Causal Banzhaf Value for Aggregate Query Explanations

Pouya Khani[1,*], Ira Assent[1]

[1]Department of Computer Science, Aarhus University, Aarhus, Denmark

Abstract

Aggregate queries are essential for summarizing data and obtaining condensed information. Explaining such queries—by identifying how specific predicates influence the result—provides deeper insights into the factors shaping query outcomes. However, existing statistical, interventional, and game theoretic explanation methods lack causal grounding, while causal methods require complete causal graphs, which are rarely available in large databases. To address this, we propose Causal Banzhaf Value (CBV): introducing causal awareness into Banzhaf values, our CBV method delivers explanations even in the absence of full causal graphs. Experiments on real world data demonstrate that CBV is computationally efficient, aligns with human intuition, and is consistent with causal explanations.

Keywords

Explainability, Query Answering, Data Analytics, Explainable Data Management

1. Introduction

The increasing reliance on data-driven decision-making in fields such as business, healthcare, and science amplifies the importance of query explanations in understanding patterns, trends, and anomalies [1, 2, 3]. Aggregate queries, such as averages or sums, play a pivotal role in summarizing high-dimensional data but pose challenges for understandability. Analysts often need explanations, such as the contributions of individual data segments or predicates[1] to understand results, especially in high-dimensional datasets where interactions and dependencies are complex [4, 5].

Example *Consider the Stack Overflow Developer Survey [6] with features such as age, developer role, education level, and salary. An analyst might pose the aggregate query*

```
SELECT AVG(Salary) FROM StackOverflow;
```

to retrieve the average salary. Still, understanding why it takes a specific value requires additional explanation; e.g., predicate `{Role = C-level Executive}` *might increase salary, while* `{Age = 25}` *might decrease it. Aggregate query explanation breaks down results into additive contributions of predicates, enabling analysts to identify key factors and make informed decisions or policy recommendations.*

However, accurately attributing the importance of individual predicates is challenging due to feature interdependencies and causal relationships. Existing techniques, such as DIFF [2] or MacroBase [4], fail to capture these dependencies, while game-theoretic methods like Shapley [7] and Banzhaf values [8] lack causal awareness [9]. Causal approaches, including XInsight [1] and CauSumX [10], rely on fully specified causal models, which are often computationally expensive and impractical to construct [3].

To address these limitations, we introduce **Causal Banzhaf Values (CBV)**, a novel approach for causally informed query explanations. CBV integrates causal knowledge of feature dependencies and employs conditional sampling [11] to estimate contributions accurately. E.g., it accounts for interdependence between `Age` and `Education`, ensuring causal information is reflected in the explanation. Unlike methods focusing on entire features, CBV evaluates the importance of predicates, offering more granular insights. For example, identifying that `Role = C-level` drives salary outcomes is more informative than attributing importance to feature `Role`. Moreover, CBV does not require a fully specified causal graph, making it suitable for scenarios where defining complete causal models is infeasible. Finally, CBV is computationally efficient, balancing accuracy with practicality for real-world applications.

2. Related Work

Statistical query explanation, e.g. [2, 4], finds associations; relational methods, e.g. [12], trace data transformations through relational operations like joins. Intervention-based approaches, e.g. [13], detect outliers in aggregate queries. All lack the ability to provide causal explanations. Game-theoretic methods like Shapley [14] and Banzhaf values [15], capture interactions between features, but assume feature independence and disregard causal relationships [9]. Also, the computational demands of simulating interventional scenarios grow exponentially with dimensionality, limiting scalability. OLAP explanations mostly focus on predefined query structures (cubes) rather than analyzing feature interactions. [16] extend OLAP operations with abstract high-level interpretability mechanisms, such as unexpectedness. [17] annotate OLAP cubes, focusing on statistical rather than causal relationships [18]. Causal approaches, such as XInsight [1] and CauSumX [10], assume fully specified causal models between all features, which are often difficult or infeasible to construct [3], or even completely unavailable, depending on the domain. Still, some knowledge about causal feature interactions is usually available, e.g., from domain knowledge or causal knowledge discovery approaches. Our approach can leverage partial causal knowledge, meaning CBV can operate even when no complete causal graph is available. This makes it more practical than fully interventional causal methods, while still incorporating causal awareness beyond statistical or game-theoretic methods.

