeasible Reasoning

Implementation and Evaluation of Entailment and Justification Algorithms

Chipo Hamayobe

chipo@cs.uct.ac.za

Artificial Intelligence Research Unit
Department of Computer Science
University of Cape Town

Abstract

In the 1950s, McCarthy proposed using formal logic in AI, establishing the foundations of **knowledge representation and reasoning**. Early systems, based on monotonic logic, struggled with exceptions since inferences could not be revised once drawn. To address this, researchers developed **non-monotonic** reasoning, enabling defeasible "common sense" inference. The **Kraus–Lehmann–Magidor (KLM) framework** emerged as a leading approach, with both syntactic and semantic formulations. Our work focuses on syntactic entailment algorithms—**Rational Closure**, **Lexicographic Closure**, and **Relevant Closure**—and develops justification methods to enhance interpretability. A web-based reasoning tool integrates computation, visualisation, and evaluation, advancing explainability and efficiency in symbolic reasoning.

1. Introduction and Background

Propositional Logic formalises reasoning by combining simple statements with defined operators, where truth depends only on base statements and operator use. It underpins formal reasoning by enabling analysis independent of natural language.

Valuation

A valuation $\boldsymbol{v}$ is a function that maps each atom in $\mathcal{P}$ to either `true` (T) or `false` (F).

Satisfiability

A valuation $\boldsymbol{v}$ satisfies an atom $\mathtt{p} \in \mathcal{P} \iff \boldsymbol{v}(\mathtt{p}) = T$, denoted by $\boldsymbol{v} \Vdash \mathtt{p}$.

Knowledge Base

A knowledge base, $\mathcal{K} \subset \mathcal{L}$, is a set of propositional formulas and is considered finite in this work.

Classical Entailment

Given a knowledge base $\mathcal{K}$ and a formula α, it is the case that α is entailed by $\mathcal{K}$, written $\mathcal{K} \models \alpha$ and read as "$\mathcal{K}$ entails α", $\iff$ for every $\boldsymbol{v} \in \mathcal{V}$ such that $\boldsymbol{v} \Vdash \mathcal{K}$ it is the case that $\boldsymbol{v} \Vdash \alpha$.

Propositional logic diverges from human reasoning: it cannot represent **typicality** with exceptions and, due to its **monotonicity**, prevents retracting conclusions when new, conflicting information arises.

2. Defeasible Reasoning and Explanation

To address the **monotonicity** limitation, the **KLM framework** formalises **non-monotonic** reasoning about typicality through a **preferential consequence relation**, supporting conclusions like *"an A is typically a B"*. Based on **KLM postulates**, relations satisfying postulates 1–7 are **preferential**, and those also meeting the 8th are **rational**. Interpreted via **preferential** or **ranked** semantics, valuations are ordered by "typicality", and a defeasible implication $\alpha \mathrel{|\!\sim} \beta$ holds when all preferred worlds satisfying α also satisfy β.

KLM Postulates

1. Reflexivity: $\dfrac{}{\alpha \to \alpha}$

2. Left Logical Equivalence: $\dfrac{\alpha \equiv \beta, \alpha \mathrel{|\!\sim} \gamma}{\beta \mathrel{|\!\sim} \gamma}$

3. Right Weakening: $\dfrac{\alpha \to \beta, \gamma \mathrel{|\!\sim} \alpha}{\gamma \mathrel{|\!\sim} \beta}$

4. And: $\dfrac{\alpha \mathrel{|\!\sim} \beta, \alpha \mathrel{|\!\sim} \gamma}{\alpha \mathrel{|\!\sim} \beta \wedge \gamma}$

5. Or: $\dfrac{\alpha \mathrel{|\!\sim} \beta, \gamma \mathrel{|\!\sim} \beta}{\alpha \vee \gamma \mathrel{|\!\sim} \beta}$

6. Cut: $\dfrac{\alpha \wedge \gamma \mathrel{|\!\sim} \beta, \alpha \mathrel{|\!\sim} \gamma}{\alpha \mathrel{|\!\sim} \beta}$

