Paper Summary

<!--META_START-->

Title: CARE: Coherent Actionable Recourse based on Sound Counterfactual Explanations

Authors: Peyman Rasouli, Ingrid Chieh Yu

DOI: https://doi.org/10.1145/nnnnnnn.nnnnnnn

Year: 2021

Publication Type: Conference

Discipline/Domain: Computer Science / Artificial Intelligence

Subdomain/Topic: Interpretable Machine Learning, Counterfactual Explanations, Actionable Recourse

Eligibility: Eligible

Overall Relevance Score: 95

Operationalization Score: 95

Contains Definition of Actionability: Yes

Contains Systematic Features/Dimensions: Yes

Contains Explainability: Yes

Contains Interpretability: Yes

Contains Framework/Model: Yes

Operationalization Present: Yes

Primary Methodology: Conceptual with empirical evaluation

Study Context: Model-agnostic counterfactual and recourse generation for classification and regression c

Geographic/Institutional Context: University of Oslo, Norway

Target Users/Stakeholders: End-users seeking actionable guidance from ML predictions; researchers in e

Primary Contribution Type: Modular explanation framework (CARE) integrating model-level and user-leve

CL: Yes

CR: Yes

FE: Yes

TI: No

EX: Partial

GA: Yes

Reason if Not Eligible: n/a

<!--META_END-->

Title: CARE: Coherent Actionable Recourse based on Sound Counterfactual Explanations

Authors: Peyman Rasouli, Ingrid Chieh Yu

DOI: https://doi.org/10.1145/nnnnnnn.nnnnnnn

Year: 2021

Publication Type: Conference

Discipline/Domain: Computer Science / Artificial Intelligence

Subdomain/Topic: Interpretable Machine Learning, Counterfactual Explanations, Actionable Recourse

Contextual Background: The paper addresses the limitations of existing counterfactual explanation m

Geographic/Institutional Context: University of Oslo, Norway

Target Users/Stakeholders: ML end-users needing recourse (e.g., loan applicants), explainable AI res

Primary Methodology: Conceptual with empirical evaluation

Primary Contribution Type: New modular framework for counterfactual and recourse generation

General Summary of the Paper

The authors propose CARE, a modular, model-agnostic explanation framework for generating actionable

Eligibility

Eligible for inclusion: **Yes**

How Actionability is Understood

Actionability is defined as satisfying global and local user/domain-specific preferences through constraint

> “A counterfactual should satisfy some global and local preferences that are domain-specific and define

> “An actionable explanation… takes into account the user’s preferences containing the name of mutable

What Makes Something Actionable

- Alignment with user-specified constraints (mutable/immutable features, allowed ranges/values)

- Preservation of feature coherency under constraints

- Feasibility in real-world terms (not recommending impossible changes)

- Respecting constraint importance (prioritizing non-violable constraints)

How Actionability is Achieved / Operationalized

- **Framework/Approach Name(s):** CARE

- **Methods/Levers:** Modular hierarchy with four modules; multi-objective optimization using NSGA-III

- **Operational Steps / Workflow:**

  1. **VALIDITY:** Enforce minimal, sparse changes to achieve the desired outcome.

  2. **SOUNDNESS:** Ensure proximity and connectedness to real, same-class data points.

  3. **COHERENCY:** Use correlation models to preserve feature relationships.

  4. **ACTIONABILITY:** Apply user-defined constraints with importance weighting.

- **Data & Measures:** Gower distance, Local Outlier Factor, HDBSCAN clustering, correlation measures

- **Implementation Context:** Model-agnostic; applicable to tabular classification/regression; handles mix

> “We propose a constraint language… and the notion of constraint importance to weigh the constraints a

> “CARE… generates actionable recourse by fulfilling the mentioned desiderata through objective functio

Dimensions and Attributes of Actionability (Authors’ Perspective)

- **CL (Clarity):** Yes — minimal, interpretable feature changes improve understandability (p. 3).

- **CR (Contextual Relevance):** Yes — proximity and connectedness ensure alignment with domain dat

- **FE (Feasibility):** Yes — coherent changes preserve real-world plausibility (p. 2–3).

- **TI (Timeliness):** No — not explicitly addressed.

- **EX (Explainability):** Partial — explanations are inherent but focus is on actionable counterfactuals, n

- **GA (Goal Alignment):** Yes — constraints ensure user goals/preferences are respected (p. 6).

- **Other Dimensions Named by Authors:** Coherency, proximity, connectedness.

Theoretical or Conceptual Foundations

- Counterfactual explanations in XAI (Wachter et al., 2017)

- Proximity and connectedness metrics (Laugel et al., 2019)

- Actionable recourse frameworks (Ustun et al., 2019; Karimi et al., 2020)

- Multi-objective optimization (NSGA-III)

Indicators or Metrics for Actionability

- Actionability cost (sum of violated constraint importance values)

- Proximity and connectedness scores to assess plausibility

- Coherency rate (preservation of feature correlations)

Barriers and Enablers to Actionability

- **Barriers:** Conflicting constraints; lack of coherent feature changes; artifacts in model space (p. 2–3).

- **Enablers:** Modular structure allowing selective enforcement of properties; weighting of constraints by

Relation to Existing Literature

The paper extends prior counterfactual explanation methods by integrating seldom-addressed properties

Summary

CARE is a modular, model-agnostic framework for generating actionable recourse grounded in sound co

Scores

- **Overall Relevance Score:** 95 — Provides explicit and nuanced definition of actionability with multiple

- **Operationalization Score:** 95 — Fully details how to implement actionability in practice through const

Supporting Quotes from the Paper

- “A counterfactual should satisfy some global and local preferences that are domain-specific and defined

- “We introduce a novel notion of actionability that can cover various constraints and prioritize different pr

- “Our proposed objective function… computes the actionability cost… according to the user’s preference

- “An actionable explanation… takes into account the user’s preferences containing the name of mutable/

Actionability References to Other Papers

- Ustun, Spangher, Liu (2019) — Actionable recourse in linear classification

- Karimi et al. (2020) — Algorithmic recourse

- Wachter et al. (2017) — Counterfactual explanations

- Laugel et al. (2019) — Proximity and connectedness in counterfactuals

- Dandl et al. (2020) — Multi-objective counterfactual explanations