LEMON: Explainable Entity Matching

Nils Barlaug
Cognite & Department of Computer Science, NTNU

COGNITE

NorwAI
Norwegian Research Center for AI Innovation

Entity Matching

The task of deciding which records refer to the same real-world entities.

title	belkin shield micra for ipod touch tint	**title**	belkin ipod touch shield micra tint-royal purple
category	mp3 accessories	**category**	cases
brand	belkin	**brand**	belkin
modelno	f8z646ttc01	**modelno**	f8z646ttc02
price	47.88	**price**	12.49

Table 1: Example of two records that need to be classified as either a match or a non-match. In this case, from the Walmart-Amazon dataset.

LEMON

State-of-the-art entity matching models [2, 3] are hard to interpret, and popular explainability methods [5, 4, 6] do not work satisfactorily out of the box for this problem. Therefore, we propose LEMON*:

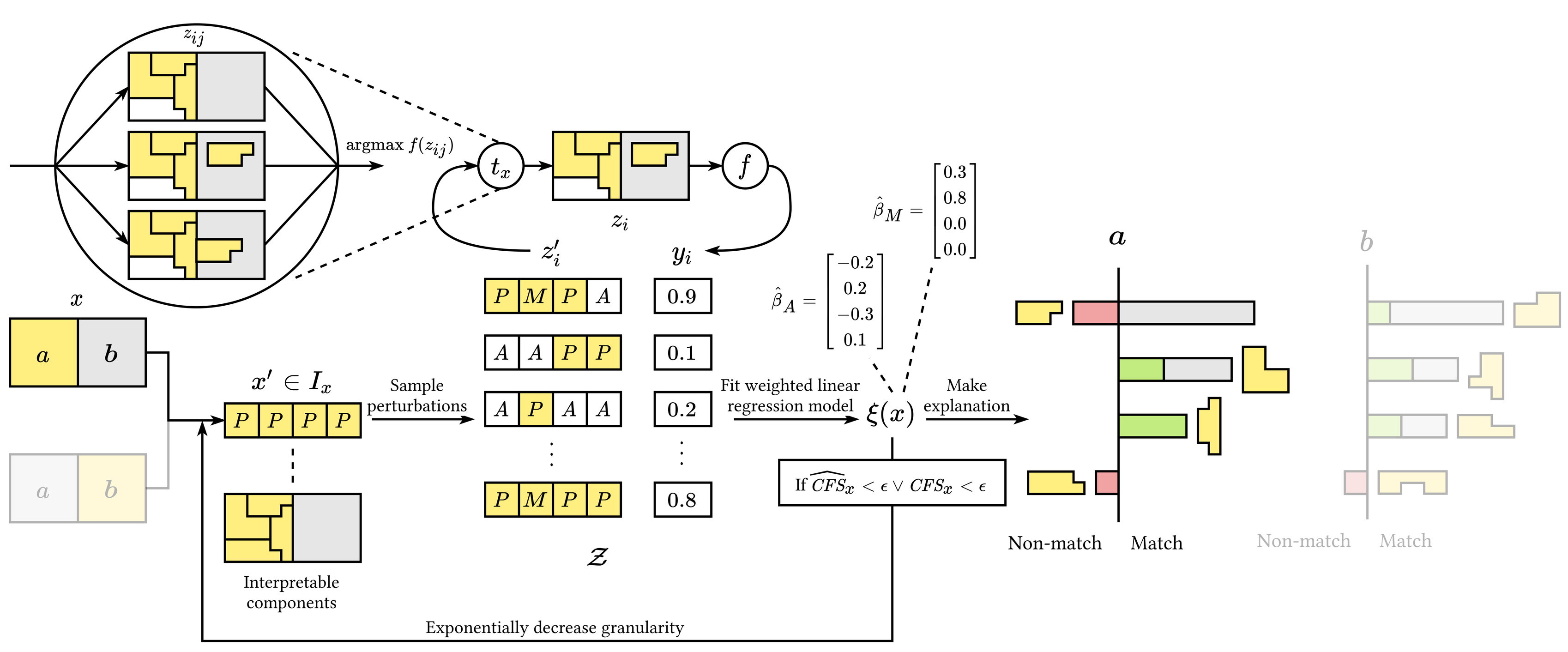

Figure 1: Overview of LEMON. The method is based on LIME [5]. Please ask me to explain or check the paper* for details.