7. Cautious Monotonicity: $\dfrac{\alpha \mathrel{|\!\sim} \gamma, \alpha \mathrel{|\!\sim} \beta}{\alpha \wedge \beta \mathrel{|\!\sim} \gamma}$

8. Rational Monotonicity: $\dfrac{\alpha \mathrel{|\!\sim} \beta, \alpha \not\mathrel{|\!\sim} \neg\gamma}{\alpha \wedge \gamma \mathrel{|\!\sim} \beta}$

Exceptionality

Given a knowledge base $\mathcal{K}$, a propositional formula $\alpha \in \mathcal{L}$ is considered exceptional for $\mathcal{K}$ if and only if $\mathcal{K} \mathrel{|\!\approx}_P \top \mathrel{|\!\sim} \neg\alpha$.

Minimal Ranked Entailment

Given a defeasible knowledge base $\mathcal{K}$, the minimal ranked interpretation satisfying $\mathcal{K}$, $\mathcal{R}^{\mathcal{K}}_{RC}$, defines an entailment relation $\mathrel{|\!\approx}$, known as the minimal ranked entailment, such that for any defeasible implication $\alpha \mathrel{|\!\sim} \beta$, $\mathcal{K} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta$ if and only if $\mathcal{R}^{\mathcal{K}}_{RC} \Vdash \alpha \mathrel{|\!\sim} \beta$.

Justification of Entailments

Given a knowledge base $\mathcal{K}$ and some classical propositional formula $\alpha \to \beta$ such that $\mathcal{K} \models \alpha \to \beta$. Then the set $\mathcal{J}$ is said to be a justification for $\alpha \to \beta$ in $\mathcal{K}$ if $\mathcal{J} \subseteq \mathcal{K}, \mathcal{J} \models \alpha \to \beta$ and for every $\mathcal{J}' \subset \mathcal{J}, \mathcal{J}' \not\models \alpha \to \beta$. The set $\mathcal{J}$ will sometimes be denoted as $\mathcal{J}^{\mathcal{K}}_{(\alpha \to \beta)}$ or $\mathcal{J}(\alpha \to \beta, \mathcal{K})$.

2.1. Base Rank

The exceptionality sequence is an iterative sequence of knowledge bases, $\mathcal{E}^{\mathcal{K}}_0, \mathcal{E}^{\mathcal{K}}_1 \ldots \mathcal{E}^{\mathcal{K}}_n$ where $\mathcal{E}^{\mathcal{K}}_0 = \mathcal{K}$ and $\mathcal{E}^{\mathcal{K}}_{i+1} = \varepsilon(\mathcal{E}^{\mathcal{K}}_i)$ for $1 \leq i \leq n$. Since $\mathcal{K}$ is finite, some n must always exist, and with the sequence $\mathcal{E}^{\mathcal{K}}_0, \mathcal{E}^{\mathcal{K}}_1, \ldots, \mathcal{E}^{\mathcal{K}}_\infty$ defined, we can then formally define the *base rank* of a formula.

Base Rank

With respect to a given defeasible knowledge base $\mathcal{K}$, $br_{\mathcal{K}}(\alpha)$, known as the base rank of a formula $\alpha \in \mathcal{L}$, is the smallest index r such that α is not exceptional in $\mathcal{E}^{\mathcal{K}}_r$, and is defined: $br_{\mathcal{K}}(\alpha) := min\{r \mid \mathcal{E}^{\mathcal{K}}_r \not\mathrel{|\!\approx}_{\mathcal{R}} \top \mathrel{|\!\sim} \neg\alpha\}$.

