# AGR: Reinforced Causal Agent-Guided Self-explaining Rationalization

**Yunxiao Zhao[1], Zhiqiang Wang[1,2*], Xiaoli Li[3], Jiye Liang[1,2], Ru Li[1,2*]**
1. School of Computer and Information Technology, Shanxi University, Taiyuan, China
2. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China
3. Institute for Infocomm Research, A*Star, Singapore
yunxiaomr@163.com, {wangzq,ljy,liru}@sxu.edu.cn, xlli@ntu.edu.sg

## Abstract

Most existing rationalization approaches are susceptible to degeneration accumulation due to a lack of effective control over the learning direction of the model during training. To address this issue, we propose a novel approach AGR (**A**gent-**G**uided **R**ationalization), guiding the next action of the model based on its current training state. Specifically, we introduce causal intervention calculus to quantify the causal effects inherent during rationale training, and utilize reinforcement learning process to refine the learning bias of them. Furthermore, we pretrain an agent within this reinforced causal environment to guide the next step of the model. We *theoretically* demonstrate that a good model needs the desired guidance, and *empirically* show the effectiveness of our approach, outperforming existing state-of-the-art methods on BeerAdvocate and HotelReview datasets.

## 1 Introduction

To explain the prediction of neural networks, selective rationalization task (Lei et al., 2016; Yu et al., 2019, 2021) has been studied in recent years. As shown in Figure 1, *it aims to select a small and human-intelligible subset* (i.e., rationale) from the input to support and explain the prediction results when yielding them. As an interpretable diagram, rationalization holds significant potential for elucidating the decision-making process of predictive models, building trust, and deriving insightful and pertinent insights (Yuan et al., 2020; Zhang et al., 2023; Deng et al., 2023).

Various approaches have been proposed for rationalization, spanning from early rationale sampling-based methods (Bao et al., 2018; Bastings et al., 2019; Paranjape et al., 2020) to the extra-component-based methods (De Cao et al., 2020; Huang et al., 2021; Yu et al., 2021; Liu et al., 2022; Yue et al., 2022; Liu et al., 2023a). These methods predominantly concentrate on improving the performance of rationalization models by either refining the sampling directly or aligning additional information beyond the rationale, resulting in impressive results. However, to the best of our knowledge, the current methods are prone to degeneration accumulation[1] since they usually do not discern whether the generator during training has produced unmeaningful or flawed rationales; instead, they directly pass them to the predictor even if generated rationales are degraded.

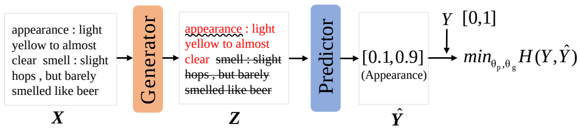


Figure 1: The standard selective rationalization, where $X, Z, \hat{Y}, Y$ represent the input text, rationale, prediction and the groundtruth label, respectively. The red text indicates the small and human-intelligible subset.

For instance, the underlined rationale in Figure 1 is degraded, as the word *appearance* alone does not reliably determine the sentiment polarity of input $X$. But the predictor overfits to this uninformative rationale and classifies the sentiment according to whether *"appearance"* is included in the rationale. Consequently, when the predictor receives degraded rationales, it steers the model towards an undesirable direction (aka., learning bias). Thus, optimizing this bias during training is crucial for ensuring the model's generalization performance.

The proposed methods (Chang et al., 2020; Zhang et al., 2023; Yue et al., 2023) fall short in considering rationalization optimization comprehensively, neglecting existing causality *during rationale learning*. Although they often employ causal theory to uncover relationships between rationale pieces, *they struggle to directly optimize*

*Corresponding author

[1]Degeneration over rationalization is a highly challenging problem, which means the predictor may overfit to meaningless rationales generated by the not yet well-trained generator (Yu et al., 2019; Liu et al., 2023b,d).



*Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 510–518
August 11-16, 2024 ©2024 Association for Computational Linguistics

*the cooperative game dynamics between the generator and predictor during training*. As shown in Figure 1, optimizing rationale from *"appearance"* to *"appearance: light yellow to almost clear"* necessitates evaluating the causal impact on target prediction, guiding the model's subsequent optimization. Thus, if we could construct a guiding signal to reward or penalize the learning behavior of the model, this would significantly reduce the model's learning bias during training, alleviating the problem of degeneration accumulation.