3. Predicate Attribution

The **Banzhaf Value** quantifies the influence of each predicate (feature-value pair) on a query outcome by computing its *marginal contribution* across all possible subsets of predicates. Given a function $\nu : 2^N \to \mathbb{R}$ that assigns a value to

DOLAP 2025: 27th International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data, co-located with EDBT/ICDT 2025, March 25, 2025, Barcelona, Spain

*Corresponding author.
pouya.khani@cs.au.dk (P. Khani); ira@cs.au.dk (I. Assent)
0000-0002-4786-0572 (P. Khani); 0000-0002-1091-9948 (I. Assent)

[1]Please note that we use the term "predicate" to refer to feature values for explanations rather than to selection predicates in SQL queries.

each subset S of predicates, representing the query result (e.g., average salary), the Banzhaf Value of predicate $p(i = u)$ (feature i with value u) is:

$$\beta_{p(i=u)} = \frac{1}{2^{n-1}} \sum_{S \subseteq N \setminus \{p(i=u)\}} \left[v(S \cup \{p(i = u)\}) - v(S) \right],$$

where $v(S \cup \{p(i = u)\}) - v(S)$ measures the **marginal contribution** of predicate $p(i = u)$ when added to subset S.

Example *Let predicates* p_1: `{Education = "Master's"}`, p_2: `{Role = "Software Engineer"}`, p_3: `{YearsCoding = 10}`, *and assumed salary:* $v(\emptyset) = 50K$, $v(\{p_1\}) = 55K$, $v(\{p_2\}) = 60K$, $v(\{p_3\}) = 65K$, $v(\{p_1, p_2\}) = 70K$, $v(\{p_1, p_3\}) = 75K$, $v(\{p_2, p_3\}) = 80K$, $v(\{p_1, p_2, p_3\}) = 85K$, we calculate Marginal Contributions (MC) for p_1: $\mathrm{MC}(\emptyset, p_1) = 55K - 50K = 5K$, $\mathrm{MC}(\{p_2\}, p_1) = 70K - 60K = 10K$, $\mathrm{MC}(\{p_3\}, p_1) = 75K - 65K = 10K$, $\mathrm{MC}(\{p_2, p_3\}, p_1) = 85K - 80K = 5K$. *Their average is* p_1*'s Banzhaf Value:*

$$\beta_{p_1} = \frac{1}{2^{3-1}}(5K + 10K + 10K + 5K) = \frac{30K}{4} = 7.5K.$$

Thus, on average, predicate `Education = "Master's"` *contributes* $7.5K$ *to salary.*

Banzhaf Value is particularly suited here because it considers subsets, so is inherently order-independent. This aligns with query explanations, where order of predicates does not influence their contribution (e.g., `{Education = Master's, Developer Role = CTO}` is equivalent to `{Developer Role = CTO, Education = Master's}`). In contrast, Shapley Value [14] relies on coalition-based marginal contributions which inherently consider order. Banzhaf Value computation is thus more efficient, as it avoids the factorial complexity of Shapley Value [15].

Banzhaf Value (BV) offers a systematic framework for attributing contributions by evaluating all possible subsets of predicates. However, its limitations become evident when applied to aggregate query explanations. BV assumes independence among predicates, overlooking dependencies such as between `Age = 10` and `Education level = Doctoral Degree`. This lack of causal awareness leads to misrepresentation of contributions. Also, BV distributes contributions symmetrically, failing to account for the hierarchical and asymmetric nature of causal chains (e.g., *Education* affects *Role*, which influences *Salary*). As a result, foundational predicates are undervalued. BV also incurs high computational overhead by evaluating all subsets exhaustively, which becomes impractical for high-dimensional datasets. CBV addresses these issues by integrating causal knowledge via partial DAGs to respect predicate dependencies, enabling accurate attribution. Unlike BV, CBV attributes contributions to specific predicate-value pairs, capturing their individual impacts while reflecting causal asymmetry. By focusing on causally valid subsets and using sampling techniques, CBV achieves efficiency without sacrificing accuracy. For predicates without causal ancestors, CBV defaults to BV, ensuring consistency.

4. Causal Banzhaf Value

Traditional Banzhaf Value fairly attributes contributions by averaging marginal effects but assumes feature independence. For instance, in a causal chain where `Education Level` influences `Job Role`, which then affects `Salary`, it treats `Education Level` and `Job Role` as independent contributors, potentially misestimating their true impact.