LEMON addresses three different challenges of applying feature attribution methods to entity matching using three distinct, but coherent, techniques:

Challenges:
1. Cross-record interactions
2. Non-match explanations
3. Variation in sensitivity

Proposed techniques:
1. Dual explanations
2. Attribution potential
3. Counterfactual granularity

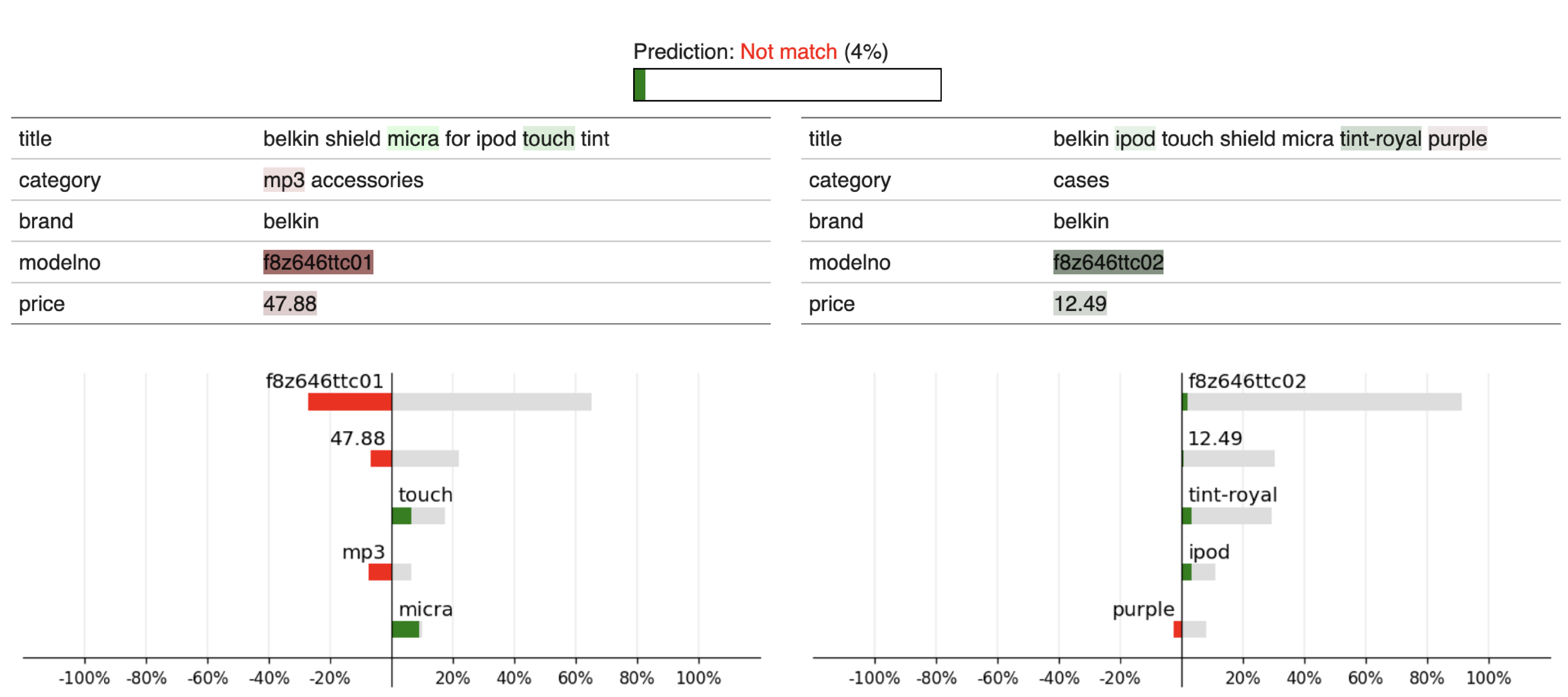

Figure 2: Example of an explanation from LEMON. In this case for a Bert-based matcher on the example from Table 1.