To address the above problems, we propose a novel rationalization method named AGR (**A**gent-**G**uided **R**ationalization), which leverages *a reinforced causal agent* to guide the cooperative game optimization *during rationale training*, as shown in Figure 2. In particular, 1) we quantify the causal effects in the rationale optimization process, and design a reinforcement learning (RL) process (e.g., *Markov decision*) to refine the learning bias during training. 2) We further pretrain an agent within reinforced causal environment to guide next actions by a system of rewards. We also theoretically illustrate that a robust model needs the desired guidance. 3) Experimental results demonstrate the effectiveness of our approach, surpassing state-of-the-art methods on BeerAdvocate and HotelReview datasets.

## 2 Problem Formulation

**Notation.** Following previous research (Liu et al., 2023b,c,d), we consider the classification problem and denote the generator and predictor as $f_G(\cdot)$ and $f_P(\cdot)$, with $\theta_g$ and $\theta_p$ representing their parameters. The input text $X = [x_1, x_2, ..., x_l] (1 \leq i \leq l)$ consists of tokens $x_i$, where $l$ is the number of tokens. The label of $X$ is a one-hot vector $Y \in \{0,1\}^c$, where $c$ is the number of categories.

**Cooperative game for rationalization**. The $f_G(\cdot)$ selects the most informative pieces from $X$ by a sequence of binary mask $M = [m_1, ..., m_l] \in \{0,1\}^l$. Then, it forms the rationale $Z = M \odot X = [m_1x_1, m_2x_2, ..., m_lx_l]$, where the informativeness of $Z$ is measured by the negative cross entropy $-H(Y, \hat{Y})$. Consequently, the $f_G(\cdot)$ and $f_P(\cdot)$ are optimized cooperatively by

$$\min_{\theta_g, \theta_p} \mathcal{H}(Y, \hat{Y} \mid f_G(X)), s.t. \hat{Y} = f_P(f_G(X)). \quad (1)$$

In addition, rationales are usually constrained by compact and coherent regularization terms $\Omega(M) = \lambda_1 \left| \frac{\|M\|_1}{l} - s \right| + \lambda_2 \sum_t |m_t - m_{t-1}|$ (Chang et al., 2020), where $s$ is a pre-defined sparsity level.

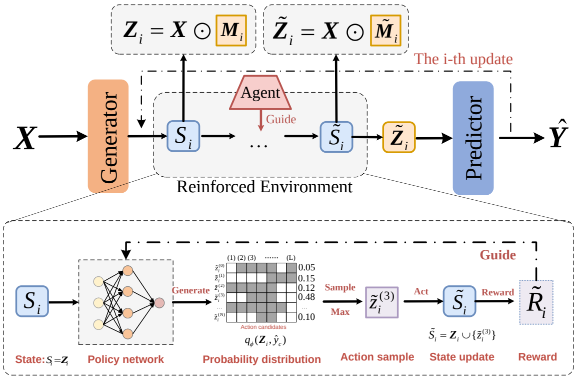


Figure 2: The architecture of AGR. $X$ and $\hat{Y}$ are the input and output. $S_i$ is the $i$-th update state of rationale, while $\widetilde{S}_i$ is the state after guidance by the agent.

## 3 Reinforced Causal Agent

In this section, we present our *reinforced causal agent*, considering both *causal effect* and *learning bias of degeneration* during rationale training.

### 3.1 Rationale Causal Attribution

Formally, we construct a rationale $\mathcal{Z}_k^*$ by maximizing an attribution metric $A(\cdot)$ in rationalization

$$\mathcal{Z}_K^* = \arg max_{\mathcal{Z}_K \subseteq X} A(\mathcal{Z}_K | \hat{y}_c), \quad (2)$$

where $A(\cdot)$ measures the contribution of each candidate $\mathcal{Z}_K$ to the target prediction $\hat{y}_c$.