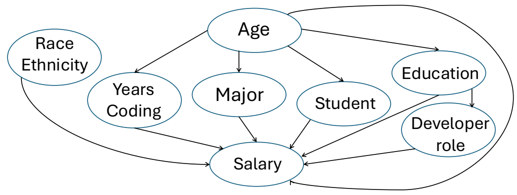


Figure 1: Partial causal DAG, adapted from the full causal model in [10] by removing one feature and several edges.

We propose the Causal Banzhaf Value (CBV), which incorporates causal knowledge into attribution by working with partial causal graphs to leverage available knowledge without requiring complete DAGs. CBV focuses on causally valid subsets to maintain consistency and employs conditional sampling to estimate contributions efficiently. This combination of partial causal integration, flexibility, and efficiency makes CBV a practical alternative to causal methods. Unlike Banzhaf value in Explainable AI [19], which focuses on features, CBV evaluates the contributions of feature values (*predicates*). This aligns with query explanation needs, where specific feature values, such as `Role = C-level`, drive outcomes. Here, causal relationships are at attribute level (e.g., Education), and contributions at predicate level (e.g., "Master's"). Predicate-level causal relationships are left for future work.

CBV incorporates causal dependencies by considering causally valid subsets of features, defined based on a partial directed acyclic graph (DAG) G: for each feature i, we identify its ancestors $A_i = \{j \in N \mid j \prec_G i\}$, where $\prec_G$ represents the causal ordering in G (Algorithm 1, line 2). Then, we identify the set of all causally valid subsets for feature i, denoted as $\mathcal{S}_i$ (Algorithm 1, line 3):

$$\mathcal{S}_i \leftarrow \{S \mid A_i \subseteq S, S \subseteq N \setminus \{i, Y\}\},$$

Thus, a subset S is valid if it contains all causal ancestors of i (A_i), and contains neither feature i itself nor the target variable Y. This ensures that attributions respect the known causal structure, while otherwise adopting the assumption of order independence as in the Banzhaf value for feature combinations where order is not known to impact outcomes. For example, $v(\texttt{Gender} = \texttt{Male} \mid \texttt{Education} = \texttt{Bachelor's}, \texttt{Age} = 25)$ is identical to $v(\texttt{Gender} = \texttt{Male} \mid \texttt{Age} = 25, \texttt{Education} = \texttt{Bachelor's})$.

The Causal Banzhaf Value (CBV) β^{C} for a predicate $p(i \text{ op } u)$ of feature i, a comparison operator op such as equality $=$, and value u is defined as:

$$\beta^{\mathrm{C}}_{p(i \text{ op } u)} = \frac{1}{2^{n-1}} \sum_{S \in \mathcal{S}_i} \sum_{\mathbf{r} \in \mathcal{R}(S)} \left[v(S_{\mathbf{r}} \cup \{p(i \text{ op } u)\}) - v(S_{\mathbf{r}}) \right],$$

where $\mathcal{R}(S)$ is the set of all possible realizations (value assignments) of features in subset S, and $v(S_{\mathbf{r}})$ is the expected value of the target variable Y conditioned on subset S with realization $\mathbf{r}$. For simplicity, we adopt the equality operator, i.e. predicate $p(i = u)$, in the following presentation.

To estimate $v(S_{\mathbf{r}})$ and $v(S_{\mathbf{r}} \cup \{p(i = u)\})$, we employ *conditional sampling* [11], which maintains feature dependencies and provides high estimation accuracy. For each subset S, realizations $\mathbf{r}^{(n)}$ are sampled from the empirical distribution $\hat{P}(S)$ (Algorithm 1, line 7), while features not in $S \cup \{i\}$ are sampled conditionally as $\mathbf{x}^{(n)} \sim \hat{P}(X \setminus (S \cup \{i\}) \mid S = \mathbf{r}^{(n)})$ (Algorithm 1, line 8). These samples allow the estimation of the expected value $v(S_\mathbf{r})$ (Algorithm 1, line 13):

$$v(S_\mathbf{r}) \leftarrow \frac{1}{M} \sum_{n=1}^{M} Y(\mathbf{r}^{(n)}, \mathbf{x}^{(n)}),$$

and similarly for $v(S_\mathbf{r} \cup \{p(i = u)\})$, where $p(i = u)$ is fixed during sampling (Algorithm 1, line 14):

$$v(S_\mathbf{r} \cup \{p(i = u)\}) \leftarrow \frac{1}{M} \sum_{n=1}^{M} Y(\mathbf{r}^{(n)}, i = u, \mathbf{x}^{(n)}_{i=u}).$$