*Code: `https://github.com/NilsBarlaug/lemon`.
*Paper: `https://arxiv.org/abs/2110.00516`.

[1] Andrea Baraldi et al. "Using Landmarks for Explaining Entity Matching Models". In: *EDBT*. OpenProceedings.org, 2021.
[2] Pradap Konda et al. "Magellan: Toward Building Entity Matching Management Systems". In: *VLDB* 9.12 (2016).
[3] Yuliang Li et al. "Deep Entity Matching with Pre-Trained Language Models". In: *VLDB* 14.1 (2020).
[4] Scott M Lundberg et al. "A Unified Approach to Interpreting Model Predictions". In: *NIPS*. Ed. by I. Guyon et al. 2017.
[5] Marco Tulio Ribeiro et al. ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *SIGKDD*. 2016.
[6] Mukund Sundararajan et al. "Axiomatic Attribution for Deep Networks". In: *ICML*. 2017.

Results

We compare against state-of-the-art for explainable entity matching [1] and popular general-purpose explainability methods [5, 4, 6] on 13 popular datasets using both a state-of-the-art classical model (results not included in poster) and a deep learning model.

Counterfactual Interpretation

- To what degree explanations communicate how the matchers decision could have been different.

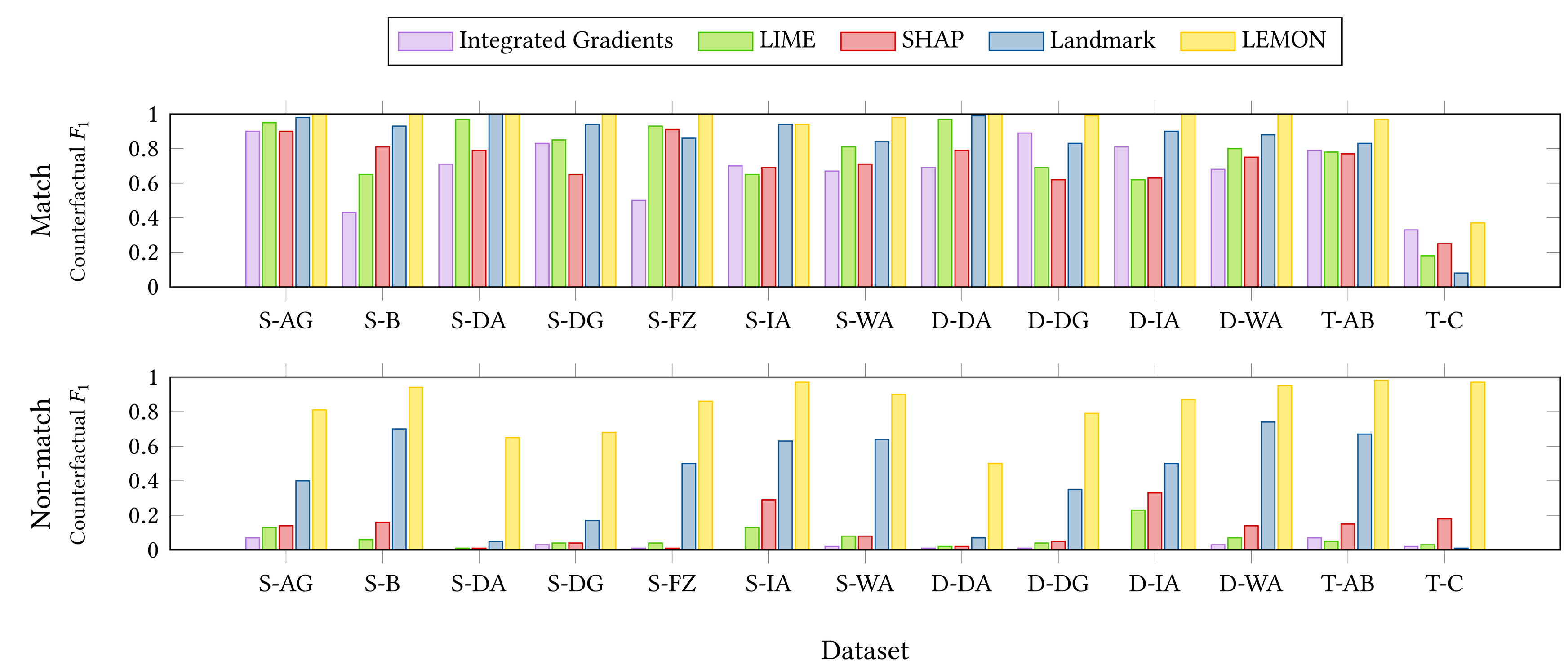

Figure 3: The counterfactual F_1 score for all evaluated explainability methods across 13 datasets using a Bert-based matcher. Higher is better.

Explanation Faithfulness

- To what degree explanations (do not) reflect the actual behavior of the matcher.

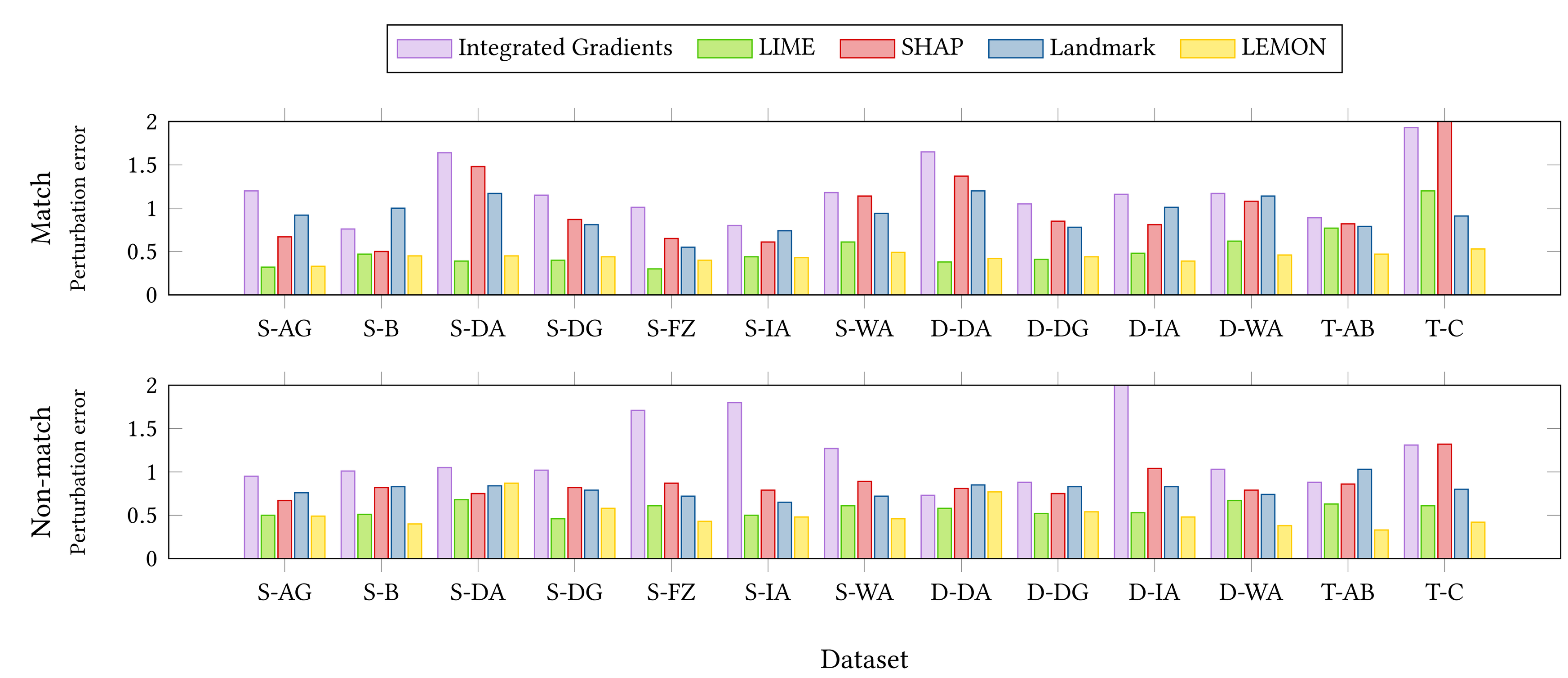

Figure 4: The perturbation error for all evaluated explainability methods across 13 datasets using a Bert-based matcher. Lower is better.