However, $A(\mathcal{Z}_K|\hat{y}_c)$ needs to be quantified. To this end, we introduce causal intervention calculus $do(\cdot)$, including $do(Z = Z_K)$ and $do(Z = \varnothing)$(Pearl, 2009; Pearl et al., 2016), and reformulate the causal contribution from $\varnothing$ to $\mathcal{Z}_K$ by mutual information,

$$A(\mathcal{Z}_K|\hat{y}_c) = I(\hat{y}_c, do(\mathcal{Z}_K)) - I(\hat{y}_c, do(\varnothing)). \quad (3)$$

### 3.2 Markov Decision Process as RL

Equation 3 illustrates the procedure for deriving $\mathcal{Z}_K$ from an initial state of zero training. However, it may generate degraded rationales at step $i$, where $0 < i < K$. Thus we need to seek for quantifiable objectives between $\mathcal{Z}_i$ and $\mathcal{Z}_{i+1}$,

$$\mathcal{Z}_{i+1} = \arg max_{\mathcal{Z}_{i+1} \in \{X \setminus \mathcal{Z}_i\}} A(\mathcal{Z}_{i+1} | \mathcal{Z}_i, \hat{y}_c). \quad (4)$$

According to Equation 3, we have the causal contribution between $\mathcal{Z}_i$ and $\mathcal{Z}_{i+1}$: $A(\mathcal{Z}_{i+1}|\mathcal{Z}_i, \hat{y}_c) = I(\hat{y}_c, do(\mathcal{Z}_{i+1})) - I(\hat{y}_c, do(\mathcal{Z}_i))$. So,

$$\begin{aligned} A(\mathcal{Z}_{i+1}|\mathcal{Z}_i, \hat{y}_c) &= -H(\hat{y}_c|\mathcal{Z}_{i+1}) + H(\hat{y}_c|\mathcal{Z}_i) \\ &= -H(\hat{y}_c|\{\mathcal{Z}_i \cup \{z_{i+1}\}\}) + H(\hat{y}_c|\mathcal{Z}_i) \\ &= -p_\theta(\hat{y}_c|\mathcal{Z}) log \frac{p_\theta(\hat{y}_c|\mathcal{Z}_i)}{p_\theta(\hat{y}_c|\{\mathcal{Z}_i \cup \{z_{i+1}\}\})}, \end{aligned} \quad (5)$$

where $H(\hat{y}_c|\mathcal{Z}_i)$ is the term of conditional entropy. As a result, Equation 5 explicitly quantifies $\mathcal{Z}_{i+1}$'s effect with previously obtained rationale $\mathcal{Z}_i$.

To further promote the cooperative game, we model the training process of rationale as a Markov decision process $\mathbb{M} = \{\mathbb{S}, \mathbb{A}, \mathbb{P}, \mathbb{R}\}$, where $\mathbb{S} = \{s_i\}$ represents set of states abstracting the process of optimizing rationale during training, and $\mathbb{A} = \{a_i\}$ indicates the set of actions. In particular, The transition dynamics $\mathbb{P}(s_{i+1}|s_i, a_{i+1})$ specify how the state $s_{i+1}$ is updated from the prior state $s_i$ by taking action $a_{i+1}$. Besides, $\mathbb{R}(s_i, a_{i+1})$ quantifies the reward obtained after taking action $a_{i+1}$ based on the prior state $s_i$. Therefore, cooperative training for rationale can be depicted as the sequence process $(s_0, a_1, r_1, s_1, ..., a_K, r_K, s_K)$, where the state $s_i$ can be formulated by $s_i = Z_i$ in the *i-th* update; $s_0 = Z_0$ can be initiated by generator $f_G(\cdot)$.

Nevertheless, the above process exhibits a limitation in its inability to detect *learning bias* at any given state $s_i$. To address this, we reformulate the sequence process as ($<s_0,\widetilde{a}_0,\widetilde{r}_0,\widetilde{s}_0>$, $a_1$, $r_1$, $<s_1,\widetilde{a}_1,\widetilde{r}_1,\widetilde{s}_1>$, ..., $a_K$, $r_K$, $<s_K,\widetilde{a}_K,\widetilde{r}_K,\widetilde{s}_K>$), where $<s_i,\widetilde{a}_i,\widetilde{r}_i,\widetilde{s}_i>$ indicates process of transitioning from state $s_i$ to $\widetilde{s}_i$ in the *i-th* update.

Given the state $s_i = Z_i$, we derive the available action space: $\widetilde{\mathbb{A}}_i = \{X \backslash Z_i\}$. The searched action can be represented as

$$\widetilde{a}_i = \widetilde{z}_i, \tag{6}$$

where $\widetilde{z}_i \in \{X \backslash Z_i\}$ indicates candidate rationale in action space. Having made the action $\widetilde{a}_i$, the state transition is to merge $\widetilde{z}_i$ into $Z_i$, i.e., $\widetilde{Z}_i = Z_i \cup \{\widetilde{z}_i\}$.