Given the following knowledge base and the query $\mathtt{pets} \mathrel{|\!\sim} \mathtt{legs}$:

$$\mathcal{K} = \{\, \mathtt{animals} \mathrel{|\!\sim} \mathtt{legs},\ \mathtt{animals} \mathrel{|\!\sim} \mathtt{wild},\ \mathtt{pets} \to \mathtt{animals},\ \mathtt{pets} \mathrel{|\!\sim} \neg\mathtt{wild} \,\}$$

$\mathcal{R}_\infty$	$\mathtt{pets} \to \mathtt{animals}$
$\mathcal{R}_1$	$\mathtt{pets} \mathrel{|\!\sim} \neg\mathtt{wild}$
$\mathcal{R}_0$	$\mathtt{animals} \mathrel{|\!\sim} \mathtt{legs},\ \ \mathtt{animals} \mathrel{|\!\sim} \mathtt{wild}$

2.2. Rational Closure

Rational closure applies prototypical reasoning, where typical instances inherit default properties, but atypical cases, such as `pets` are `animals`, do not.

Rational Closure

Given a knowledge base $\mathcal{K}$, a defeasible implication $\alpha \mathrel{|\!\sim} \beta$ is said to be in the rational closure of $\mathcal{K}$ (written $\mathcal{K} \mathrel{|\!\approx}_{RC} \alpha \mathrel{|\!\sim} \beta$), if and only if $br_{\mathcal{K}}(\alpha) < br_{\mathcal{K}}(\alpha \wedge \neg\beta)$ or $br_{\mathcal{K}}(\alpha) = \infty$.

- *Discarded ranks*: $\mathcal{R}_0 = \{\mathtt{animals} \mathrel{|\!\sim} \mathtt{legs}, \mathtt{animals} \mathrel{|\!\sim} \mathtt{wild}\}$
- *Entailment*: $\mathcal{K} \not\mathrel{|\!\approx}_{RC} \mathtt{pets} \mathrel{|\!\sim} \mathtt{legs}$

2.3. Lexicographic Closure

Lexicographic closure, a more adventurous form of defeasible entailment, permits atypical cases like platypuses to inherit most typical mammalian properties, unlike rational closure.

Lexicographic Closure

Given a knowledge base $\mathcal{K}$, and a defeasible implication $\alpha \mathrel{|\!\sim} \beta$, it is the case that $\alpha \mathrel{|\!\sim} \beta$ is in the *Lexicographic Closure* of $\mathcal{K}$, denoted $\mathcal{K} \mathrel{|\!\approx}_{LC} \alpha \mathrel{|\!\sim} \beta$, if and only if for any basis $\mathcal{K}^{\mathcal{B}} \subseteq \mathcal{K}$ of α, $\overrightarrow{\mathcal{K}^{\mathcal{B}}} \cup \{\alpha\} \models \beta$.

- *Discarded statements*: $\{\mathtt{animals} \mathrel{|\!\sim} \mathtt{wild}\}$
- *Entailment*: $\mathcal{K} \mathrel{|\!\approx}_{LC} \mathtt{pets} \mathrel{|\!\sim} \mathtt{legs}$
- *Deciding*: $\mathcal{D} = \{\mathtt{pets} \to \mathtt{animals}, \mathtt{pets} \mathrel{|\!\sim} \neg\mathtt{wild}, \mathtt{animals} \mathrel{|\!\sim} \mathtt{legs}\}$
- *Justifications*: $\mathcal{J}_1 = \{\mathtt{pets} \to \mathtt{animals}, \mathtt{animals} \mathrel{|\!\sim} \mathtt{legs}\}$

2.4. Relevant Closure

To define how to compute the **relevant** partition $\langle \mathcal{R}, \mathcal{R}^- \rangle$ for a query $\alpha \mathrel{|\!\sim} \beta$: the antecedent α determines the **relevant** partition. Based on the reasoning behind **Relevant Closure**, $\mathcal{R}$ should include exactly those statements used to derive $\neg\alpha$.

Basic Relevant Closure

For a statement α and knowledge base $\mathcal{K}$, let $\mathcal{J}^{\mathcal{K}}_{bas}(\alpha) = \{\mathcal{J} \mid \mathcal{J}$ is an ε-justification w.r.t. $\mathcal{K}\}$. Then $\alpha \mathrel{|\!\sim} \beta$ is said to be in the Basic Relevant Closure of $\mathcal{K}$ if it is in the Relevant Closure of $\mathcal{K}$ w.r.t. $\bigcup \mathcal{J}_{bas} \subseteq \mathcal{K}$.

