Case-Based Explanation: Making the Implicit Explicit

David Leake*

Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana, 47408 U.S.A.

Abstract

Case-based explanation (CBE) is seen by many as a compelling method for explaining black-box systems, and is advocated and pursued in substantial research in the CBR community. A 2022 position paper by Jonathan Dodge, "The Case Against Case-Based Explanation," takes a contrasting view, arguing that the use of CBE should be limited to explaining processes based on k-NN or similar approaches. This position paper takes Dodge's points as a jumping-off point to examine the nature and applicability of case-based explanation and to make explicit some of the premises which are fundamental to successful case-based explanation. It considers requirements for the domains to which CBE should be applied, the possible variants of CBE, and the knowledge it requires—both in the system and in the recipient of the explanation.

Keywords

Case-based explanation, Dimensions of explanation presentation, Explanation goals

1. Introduction

Data-driven machine learning methods such as learning with deep neural networks have achieved impressive task performance and are having great practical impact. As AI systems are applied for high-stakes tasks such as medical decision-making, legal sentencing, and loan approvals, being able to explain such systems becomes socially important; with passage of the EU General Data Protection Regulation with its "right to explanation", explanation became a legal necessity. This has led to extensive effort in explainable AI (e.g., [1]). A major thrust in this work is on augmenting black-box systems with explanation capabilities. This has led to strong interest in using prior cases to explain results of black-box AI systems, both by presenting similar cases with similar outcomes (e.g., [2]), and by presenting semifactual and counterfactual cases to help illuminate the factors relevant to a result [3].

From the early days of CBR, presentation of cases has been seen as a natural way to explain CBR system decisions (e.g., [4]). CBR systems for tasks such as decision-making, design, and planning have provided their users with the cases on which their decisions have been based, sometimes elaborating on the factors underlying processes such as similarity assessment [5].

Given the complexity and richness of explanation [6], it is clear that no explanation method will be a panacea. Accordingly, a fundamental question concerns the limitations and scope of

ICCBR XCBR'22: 4th Workshop on XCBR: Case-based Reasoning for the Explanation of Intelligent Systems at ICCBR-2022, September, 2022, Nancy, France

*Corresponding author.
leake@indiana.edu (D. Leake)
https://homes.luddy.indiana.edu/leake (D. Leake)
0000-0002-8666-34163 (D. Leake)

applicability for case-based explanation methods. In the 2022 position paper "The Case Against Case-Based Explanation" [7], Jonathan Dodge argues against the use of case presentation as a general means of explanation, with the conjecture: "It is a bad idea to deploy Case-Based Explanations when not using K-Nearest-Neighbors, or similar". Dodge notes that this position is not aimed at case-based reasoning, noting that CRR is a richer process than CBE as he describes and that CBR is compelling to people. This raises interesting questions about what case-based explanation means, how it can leverage CBR, and the range and value of uses of cases for explanation. This position paper presents an initial perspective on those questions.[1].

Dodge illustrates his meaning of case-based explanation (CBE) with two examples of explaining by similar cases. In one, a classification is explained by presenting a user with a similar example that has the same classification; in the other, by stating that the current classification is the majority classification of a set of matching examples. Beyond such methods, there is a breadth of research on explainable case-based reasoning (XCBR) which spans not only presentation of similar cases but also methods such as explaining by generating synthetic cases [9], retrieval of cases by twin systems [2], and explanation based on generating semifactual or counterfactual cases [3]. For reasons of space the paper focuses primarily on explanation by presenting similar cases from prior data, aiming to illuminate a richer view of how those cases can be used or framed, based on lessons from CBR/XCBR and how cases can leverage human experience. It also presents a start at categorizing the ways cases can be used to explain similar outcomes.

The paper makes two main contributions: first, to illuminate the variants and scope of applicability of case presentation as a form of explanation, and second, to advocate for a vision of CBE that is closer to CBR to address these issues—which in turn suggests research areas for XCBR. It aims to clarify the problems for which CBE is suitable and to underline that, reflecting the CBR cognitive model [10], good CBE must support the recipient's case-based reasoning and goals.

2. Dodge's Arguments Against (Basic) Case-Based Explanation

Dodge presents four arguments against case-based explanation: unfavorable user response ("Users seem to dislike CBE"), issues in achieving effective case representations and similarity measures ("CBE relies on weak semantic linkage"), the sorts of explanations CBE can provide ("CBE is epistemically outmatched"), and that he sees the use of CBE as restricted to domains for which training data is available and presenting cases does not violate privacy ("CBE is restrictive"). Examining these illuminates aspects of how effective case-based explanation works, when and how CBE should be applied, and potential future CBE and CBR research. We highlight selected points from each in turn, discussing lessons they suggest.