To assess the effectiveness of the action $\widetilde{a}_i$ in mitigating the learning bias of the model, the reward $\widetilde{\mathbb{R}}_i(\widetilde{s}_i, \widetilde{a}_i)$ at state $s_i$ can be formulated as follows:

$$\widetilde{\mathbb{R}}_i = \begin{cases} A(\widetilde{z}_i|Z_i, \hat{y}_c^*) + 1, & iff_P(Z_i \cup \{\widetilde{z}_i\}) = \hat{y}_c^* \\ A(\widetilde{z}_i|Z_i, \hat{y}_c^*) - 1, & otherwise. \end{cases} \tag{7}$$

According to Equation 5, although we can quantify the probabilities at states $\widetilde{s}_i$ and $s_i$, and present the relevant reward $\widetilde{\mathbb{R}}_i$, obtaining $y_c^*$ poses a challenge.

### 3.3 Pretrained Agent

To address the limitation, we propose a *reinforced causal agent* in the aforementioned causal and reinforcement learning framework to better align the probability distribution of the target prediction and theoretically justify the creation of an auxiliary agent targeting $\hat{y}_c$.

**Pretrained Embedding.** We pretrain the auxiliary $agent$, denoted as $f_A(\cdot)$, with

$$\theta_A^* = arg \min_{\theta_A} \mathcal{H}(Y, \hat{Y}|X), s.t. \hat{Y} = f_A(X), \tag{8}$$

where $\theta_A$ represents the parameters of the $agent$, and $\theta_A^*$ denotes the optimal solution.

**Theorem Analysis.** Assuming $X$, $Z$, $Y$, and $\mathcal{A}$ as random variables in rationalization representing the input, rationale, label, and auxiliary variable, respectively, we propose:

**Lemma 1.** *Given X, Z, Y, $\hat{Y} = f_P(f_G(X))$. Existing a guiding variable $\mathcal{A}$ could enable the predictor $f_P(\cdot)$ to achieve good predictions. That is, a solution for $\mathcal{A}$ exists, and X is a solution of $\mathcal{A}$.*

The proof is provided in Appendix A. Lemma 1 suggests that constructing an auxiliary variable $\mathcal{A}$ aligned with $X$ for rationalization contributes to the learning of a good prediction.

## 4 Agent-Guided Rationalization

As depicted in Figure 2, following the establishment of the environment for the reinforced causal agent, we delineate the construction and training of the policy network $q_\phi$.

### 4.1 Policy Network Architecture

It takes the pair of intermediate state $\mathcal{Z}_i$ and $\hat{y}_c$ provided by $f_A(\cdot)$ as input. Formally,

$$\widetilde{z}_i \sim q_\phi(\mathcal{Z}_i, \hat{y}_c), \tag{9}$$

where $\theta_\phi$ is the trainable parameters of the policy network, and $\widetilde{z}_i$ is generated according to the probability of next action $\mathbb{P}_\phi(\widetilde{z}_i|\mathcal{Z}_i, \hat{y}_c)$.

**Representation learning of action candidates.** With the space of action candidates $\widetilde{\mathbb{A}}_i = X \backslash \mathcal{Z}_i$, our policy network first learns the representation for each action candidate $\widetilde{a}_i^{(j)} (0 < j < N)$, where $N$ is the number of candidates.

Then, we employ the encoder to encode $X \backslash \mathcal{Z}_i$ for obtaining the action representation of $\widetilde{z}_i$ by

$$e_{\widetilde{z}_i} = encoder(X \backslash \mathcal{Z}_i), \tag{10}$$

utilizing bidirectional Gated Recurrent Units (GRUs) (Cho et al., 2014) as the encoder.