Here, $Y(\mathbf{r}^{(n)}, \mathbf{x}^{(n)})$ is the estimated value of target variable Y given a sampled realization $\mathbf{r}^{(n)}$ of the subset S (Algorithm 1, line 9), and $\mathbf{x}^{(n)}$, which includes the sampled values for all features not in $S \cup \{i\}$. This term reflects the target outcome based on the sampled configuration of the subset S and the conditionally sampled remaining features. Similarly, $Y(\mathbf{r}^{(n)}, i = u, \mathbf{x}^{(n)}_{i=u})$ captures target variable Y under the same realization of S, but with feature i explicitly set to value u (Algorithm 1, line 10). The term $\mathbf{x}^{(n)}_{i=u}$ corresponds to the sampled values of the remaining features conditioned on $S = \mathbf{r}^{(n)}$ and $i = u$. By setting $i = u$, this estimate reflects the impact of the specific value u for feature i on the target variable Y, considering the dependencies defined in the data distribution (Algorithm 1, line 11).

Example *Respecting causal dependencies (Fig. 1), for e.g.* `Education = Master's`*, CBV finds causally valid subsets* $\mathcal{S}_{Education}$*, and for each* $S \in \mathcal{S}_{Education}$*, marginal contribution* $v(S_\mathbf{r} \cup \{p(\texttt{Education = Master's})\}) - v(S_\mathbf{r})$*, where* $v(S_\mathbf{r})$ *is the average salary conditioned on* $\mathbf{r}$*, the realization of predicates in S. CBV aggregates these contributions across all subsets and realizations to quantify causally consistent importance.*

Enumerating all subsets and performing conditional sampling can be computationally intensive, especially for high-dimensional datasets. To address this, We propose to approximate CBV by employing Monte Carlo sampling [20], which involves randomly selecting a suitable number of subsets M from $\mathcal{S}_i$ and estimating the corresponding expected values. This approach balances efficiency and accuracy (Alg. 1).

Algorithm 1 CBV

Require: Dataset D, features $N = \{1, 2, \ldots, n\}$, target variable Y; partial causal DAG G; number of samples M

Ensure: Contributions $\beta^{\text{CBV}}_{p(i=u)}$ for all features $i \in N$ and unique values for that feature $u \in U_i$

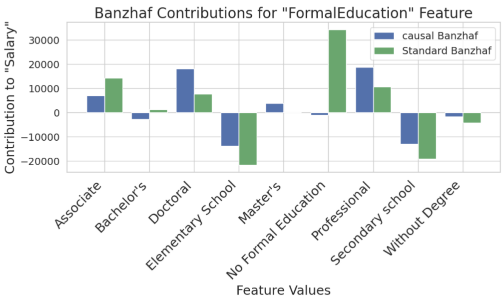
1: **for** $i \in N$ **do**
2: $\quad A_i \leftarrow \{j \in N \mid j \prec_G i\}$
3: $\quad \mathcal{S}_i \leftarrow \{S \mid A_i \subseteq S, S \subseteq N \setminus \{i, Y\}\}$
4: $\quad$ **for** $u \in U_i$ **do**
5: $\quad\quad$ **for** $S \in \mathcal{S}_i$ **do**
6: $\quad\quad\quad$ **for** $m = 1$ to M **do**
7: $\quad\quad\quad\quad \mathbf{r}^{(n)} \sim \hat{P}(S)$
8: $\quad\quad\quad\quad \mathbf{x}^{(n)} \sim \hat{P}(X_{\setminus (S \cup \{i\})} \mid S = \mathbf{r}^{(n)})$
9: $\quad\quad\quad\quad Y^{(n)} = Y(\mathbf{r}^{(n)}, \mathbf{x}^{(n)})$
10: $\quad\quad\quad\quad \mathbf{x}^{(n)}_{i=u} \sim \hat{P}(X_{\setminus (S \cup \{i\})} \mid S = \mathbf{r}^{(n)},\ p(i = u))$
11: $\quad\quad\quad\quad Y^{(n)}_{p(i=u)} = Y(\mathbf{r}^{(n)}, p(i = u), \mathbf{x}^{(n)}_{i=u})$
12: $\quad\quad\quad$ **end for**
13: $\quad\quad\quad v(S_\mathbf{r}) \leftarrow \frac{1}{M} \sum_{n=1}^{M} Y^{(n)}$
14: $\quad\quad\quad v(S_\mathbf{r} \cup \{p(i = u)\}) \leftarrow \frac{1}{M} \sum_{n=1}^{M} Y^{(n)}_{p(i=u)}$
15: $\quad\quad\quad \beta^{\text{CBV}}_{p(i=u)} \leftarrow \frac{1}{2^{n-1}} [v(S_\mathbf{r} \cup \{p(i = u)\}) - v(S_\mathbf{r})]$
16: $\quad\quad$ **end for**
17: $\quad$ **end for**
18: **end for**