User response to CBE: The CBR process is commonly seen as natural for people [4]. Human subjects studies have supported case-based explanation (e.g., [11, 12]), and explanation by counterfactuals and semifactuals [3], though more evaluation is needed. Dodge's point on preferred

[1]The paper's title and approach are inspired by Janet Kolodner's "Making the Implicit Explicit: Clarifying the Principles of Case-Based Reasoning" [8]

explanation styles is largely based on studies of the use of CBE for assessing fairness-related questions for decision-making, primarily by Binns et al. [13] examining human judgments of alternative explanation methods for making AI decision-making fair, accountable, and transparent. Binns et al. compare four explanation methods: a demographic method, sensitivity, input influence, and cases. They compare them for two tasks, automobile insurance rating and loan qualification, for scenarios with negative outcomes. They state:

> Case-based explanations result in lower perceptions of appropriateness, fair process perception, and (in the loans case) deservedness, consistently compared to sensitivity based styles and occasionally compared to other styles. This is an effect primarily observed... in the within subject study design, indicating that the act of comparison in a particular scenario is important for these differences to become apparent.

That this effect appears primarily when other explanations can be compared might raise some question of impact for applications relying on a single method. However, more important to understanding CBE is how CBE fits their test domain. A sample domain example explains why applicant's submission for insurance was declined for a low-cost rating, with the explanation:

> This decision was based on thousands of similar cases from the past. For example, a similar case to yours is a previous customer, Claire. She was 38 years old, with 18 years of driving experience, drove 850 miles per month, occasionally exceeded the speed limit, and 25% of her trips took place at night. Clair was involved in one accident in the following year.

For accident prediction, the stochastic nature of accidents means that the basic CBR premise of "similar problems have similar solutions" does not hold. Thus to the extent a subject treats this as suggesting a case-based prediction, it is not compelling and skepticism is justified. On the other hand, CBR would be expected to be more accurate for a more deterministic task domain such as real estate appraisal, and even more so for domains with a fuller causal characterization. This suggests one requirement for effective CBE:

- **If CBR isn't appropriate, CBE shouldn't be expected to be**

Likewise, CBR quality depends on the quality of the case base. Binns et al. report that some subjects expressed that the coverage of examples appeared insufficient. Establishing whether case base coverage is sufficient is a core concern of CBR, reflected in extensive work on case-base competence [14]. This suggests another requirement for successful CBE:

- **Explaining system competence matters**

Issues in Case Representation and Similarity Measures: A second point by Dodge is that CBE relies on case distance, which may only weakly reflect semantic characteristics of the domain. He correctly observes that the meaning of a given distance value may vary at different points in the space, and that it is possible not all relevant features may be included in a case representation. These are challenges for CBR, and care is required. However, CBR research has successfully devised feature representation schemes in many domains, and similarity may itself be explained, so this does not preclude the use of CBE.

	Case coupled to system reasoning	Case relevance Explained
CBE_0 (Dodge)		
CBE_C (E.g., twins)	✓	
CBE_R		✓
CBE_{CR}	✓	✓

Table 1
Explanatory case presentation variants

The Types of Explanations CBR May Provide: Dodge correctly observes that CBE does not provide a proof entailing an outcome and asks what sort of evidence CBR provides, stating that CBE describes an outcome, but does not provide justification. However, one may distinguish different possible explanations provided by CBE, each useful in the right context:

- If the explainer is confident of the accuracy of the prior case, and if the domain satisfies the CBR assumption of problem-solution regularity [15], CBE can (on average) provide assurances about the outcome (but not process) of decision of the system being explained.
- If the recipient is familiar with the domain, and/or the CBE system can explain its retrieval (e.g., as in explaining similarity criteria or presenting cases to delimit decision boundaries), or relevant adaptations, it can empower the recipient to assess result quality.
- If the similarity criteria of CBE retrieval are coupled to the factors used in decision-making by a black-box system, as in CBR twin systems [2], the CBE process both explains the result and provides a level of explanation of the black box process.

Data Access Constraints on CBR: Dodge also expresses concern that "CBE is restrictive" because it requires domains in which one may access the training data, and there may be privacy concerns for doing so. Privacy concerns have received relatively little research attention (cf. [16]) so suggest an avenue for research. However, as other explanation methods can have their own drawbacks (e.g., unintentionally revealing proprietary learned insights), there is no magic bullet.

3. Facets of Case-Based Explanation

The previous section illustrated that CBE itself is not unitary—there are possible variations in what CBE presents to an explanation recipient, and the fit between a variant and (1) the task domain, and (2) explainer needs to support reasoning, will determine suitability of CBE.

Table 1 summarizes four possible variants of similar case CBE explanation. CBE_0, the basic variant described by Dodge, refers to explaining system outcomes based on cases retrieved using similarilty criteria not necessarily related to factors affecting system processing. A second form, CBE_1 explains by retrieved cases that are coupled to the reasoning of the system being explained, as in CBR twin systems [2]. Alternatively, CBE_0 can present cases based on criteria not necessarily related to those of the system being explained, but have its own similarity criteria explained (CBE_R). Finally, both augmentations of CBE_0 can be applied, giving CBE_{CR}.