**Sampling of action.** The policy network aims to select a singular action $\widetilde{a}_i = \widetilde{z}_i$ from the search space, prioritizing its relevance to the current state $s_i = \mathcal{Z}_i$. This selection process is modeled as:

$$p_{\widetilde{z}_i} = MLP([\boldsymbol{e}_{\widetilde{z}_i}; \boldsymbol{e}_{\mathcal{Z}_i}]), \tag{11}$$

where $e_{\mathcal{Z}_i}$ indicates the current rationale's representation. The selection probability for each action candidate within $\widetilde{\mathbb{A}}_i$ is computed using

$$\mathbb{P}_\phi(\widetilde{z}_i|\mathcal{Z}_i, \hat{y}_c) = softmax_{\widetilde{\mathbb{A}}_i}(p_{\widetilde{z}_i}), \tag{12}$$

where $\phi$ is the parameters collected of MLP.

| Methods | S | Appearance | | | Aroma | | | Palate | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 | P | R | F1 |
| RNP (Lei et al., 2016) | 20 | 39.4 | 44.9 | 42.0 | 37.5 | 51.9 | 43.5 | 21.6 | 38.9 | 27.8 |
| HardKuma (Bastings et al., 2019) | 20 | 64.9 | 69.2 | 67.0 | 37.0 | 55.8 | 44.5 | 14.6 | 22.3 | 17.7 |
| IB (Paranjape et al., 2020) | 20 | 59.3 | 69.0 | 63.8 | 38.6 | 55.5 | 45.6 | 21.6 | 48.5 | 29.9 |
| INVRAT (Chang et al., 2020) | 20 | 58.9 | 67.2 | 62.8 | 29.3 | 52.1 | 37.5 | 24.0 | 55.2 | 33.5 |
| DARE (Yue et al., 2022) | 20 | 63.7 | 71.8 | 67.5 | 41.0 | 61.5 | 49.3 | 24.4 | 54.9 | 33.8 |
| FR (Liu et al., 2022) | 20 | 74.9 | 84.9 | 79.6 | 58.7 | 73.3 | 65.2 | 36.6 | 59.4 | 45.3 |
| Inter-RAT (Yue et al., 2023) | 20 | 62.0 | 76.7 | 68.6 | 44.2 | 65.4 | 52.8 | 26.3 | 59.1 | 36.4 |
| MGR (Liu et al., 2023b) | 20 | 76.3 | 83.6 | 79.8 | 64.4 | 81.3 | 71.9 | 47.1 | 73.1 | 57.3 |
| AGR(Ours) | 20 | **83.7** | **87.5** | **85.6** | **67.5** | **81.4** | **73.8** | **47.6** | **77.7** | **59.0** |

Table 1: Results on BeerAdvocate, where **Bold** text indicates the best experimental results across different methods.

| Methods | Appearance | | | | Appearance | | | | Appearance | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | S | P | R | F1 | S | P | R | F1 | S | P | R | F1 |
| RNP | 10 | 32.4 | 18.6 | 23.6 | 20 | 39.4 | 44.9 | 42.0 | 30 | 24.2 | 41.2 | 30.5 |
| DARE | 10 | 63.9 | 42.8 | 51.3 | 20 | 63.7 | 71.8 | 67.5 | 30 | 45.5 | 80.6 | 58.1 |
| FR | 10 | 70.4 | 42.0 | 52.6 | 20 | 74.9 | 84.9 | 79.6 | 30 | 50.6 | 81.4 | 62.3 |
| Inter-RAT | 10 | 66.0 | 46.5 | 54.6 | 20 | 62.0 | 76.7 | 68.6 | 30 | 48.1 | 82.7 | 60.8 |
| MGR | 10 | 87.5 | 51.7 | 65.0 | 20 | 76.3 | 83.6 | 79.8 | 30 | 57.2 | 93.9 | 71.1 |
| AGR | 10 | **83.5** | **54.9** | **66.2** | 20 | **83.7** | **87.5** | **85.6** | 30 | **59.7** | **94.3** | **73.1** |

Table 2: The different sparsity results on BeerAdvocate.

