RES: An Interpretable Replicability Estimation System for Research Publications

Zhuoer Wang*, Qizhang Feng*, Mohinish Chatterjee*, Xing Zhao, Yezi Liu, Yuening Li, Abhay Kumar Singh, Frank M. Shipman, Xia Hu, James Caverlee

TEXAS A&M UNIVERSITY | RICE

Abstract

Reliable and faithful research is the cornerstone of breakthrough advancements and disruptive innovations. Assessing the credibility of scientific findings and claims in research publications has long been a time-consuming and challenging task for researchers and decision-makers. We introduce RES - an intelligent system that assists humans in analyzing the credibility of scientific findings and claims in research publications in the field of social and behavioral sciences by estimating their replicability. The pipeline of RES consists of four major modules that perform feature extraction, replicability estimation, result explanation, and sentiment analysis respectively. Our evaluation based on human experts' assessments suggests that the RES has achieved adequate performance.
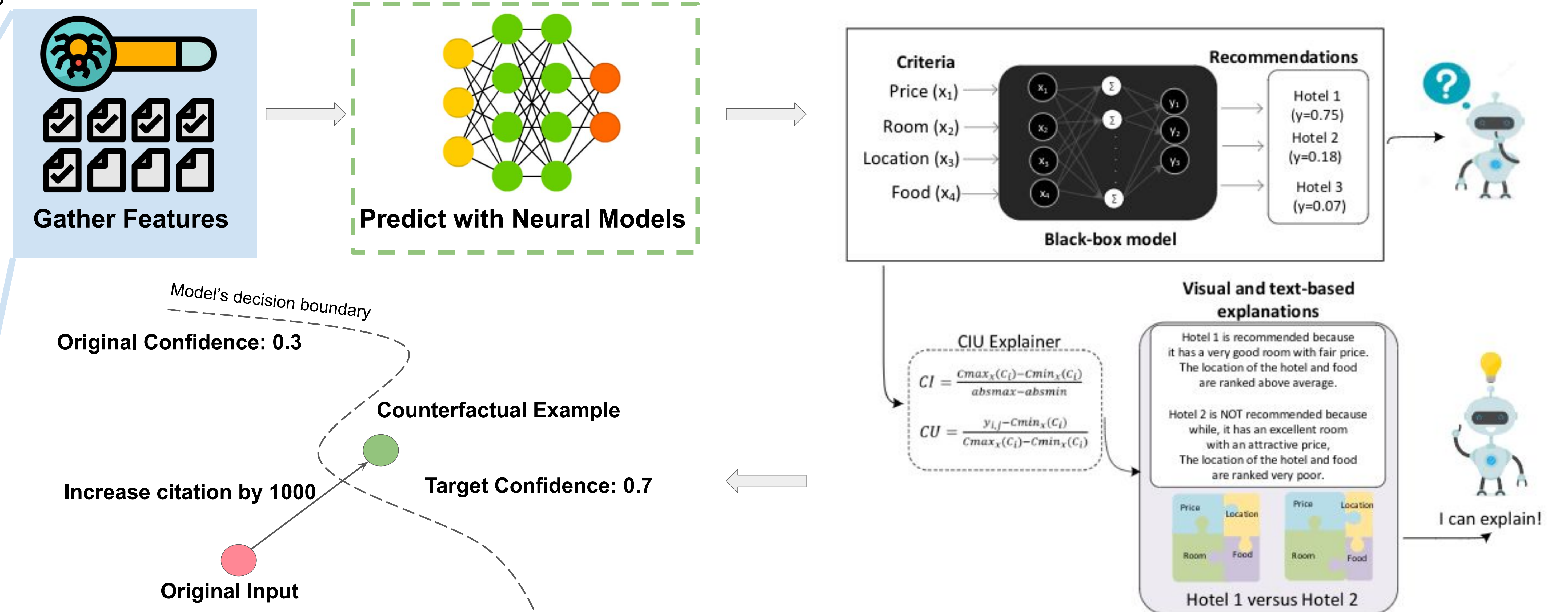
To estimate the replicability of research publications, RES exploits an extensive set of intrinsic and extrinsic features associated with each publication. Intrinsic features typically formalize a fundamental profile of the research that determines its replicability while extrinsic features are usually indirect clues of the research's replicability. After data preparation, the RES feeds those data to the trained replicability estimation model.