5. Experimental Evaluation

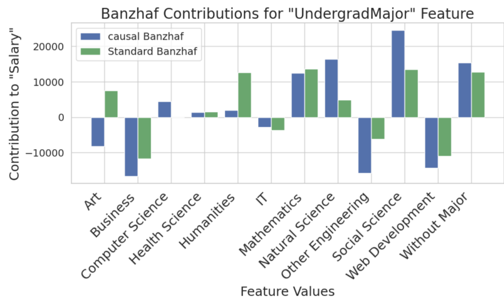
We implement Banzhaf Value (BV) [14] and Causal Banzhaf Value (CBV) using PyTorch with GPU acceleration on an NVIDIA T4 GPU, with 500 Monte Carlo samples for marginal contribution estimates via conditional sampling. Experiments are conducted on Stack Overflow Developer Survey [6], offering insights into developer demographics, education, roles, and salaries. A partial causal DAG is created from the complete version in [10] by excluding one feature and several edges, see Figure 1.

Figure 2 illustrates BV and CBV contributions for different features (plots (a)-(g)), and for different equality and range predicates. For features without ancestors in the DAG (*race/ethnicity, age* in Fig. 2g, Fig. 2f), BV and CBV produce identical results, as causal dependencies are absent. We now discuss cases where CBV and BV differ.

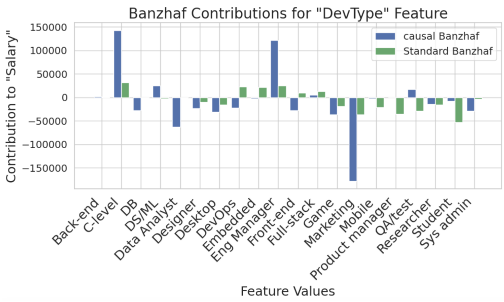
For *FormalEducation* (Fig. 2a), BV underestimates contributions of advanced degrees (*Doctoral*, *Master's*, *Professional*) and overestimates those of elementary and secondary school by treating all effects as direct, ignoring downstream roles like *developer role* and upstream factors like *Age*, which significantly drive salaries. A notable example is "No Formal Education", which BV incorrectly attributes as highly important for higher salaries. CBV reveals that its apparent importance is actually due to its causal ancestor *Age*—many individuals in the dataset with no formal education are older and hold senior roles. For *UndergradMajor* (Fig. 2b), BV overestimates contributions for fields like *Mathematics* and *Social Science*, which often affect salaries indirectly through roles or skills. CBV reduces these contributions while increasing contributions for fields like *Computer Science*, which have direct links to high-paying roles, aligning better with domain knowledge in the causal model. For *DevType* (Fig. 2c), BV misses the importance of roles like *Marketing* and *C-level*. CBV redistributes contributions, assigning higher values to roles like *C-level*, influenced by education and age (experience), and lower values to roles like *Marketing*. While negative attribution to "Marketing", and minor attribution to "Student" may be surprising, it actually captures the underlying causality: *Age* is ancestor of *DevType*. Students tend to be young, and with limited work experience, which is captured by CBV. People in *marketing* roles exhibit a different distribution concerning age and experience, where others with similar age and experience tend to have higher salaries. Unlike BV, CBV recognizes this pattern. For *YearsCoding* (Fig. 2d), BV attributes all contributions directly to coding experience, inflating the importance of mid-level ranges (*12–14 years, 15–17 years*). CBV incorporates the causal dependency of *YearsCoding* on *Age*, recognizing that older individuals naturally accumulate more experience, which indirectly influences salary. This adjustment results in more accurate contributions. For *Student* status (Fig. 2e), BV suggests similar contributions for *full-time* and *part-time students*, which contradicts domain knowledge, as full-time students typically have less time for work and lower salaries. This issue arises because BV ignores *Age* as an ancestor of *student status* in the DAG, which appears to be the main cause for salary levels, whereas CBV contributions are in