| | Methods | S | P | R | F1 |
|---|---|---|---|---|---|
| Location | RNP (Lei et al., 2016) | 10.9 | 43.3 | 55.5 | 48.6 |
| | CAR (Chang et al., 2019) | 10.6 | 46.6 | 58.1 | 51.7 |
| | DMR (Huang et al., 2021) | 10.7 | 47.5 | 60.1 | 53.1 |
| | A2R (Yu et al., 2021) | 8.5 | 43.1 | 43.2 | 43.1 |
| | MGR (Liu et al., 2023b) | 9.7 | 52.5 | 60.5 | 56.2 |
| | AGR(Ours) | 9.3 | **54.9** | **60.5** | **57.6** |
| | | S | P | R | F1 |
| Service | RNP (Lei et al., 2016) | 11.0 | 40.0 | 38.2 | 39.1 |
| | CAR (Chang et al., 2019) | 11.7 | 40.7 | 41.4 | 41.1 |
| | DMR (Huang et al., 2021) | 11.6 | 43.0 | 43.6 | 43.3 |
| | A2R (Yu et al., 2021) | 11.4 | 37.3 | 37.2 | 37.2 |
| | MGR (Liu et al., 2023b) | 11.8 | 45.0 | 46.4 | 45.7 |
| | AGR(Ours) | 12.3 | **45.9** | **49.3** | **47.6** |
| | | S | P | R | F1 |
| Cleanliness | RNP (Lei et al., 2016) | 10.6 | 30.5 | 36.0 | 33.0 |
| | CAR (Chang et al., 2019) | 9.9 | 32.3 | 35.7 | 33.9 |
| | DMR (Huang et al., 2021) | 10.3 | 31.4 | 36.4 | 33.7 |
| | A2R (Yu et al., 2021) | 8.9 | 33.2 | 33.3 | 33.3 |
| | MGR (Liu et al., 2023b) | 10.5 | 37.6 | 44.5 | 40.7 |
| | AGR(Ours) | 10.3 | **39.0** | **45.5** | **42.0** |

Table 3: The experimental results on HotelReview.

### 4.2 Policy Gradient Training

Since discrete sampling within the policy network blocks gradients, we adopt policy gradient-based training framework REINFORCE (Sutton et al., 1999). The objective $\max_{\Omega}(\mathbb{L})$ is as follows:

$$\max_{\phi} \mathbb{E}_{\mathcal{Z}_i \in \widetilde{\mathbb{A}}_i} \mathbb{E}_i[\widetilde{\mathbb{R}}(\mathcal{Z}_i, \widetilde{z}_i) log \mathcal{P}_{\phi}(\widetilde{z}_i | \mathcal{Z}_i, \hat{y}_c)]. \quad (13)$$

The final task loss is a jointly optimized objective:

$$\min_{\theta_g, \theta_p} \mathcal{H}(Y, \hat{Y}) + \Omega(M) - \Omega(\mathbb{L}), s.t. \hat{Y} = f_P(f_G(X)) \quad (14)$$

## 5 Experiments

### 5.1 Datasets, Baselines and Evaluation Metrics

**Datasets.** We compare AGR using BeerAdvocate (McAuley et al., 2012) and HotelReview (Wang et al., 2010) datasets, which are two multi-aspect sentiment classification datasets widely used in rationalization. Following existing work, we obtain the data in the same way as Yue et al. (2023) for BeerAdvocate, and we preprocess HotelReview dataset in the same way as Huang et al. (2021) and Liu et al. (2023b).

**Baselines.** We compare with *eight* models for BeerAdvocate, including three *sampling-based methods*: **RNP** (Lei et al., 2016), **HardKuma** (Bastings et al., 2019), **Information Bottleneck (IB)** (Paranjape et al., 2020), and three *extra-component-based methods:* **DARE** (Yue et al., 2022), **FR** (Liu et al., 2022), **MGR** (Liu et al., 2023b), and two *causal-based methods:* **INVRAT** (Chang et al., 2020), **Inter-RAT** (Yue et al., 2023). For HotelReview dataset, we compare with *five* models, including **RNP** (Lei et al., 2016), **CAR** (Chang et al., 2019), **DMR** (Huang et al., 2021), **A2R** (Yu et al., 2021), and **MGR** (Liu et al., 2023b).

**Evaluation Metrics.** Following (Huang et al., 2021; Yu et al., 2021; Yue et al., 2023; Liu et al., 2023b), we focus on the quality of rationales, and adopt Precision (P), Recall (R), and F1-score (F1) as metrics. We perform the best results on the validation set before testing on the test set. The Appendix B provides further details in this section.

### 5.2 Performance Comparison

**Results on BeerAdvocate.** As shown in Table 1, our proposed method AGR outperforms all the eight baselines in terms of three aspects for BeerAdvocate dataset. Furthermore, in sparsity experiments (Table 2), AGR consistently outperforms the latest state-of-the-art results, affirming its effectiveness for selective rationalization.