Minimal Relevant Closure

For some set of justifications $\mathcal{J} \subseteq \mathcal{K}$, let $\mathcal{J}^{\mathcal{K}}_{min} = \{\alpha \mathrel{|\!\sim} \beta \mid r_{\mathcal{K}}(\alpha) \leq r_{\mathcal{K}}(\gamma)$ for every $\gamma \mathrel{|\!\sim} \sigma \in \mathcal{J}$. For a statement α, let $\mathcal{J}^{\mathcal{K}}_{min}(\alpha) = \bigcup_{\mathcal{J} \in \mathcal{J}^{\mathcal{K}}\alpha} \mathcal{J}^{\mathcal{K}}_{min}(\alpha)$. Then $\alpha \mathrel{|\!\sim} \beta$ is said to be in the Minimal Relevant Closure of $\mathcal{K}$ if it is in the Relevant Closure of $\mathcal{K}$ w.r.t. $\bigcup \mathcal{J}^{\mathcal{K}}_{min}(\alpha)$.

- *Relevance*: $\mathcal{R}^{bas} = \{\mathtt{animals} \mathrel{|\!\sim} \mathtt{wild}, \mathtt{pets} \mathrel{|\!\sim} \neg\mathtt{wild}\}$, $\mathcal{R}^{min} = \{\mathtt{animals} \mathrel{|\!\sim} \mathtt{wild}\}$
- *Discarded statements*: $\{\mathtt{animals} \mathrel{|\!\sim} \mathtt{wild}\}$
- *Entailment*: $\mathcal{K} \mathrel{|\!\approx}_{LC} \mathtt{pets} \mathrel{|\!\sim} \mathtt{legs}$
- *Deciding*: $\mathcal{D} = \{\mathtt{pets} \to \mathtt{animals}, \mathtt{pets} \mathrel{|\!\sim} \neg\mathtt{wild}, \mathtt{animals} \mathrel{|\!\sim} \mathtt{legs}\}$
- *Justifications*: $\mathcal{J}_1 = \{\mathtt{pets} \to \mathtt{animals}, \mathtt{animals} \mathrel{|\!\sim} \mathtt{legs}\}$

3. Implementation of Algorithms

The ***klm-algorithms*** system employs a *modular, layered architecture* that promotes clarity, scalability, and maintainability. It features a web-based interface linked to a *backend* responsible for logical processing and algorithm execution, integrating external libraries to enhance reasoning and explanation capabilities.

By maintaining a clear separation between the interface, controllers, and domain logic, the system efficiently processes user input, executes defeasible reasoning algorithms, and generates coherent explanatory feedback.

klm-algorithms is available at: https://klm-algorithms.fly.dev

An example of the implemented algorithms is *DefeasibleJustification*, which computes minimal justification sets for a given entailment **deciding knowledge base** $\mathcal{D}$ for a defeasible query $\alpha \mathrel{|\!\sim} \beta$:

Algorithm .1: `UniversalDefeasibleJustification`

Input: A defeasible knowledge base $\mathcal{K}$ and a defeasible query $\alpha \mathrel{|\!\sim} \beta$
Output: The set of all justification sets $\{\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3, \ldots\}$

```
(holds, D) := KLMDefeasibleEntailment(K, α |~ β);
if holds = false then
    return ∅;
end
return ComputeAllJustifications(D, α |~ β);
```

4. Evaluation of Algorithms

This study evaluates the logical soundness and computational feasibility of the algorithms to clarify their role in defeasible reasoning, particularly amid their growing relevance in artificial intelligence. The analysis focuses on key performance metrics such as efficiency and entailment accuracy, which determine practical applicability.

We evaluate algorithm performance on knowledge bases that vary in statement count, rank depth, and distribution pattern. While comparative studies indicate that these algorithms often mirror intuitive reasoning, they continue to face scalability challenges in large knowledge bases, underscoring the need for future work on interpretable and scalable solutions.