An important alternative perspective is explanation by counterfactual cases, which has become the subject of intense research interest—As of 2022, with over 100 counterfactual explanation methods in the literature [3]

Finally, it is clear that in many contexts goals strongly affect what constitutes a "good" explanation [6, 17]. This affects the relevance of the cases presented—and which cases will be good explanations in a given context.

4. Conclusions

Case-based explanation is powerful for explaining black-box systems. However, it is not a single unitary approach, nor is is a panacea. Taking recent criticisms of CBE as a starting point, this position paper has argued that the question should not be "is CBE good or bad?", but rather what are the different forms of CBE, and when is each appropriate.

The paper makes explicit some requirements that are implicit in CBR approaches to CBE:

- CBE is only appropriate if CBR is too—for domains in which similar problems predict similar solutions, with quality cases and competent case bases
- The convincingness of CBR depends on the recipient accepting the CBR process including original case quality and system competence
- CBE could be used either (a) to establish trust in an answer or (b) to establish trust in a system. These have different requirements; for (a), any of the CBE variants are sufficient, for (b), only CBE_C or CBE_{CR}

The effectiveness of case-based explanation may depend on going beyond presentation of cases to support the recipient in full CBR by explaining similarity and adaptation as well. This view of CBR also suggests the conjecture that the better subjects are at performing CBR in a domain, the wider the range of case-based explanations they will find useful and the more compelling they will find CBE.

Acknowledgment

This work was funded by the the Department of the Navy, Office of Naval Research (Award N00014-19-1-2655). The author thanks the anonymous reviewers for insightful comments.

References

[1] D. Gunning, D. W. Aha, DARPA's explainable artificial intelligence (XAI) program, AI Magazine 40 (2019) 44–58.

[2] E. M. Kenny, M. T. Keane, Twin-systems to explain artificial neural networks using case-based reasoning: Comparative tests of feature-weighting methods in ANN-CBR twins for XAI, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2019, pp. 2708–2715.

[3] M. T. Keane, E. M. Kenny, E. Delaney, B. Smyth, If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual XAI techniques, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, ijcai.org, 2021, pp. 4466–4474.
[4] D. Leake, CBR in context: The present and future, in: D. Leake (Ed.), Case-Based Reasoning: Experiences, Lessons, and Future Directions, AAAI Press, Menlo Park, CA, 1996, pp. 3–30. Http://www.cs.indiana.edu/˜leake/papers/a-96-01.html.
[5] S. Massie, S. Craw, N. Wiratunga, A visualisation tool to explain case-base reasoning solutions for tablet formulation, in: Proceedings of the 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer-Verlag, Berlin, 2004.
[6] F. Sormo, J. Cassens, A. Aamodt, Explanation in case-based reasoning—perspectives and goals, Artificial Intelligence Review 24 (2005) 109–143.
[7] J. Dodge, The case against case-based explanation, in: IUI-WS 2022 Workshops at the International Conference on Intelligent User Interfaces (IUI, volume 3124 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 175–180.
[8] J. Kolodner, Making the implicit explicit: Clarifying the principles of case-based reasoning, in: D. Leake (Ed.), Case-Based Reasoning: Experiences, Lessons, and Future Directions, AAAI Press, Menlo Park, CA, 1996, pp. 349–370.
[9] C. Nugent, P. Cunningham, A case-based recommender for black-box systems, Artificial Intelligence Review 24 (2005) 163–178.
[10] D. Leake, Cognition as case-based reasoning, in: W. Bechtel, G. Graham (Eds.), A Companion to Cognitive Science, Blackwell, Oxford, 1998, pp. 465–476.
[11] P. Cunningham, D. Doyle, J. Loughrey, An evaluation of the usefulness of case-based explanation, in: Case-Based Reasoning Research and Development: Proceedings of the Fifth International Conference on Case-Based Reasoning, ICCBR-03, Springer-Verlag, Berlin, 2003, pp. 122–130.
[12] D. Doyle, P. Cunningham, P. Walsh, An evaluation of the usefulness of explanation in a case-based reasoning system for decision support in bronchiolitis treatment, Computational Intelligence 22 (2006) 269–281.
[13] R. Binns, M. Van Kleek, M. Veale, U. Lyngs, J. Zhao, N. Shadbolt, 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–14.
[14] B. Smyth, M. Keane, Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems, in: Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Mateo, 1995, pp. 377–382.
[15] D. Leake, D. Wilson, When experience is wrong: Examining CBR for changing tasks and environments, in: Proceedings of the Third International Conference on Case-Based Reasoning, Springer Verlag, Berlin, 1999, pp. 218–232.
[16] H. Montenegro, W. Silva, A. Gaudio, M. Fredrikson, A. Smailagic, J. S. Cardoso, Privacy-preserving case-based explanations: Enabling visual interpretability by protecting privacy, IEEE Access 10 (2022) 28333–28347.
[17] D. Leake, Goal-based explanation evaluation, Cognitive Science 15 (1991) 509–545.