**Results on HotelReview.** Table 3 shows that our model once again obtains the best performance

Table 4: Examples of generated rationales. Human-annotated rationales are underlined. Rationales from three models are highlighted in blue and are denoted as $Z_1$, $Z_2$ and $Z_3$ respectively.

| **FR (2022)** | **MGR (2023b)** | **AGR (Ours)** |
|---|---|---|
| **Aspect:** Beer-Appearance<br>**Label:** Positive, **Pred:** Positive<br>**Text:** i picked this beer up on a whim as i was in the mood for a good coffee stout and the siren-like figure somehow told me this is the beer for you . a bit freaky , but i went with it . i was impressed from the very first pour . like any stout , the color is a dark molasses black . but ... the head was thick and dense with good retention . the coffee aroma was intense ! the roasted goodness almost overwhelms my sense of smell .the roasted coffee flavors are the first things that i could taste along with hints of chocolate . however , i can tell there 's more complexity here than my palette can decipher . the coffee flavors bring bitterness but it 's not over powering as the sweetness of the malt cuts the bitterness quite nicely the beer has carbonation but once the bubbles have escaped the beer gives a creamy , velvety feel and finish . the alcohol was very well hidden in this beer which is scary ... | **Aspect:** Beer-Appearance<br>**Label:** Positive, **Pred:** Positive<br>**Text:** i picked this beer up on a whim as i was in the mood for a good coffee stout and the siren-like figure somehow told me this is the beer for you . a bit freaky , but i went with it . i was impressed from the very first pour . like any stout , the color is a dark molasses black . but ... the head was thick and dense with good retention . the coffee aroma was intense ! the roasted goodness almost overwhelms my sense of smell .the roasted coffee flavors are the first things that i could taste along with hints of chocolate . however , i can tell there 's more complexity here than my palette can decipher . the coffee flavors bring bitterness but it 's not over powering as the sweetness of the malt cuts the bitterness quite nicely the beer has carbonation but once the bubbles have escaped the beer gives a creamy , velvety feel and finish . the alcohol was very well hidden in this beer which is scary ... | **Aspect:** Beer-Appearance<br>**Label:** Positive, **Pred:** Positive<br>**Text:** i picked this beer up on a whim as i was in the mood for a good coffee stout and the siren-like figure somehow told me this is the beer for you . a bit freaky , but i went with it . i was impressed from the very first pour . like any stout , the color is a dark molasses black . but ... the head was thick and dense with good retention . the coffee aroma was intense ! the roasted goodness almost overwhelms my sense of smell .the roasted coffee flavors are the first things that i could taste along with hints of chocolate . however , i can tell there 's more complexity here than my palette can decipher . the coffee flavors bring bitterness but it 's not over powering as the sweetness of the malt cuts the bitterness quite nicely the beer has carbonation but once the bubbles have escaped the beer gives a creamy , velvety feel and finish . the alcohol was very well hidden in this beer which is scary ... |

| Methods | Appearance | | | |
|---|---|---|---|---|
| | S | P | R | F1 |
| AGR | 20 | 83.7 | 87.5 | 85.6 |
| -w/o *causal.* | 20 | 81.5 | 87.8 | 84.5 |
| -w/o *embedd.* | 20 | 81.9 | 86.9 | 84.3 |
| -w/o *both* | 20 | 74.3 | 85.2 | 79.4 |

Table 5: Ablation studies on the BeerAdvocate.

across all multi-aspects datasets consistently.

**Ablation Studies.** To further verify the effectiveness of AGR, we conduct the ablation experiments. As depicted in Table 5, removing either the optimized objective of causal effectiveness (referred to as *causal.*), the pretrained agent embedding (referred to as *embedd.*), or *both*, results in a notable decline in AGR's performance, underscoring the critical roles played by our proposed key components in AGR method.

**Further Analyses.** Firstly, we compare AGR with FR and MGR, providing the visualized examples. For example, we can observe from Table 4 that although all three methods are able to focus on the appearance aspect, FR and MGR still exhibit some degeneration (since the selective rationale still has some distance from the target prediction). However, AGR utilizes causal calculus to capture the causal variations between $Z_1$ and $Z_2$, as well as between $Z_2$ and $Z_3$, regarding the target prediction, thereby gradually mitigating this degeneration during the training process. The Appendix C presents more visualized examples. Secondly, similar to (Liu et al., 2023b), we also compare the complexity of AGR with other models. As shown in Table 6, we can see that the complexity of AGR has been somewhat improved compared to latest work; however, there is still room for further improvement. This will be a key focus of future research.

| | RNP | FR | AGR | CAR |
|---|---|---|---|---|
| modules<br>parameters | 1gen+1pred<br>2× | 1gen+1pred<br>2× | 1gen+1pred+1agent<br>**3×** | 1gen+2pred<br>3× |
| | DARE | CAR | DMR | MGR |
| modules<br>parameters | 1gen+1pred+guider<br>3× | 1gen+2pred<br>3× | 1gen+3pred<br>4× | 3gen+1pred<br>4× |

Table 6: The complexity of different models. "gen": generator. "pred": predictor.