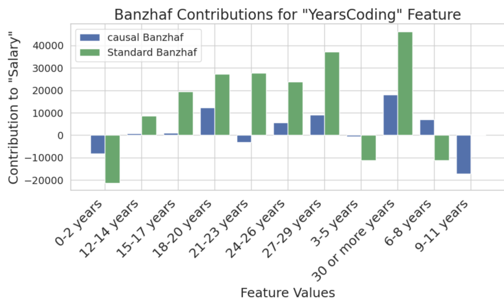
(a) Predicates in Formal Education



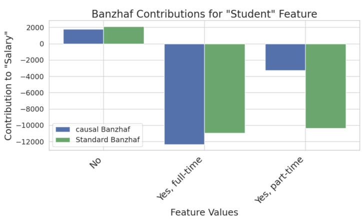
(b) Predicates in Undergraduate Major



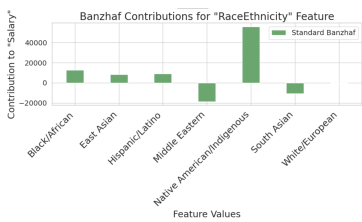
(c) Predicates in Developer Type



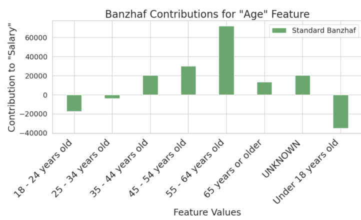
(d) Predicates in Years Coding



(e) Predicates in Student



(f) Predicates in Race / Ethnicity



(g) Predicates in Age

Figure 2: CBV and BV Stack Overflow predicate importance for salary. CBV obtains predicate importance in line with the partial causal model in Figure 1. Plots (a)-(e) show CBV finds notably different importance attributions where BV may result in misleading conclusions. For features in plots (f) and (g) without any ancestors in the causal model, BV and CBV results are identical (thus single bar only).

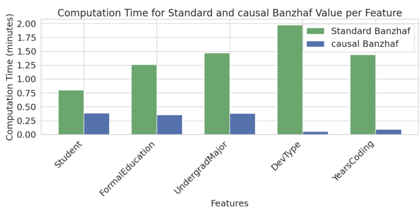


Figure 3: Runtime of predicate explanations aggregated per feature. CBV consistently outperforms BV.

line with domain knowledge. For race/ethnicity (*RaceEthnicity*, Fig. 2f) and Age (*Age*, Fig. 2g), BV and CBV results coincide exactly, as the DAG has no causal incoming edges for these features. Both methods highlight disparities in salary outcomes for e.g. *Native Americans* and higher age.

CBV is much faster than BV (Fig. 3). By focusing only on causally valid subsets, CBV avoids unnecessary computations on infeasible configurations, substantially reducing runtime. This improvement is especially notable for features with many causal dependencies, such as *developer role*. Even with Monte Carlo sampling applied to both methods, CBV is faster as it limits sampling and evaluation to valid combinations, reducing overhead and effort. For features *age* and *race/ethnicity*, without causal parents, BV and CBV produce identical results represented as a single bar for both.

6. Conclusion and Future Work

CBV presents a novel efficient causally consistent method for predicate attribution for aggregate query explanations, addressing the limitations of traditional game-theoretic or causal methods. CBV integrates partial causal knowledge without requiring complete causal graphs, making it a valuable contribution in practice. In real-world applications where causal knowledge is incomplete, the quality of explanations naturally depends on how well the available causal information reflects actual data-generating mechanisms. While conditional sampling preserves feature dependencies and enhances accuracy, it introduces computational overhead, particularly in high-dimensional settings. CBV could be extended to handle continuous domains directly, without the need to define predicate ranges (e.g., for age).