User Study

- To what degree explanations help (laymen) users successfully manipulate a record pair to flip the matchers decision. After showing the users an explanation, we ask them "What do you think is the smallest change to the records that would convince the matcher otherwise?"

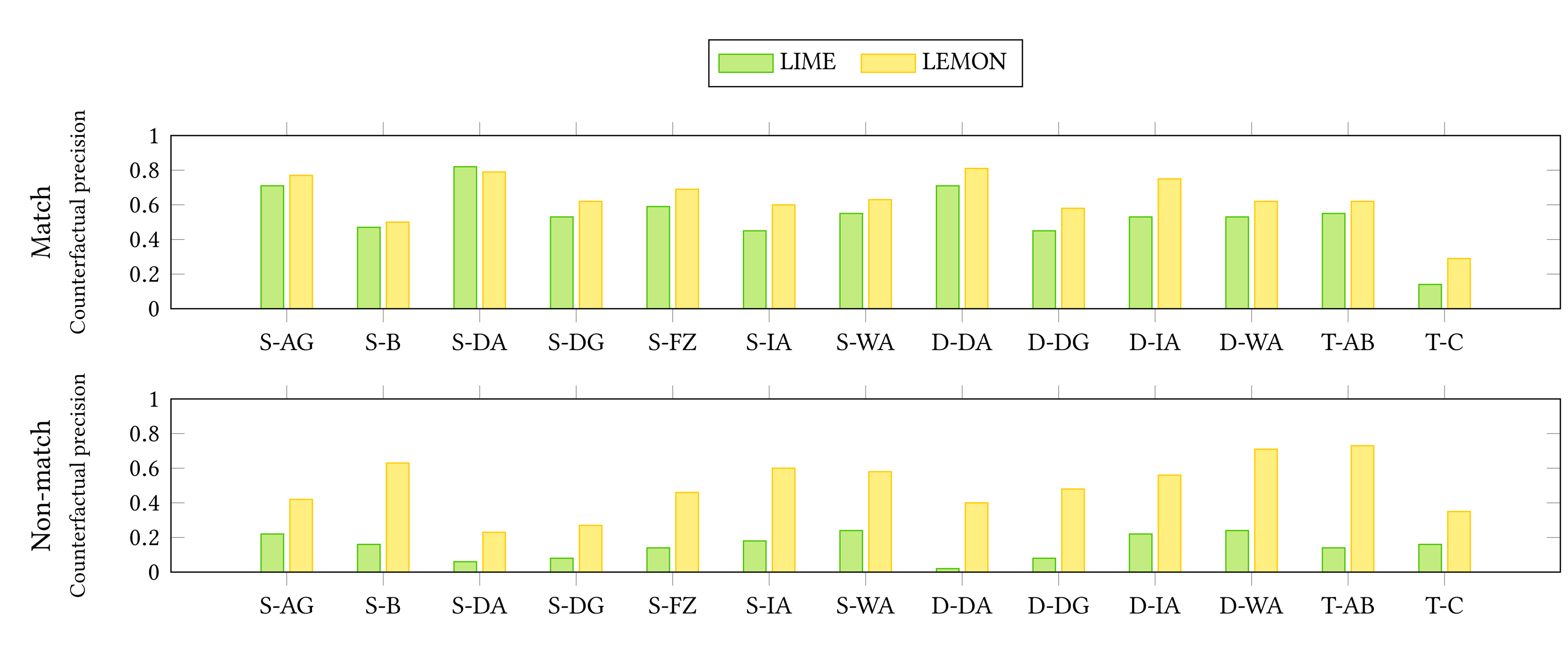

Figure 5: The counterfactual precision of real users for LIME and LEMON across 13 datasets using a Bert-based matcher. Higher is better.