- The model consists of a **neural structured data regressor** and a pretrained language model **SciBERT**[1].
- The training set contains **2,400 samples** and their **replicability scores** made available through the SCORE program[2].
- The training objectives of the model are **minimizing the mean squared error** and **maximizing the ranking correlation**.
- The model achieved **0.137 RMSE** and **0.32 Spearman ranking correlation** on the test set.

[1] Beltagy, I.; Lo, K.; and Cohan, A. 2019. SciBERT: A Pre- trained Language Model for Scientific Text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP- IJCNLP), 3615–3620.

[2] Fraser, H.; Bush, M.; Wintle, B.; Mody, F.; Smith, E. T.; Hanea, A.; Gould, E.; Hemming, V.; Hamilton, D. G.; Rumpff, L.; and et al. 2021. Predicting reliability through structured expert elicitation with repliCATS (Collaborative Assessments for Trustworthy Science)

Type	Category	Examples
Extrinsic	Bibliometrics	Citation Count, Citation Velocity, Influential Citation Count
Extrinsic	Author Profile	Total Publications, Total Citations, h-index
Extrinsic	Venue Profile	Scholarly Output, Source Normalized Impact, Journal Ranking
Intrinsic	Experiment Stats	Significance, p-value, Sample Size
Intrinsic	Textual Content	Claim Abstract, Hypothesis, Test Specifications

- **Positive-Consistency**: … Thus, the meta-analytic results suggest that these analyses also successfully replicated *Eastwick and Finkel (2008)* …
- **Negative-Inconsistency**: … These results failed to replicate *Eitam et al.'s (2008)* findings even though the same goal motivation procedures they employed were used in this experiment. …
- **Positive-Similar_Findings**: … bears a strong similarity to the inhibitory effect of having many near neighbors, as found in studies by Mirman and colleagues (Chen & Mirman, 2012; Mirman, 2011; *Mirman & Magnuson, 2008*), as well the findings of Nelson, Bennett, Gee, Schreiber, & McKinney (1993) and Storkel and Adlof (2009)...
- **Negative-Limitations**: … models that assume that positional representations are wholly absolute (*G. D. A. Brown et al., 2000*; Burgess & Hitch, 1999; Lewandowsky & Farrell, 2008b) have been argued unable to handle the results of …

Gather Features

Predict with Neural Models

Criteria — Price (x_1), Room (x_2), Location (x_3), Food (x_4)

Black-box model

Recommendations — Hotel 1 (y=0.75), Hotel 2 (y=0.18), Hotel 3 (y=0.07)

Visual and text-based explanations

CIU Explainer

$$CI = \frac{Cmax_x(C_i) - Cmin_x(C_i)}{absmax - absmin}$$

$$CU = \frac{y_{i,j} - Cmin_x(C_i)}{Cmax_x(C_i) - Cmin_x(C_i)}$$

Hotel 1 is recommended because it has a very good room with fair price. The location of the hotel and food are ranked above average.

Hotel 2 is NOT recommended because while, it has an excellent room with an attractive price, The location of the hotel and food are ranked very poor.

Hotel 1 versus Hotel 2

I can explain!

Model's decision boundary

Original Confidence: 0.3

Counterfactual Example

Target Confidence: 0.7

Increase citation by 1000

Original Input

Explanation via Counterfactual Examples

Explanation via Contextual Importance and Utility*

| Using | a | large | and | unique | patent-merger | data | set | over | the | period | 1984 | to | 2006 |
| we | show | that | companies | with | large | patent | portfolios | and | low | R&D | expenses | are | acquirers |

Explanation via Additive Decomposition

Downstream Sentiment Analysis

Positive-Consistent

"Results from Study 1 replicated previous findings that acting in an explanatory vacuum triggers negative affect and a tendency to confabulate (Bar-Anan et al., 2010; Parks-Stamm et al., 2010)."

*Contextual Importance and Utility image credit: Anjomshoae, Sule, Kary Främling, and Amro Najjar. "Explanations of black-box model predictions by contextual importance and utility." International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems. Springer, Cham, 2019.