Why this and not that?
A Logic-based Framework for Contrastive Explanations

Tobias Geibinger, **Reijo Jaakkola**, Antti Kuusisto,
Xinghan Liu and Miikka Vilander

Tampere University & TU Wien

September 3, 2025

Overview

Table of Contents

Contrastive Explanation

- Answers a question of the form "Why P and not Q?"
  - Why is this Iris flower a Setosa and not a Versicolor?
  - Why did Bob get a loan and I did not?
  - Why was this animal classified as a dog and not a cat?
  - ⋮
- Miller (2019) argues that people tend to prefer a contrastive explanation even to a **why**-question.

This work

- We formalize several notions of contrastive explanation as computational problems.
- In our definitions we assume a background logic $\mathcal{L}$.

Table of Contents

O-contrast

- **Question:** given two objects s, s' and properties φ, ψ, we ask

$$\text{Why } s \models \varphi \wedge \neg\psi \text{ but } \quad s' \models \neg\varphi \wedge \psi \text{ ?}$$

  In Van Bouwel and Weber (2002) this question is called an O-contrast.

- **Example:** Why is flower s a Setosa while flower s' is a Versicolor?

Contrastive Explanation Problem

- As an **answer** to O-contrast we propose a tuple

$$\theta, \theta', \chi$$

of formulas of $\mathcal{L}$ such that

$$s \models \theta \wedge \chi \models \varphi \wedge \neg\psi$$

and

$$s' \models \theta' \wedge \chi \models \neg\varphi \wedge \psi$$

- This becomes an **optimization problem** if we require θ, θ' and χ to be as short as possible.
- **Example:** Flower s is a Setosa because its petals width and length are < 2.45cm and < 1.75cm respectively while flower s' is a Versicolor because its petal width is ≥ 2.45cm and petal length is < 1.75cm.

Table of Contents

P-contrast

- **Question:** given an object s and properties φ, ψ, we ask

$$\text{Why } s \models \varphi \wedge \neg\psi \text{ ?}$$

  In Van Bouwel and Weber (2002) this question is called a P-contrast.
- **Example:** Why is the Iris flower s a Setosa and not a Versicolor?

Counterfactual Contrastive Explanation Problems

- Again, we propose a minimal tuple

$$\theta, \theta', \chi$$

of formulas of $\mathcal{L}$ as an answer.

- There are two natural choices as to what require from them.

First option:

$$s \models \theta \wedge \chi \models \varphi \wedge \neg\psi$$

and

$$\theta' \wedge \chi \models \neg\varphi \wedge \psi$$

Counterfactual Contrastive Explanation Problems

- Again, we propose a minimal tuple

$$\theta, \theta', \chi$$

of formulas of $\mathcal{L}$ as an answer.

- There are two natural choices as to what require from them.

First option:

$$s \models \theta \wedge \chi \models \varphi \wedge \neg\psi$$

and

$$\theta' \wedge \chi \models \neg\varphi \wedge \psi$$

Counterfactual Contrastive Explanation Problems

- Again, we propose a minimal tuple

$$\theta, \theta', \chi$$

of formulas of $\mathcal{L}$ as an answer.

- There are two natural choices as to what require from them.

Second option:

$$s \equiv \theta \wedge \chi \models \varphi \wedge \neg\psi$$

and

$$\theta' \wedge \chi \models \neg\varphi \wedge \psi$$

- This option is analogous with the notion of CXp from Ignatiev et al. (2020).

Table of Contents

Global contrast

- **Question**: given properties φ, ψ, we ask

  What is the difference between φ and ψ ?

- **Example:** What is the difference between a Setosa and a Versicolor?

Global Contrastive Explanation Problem

- In analogue with earlier definitions, a possible answer is a minimal triple θ, θ', χ such that

$$\theta \wedge \chi \equiv \varphi \wedge \neg\psi$$

and

$$\theta' \wedge \chi \equiv \neg\varphi \wedge \psi.$$

- Intuitively, θ, θ' capture all the differences between φ and ψ, while χ covers their common properties.

Minimal Separator Problem

- In this variant we are asking for the shortest **interpolant** θ such that

$$\varphi \models \theta \models \neg\psi$$

- **Example:** Every Setosa has petal width < 2.45cm and every Versicolor has petal width ≥ 2.45cm.

Table of Contents

Conclusions

- We introduced a logic-based framework for contrastive explanations.
- Formalized different central types of contrastive questions (O-contrasts, P-contrasts and Global contrasts) as computational problems.
- In the paper we also study their computational complexity and also provide implementations for them using Answer Set Programming.

Thanks!

References

Ignatiev, A., Narodytska, N., Asher, N., and Marques-Silva, J. (2020). From contrastive to abductive explanations and back again. In Baldoni, M. and Bandini, S., editors, *AIxIA 2020 - Advances in Artificial Intelligence - XIXth International Conference of the Italian Association for Artificial Intelligence, Virtual Event, November 25-27, 2020, Revised Selected Papers*, volume 12414 of *Lecture Notes in Computer Science*, pages 335–355. Springer.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38.

Van Bouwel, J. and Weber, E. (2002). Remote causes, bad explanations? *Journal for the Theory of Social Behaviour*, 32(4):437 – 449.