7. Acknowledgments

This work is part of RELAX (Grant Agreement No. 101072456), supported by Marie Skłodowska-Curie Actions.

References

[1] P. Ma, R. Ding, S. Wang, S. Han, D. Zhang, Xinsight: explainable data analysis through the lens of causality, Proceedings of the ACM on Management of Data 1 (2023) 1–27.

[2] F. Abuzaid, P. Kraft, S. Suri, E. Gan, E. Xu, A. Shenoy, A. Ananthanarayan, J. Sheu, E. Meijer, X. Wu, et al., Diff: a relational interface for large-scale data explanation, The VLDB Journal 30 (2021) 45–70.

[3] A. Meliou, W. Gatterbauer, K. F. Moore, D. Suciu, The complexity of causality and responsibility for query answers and non-answers, arXiv preprint arXiv:1009.2021 (2010).

[4] P. Bailis, E. Gan, S. Madden, D. Narayanan, K. Rong, S. Suri, Macrobase: Prioritizing attention in fast data, in: Proceedings of the 2017 ACM International Conference on Management of Data, 2017, pp. 541–556.

[5] Z. Miao, Q. Zeng, B. Glavic, S. Roy, Going beyond provenance: Explaining query answers with pattern-based counterbalances, in: Proceedings of the 2019 International Conference on Management of Data, 2019, pp. 485–502.

[6] Stack Overflow, Developer Survey, 2021. URL: https://survey.stackoverflow.co/2021.

[7] L. S. Shapley, A value for n-person games, Contributions to the Theory of Games 2 (1953) 307–317.

[8] J. F. Banzhaf III, Weighted voting doesn't work: A mathematical analysis, Rutgers Law Review 19 (1965) 317–343.

[9] T. Heskes, E. Sijben, I. G. Bucur, T. Claassen, Causal shapley values: Exploiting causal knowledge to explain individual predictions of complex models, Advances in neural information processing systems 33 (2020) 4778–4789.

[10] B. Youngmann, M. Cafarella, A. Gilad, S. Roy, Summarized causal explanations for aggregate views, Proceedings of the ACM on Management of Data 2 (2024) 1–27.

[11] S. Chakraborty, E. Fischer, Y. Goldhirsh, A. Matsliah, On the power of conditional samples in distribution testing, in: Proceedings of the 4th conference on Innovations in Theoretical Computer Science, 2013, pp. 561–580.

[12] C. Li, Z. Miao, Q. Zeng, B. Glavic, S. Roy, Putting things into context: Rich explanations for query answers using join graphs, in: Proceedings of the 2021 International Conference on Management of Data, 2021, pp. 1051–1063.

[13] E. Wu, S. Madden, Scorpion: explaining away outliers in aggregate queries, Proc. VLDB Endow. 6 (2013) 553–564. URL: https://doi.org/10.14778/2536354.2536356. doi:10.14778/2536354.2536356.

[14] D. Deutch, N. Frost, B. Kimelfeld, M. Monet, Computing the shapley value of facts in query answering, in: Proceedings of the 2022 International Conference on Management of Data, 2022, pp. 1570–1583.

[15] O. Abramovich, D. Deutch, N. Frost, A. Kara, D. Olteanu, Banzhaf values for facts in query answering, Proceedings of the ACM SIGMOD International Conference on Management of Data 2 (2024) 1–26.

[16] P. Vassiliadis, P. Marcel, S. Rizzi, Beyond roll-up's and drill-down's: An intentional analytics model to reinvent olap, Information Systems 84 (2019) 147–168.

[17] M. Francia, M. Golfarelli, S. Rizzi, Describing and assessing cubes through intentional analytics, in: Proceedings of the 24th International Conference on Extending Database Technology (EDBT), 2021, pp. 3–14.

[18] M. Francia, S. Rizzi, P. Marcel, Explaining cube measures through intentional analytics, Information Systems 121 (2024) 102338.

[19] A. Karczmarz, T. Michalak, A. Mukherjee, P. Sankowski, P. Wygocki, Improved feature importance computation for tree models based on the banzhaf value, in: Uncertainty in Artificial Intelligence, PMLR, 2022, pp. 969–979.

[20] A. Shapiro, Monte carlo sampling methods, Handbooks in operations research and management science 10 (2003) 353–425.