## 6 Conclusion

In this paper, we propose AGR, a reinforced causal agent-based rationalization approach to guide the cooperative game optimization during rationale training. Our theoretical insights underscore the necessity of this guidance signal for accurate predictions. Empirical evaluations on two widely-used benchmarks indicate the effectiveness of our proposed approach, surpassing existing state-of-the-art methods for selective rationalization.

## Limitations

There are still some limitations that need further improvement in the future. Firstly, optimizing cooperative game of rationalization during training brings great significance to the model performance, but how to more efficiently search for meaningful actions within a larger search space for good rationales remains the next direction to explore. Nextly, this work does not involve the debiasing techniques of data-level. Considering the debiasing technique may be a good way to further improve the results. In addition, as the latest research (Chen et al., 2022; Liu et al., 2023a,b) has shown that it is still a challenging task to finetune pretrained language models on the cooperative game framework. Therefore, how to incorporate the cooperative framework and (large) language models is a research interest.

## Ethics Statement

This paper does not involve the presentation of a new dataset and the utilization of demographic or identity characteristics information.

## Acknowledgements

We would like to thank all the anonymous reviewers for their valuable feedback. This work was supported by the National Natural Science Foundation of China (Nos.62376144, 62272285, 62076155) and the Science and Technology Cooperation and Exchange Special Project of Shanxi Province (No.202204041101016).

## References

Yujia Bao, Shiyu Chang, Mo Yu, and Regina Barzilay. 2018. Deriving machine attention from human rationales. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1903–1913, Brussels, Belgium. Association for Computational Linguistics.

Jasmijn Bastings, Wilker Aziz, and Ivan Titov. 2019. Interpretable neural predictions with differentiable binary variables. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2963–2977, Florence, Italy. Association for Computational Linguistics.

Shiyu Chang, Yang Zhang, Mo Yu, and Tommi Jaakkola. 2019. A game theoretic approach to class-wise selective rationalization. *Advances in neural information processing systems*, 32.

Shiyu Chang, Yang Zhang, Mo Yu, and Tommi Jaakkola. 2020. Invariant rationalization. In *International Conference on Machine Learning*, pages 1448–1458. PMLR.

Howard Chen, Jacqueline He, Karthik Narasimhan, and Danqi Chen. 2022. Can rationalization improve robustness? In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3792–3805, Seattle, United States. Association for Computational Linguistics.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Nicola De Cao, Michael Sejr Schlichtkrull, Wilker Aziz, and Ivan Titov. 2020. How do decisions emerge across layers in neural models? interpretation with differentiable masking. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3243–3255, Online. Association for Computational Linguistics.

Zhiying Deng, Jianjun Li, Zhiqiang Guo, and Guohui Li. 2023. Multi-aspect interest neighbor-augmented network for next-basket recommendation. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Yongfeng Huang, Yujun Chen, Yulun Du, and Zhilin Yang. 2021. Distribution matching for rationalization. In *AAAI Conference on Artificial Intelligence*.

Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, San Diega, CA, USA.

Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2016. Rationalizing neural predictions. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 107–117, Austin, Texas. Association for Computational Linguistics.

Wei Liu, Haozhao Wang, Jun Wang, Zhiying Deng, YuanKai Zhang, Cheng Wang, and Ruixuan Li. 2023a. Enhancing the rationale-input alignment for self-explaining rationalization. *arXiv preprint arXiv:2312.04103*.

Wei Liu, Haozhao Wang, Jun Wang, Ruixuan Li, Xinyang Li, YuanKai Zhang, and Yang Qiu. 2023b. MGR: Multi-generator based rationalization. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12771–12787, Toronto, Canada. Association for Computational Linguistics.

Wei Liu, Haozhao Wang, Jun Wang, Ruixuan Li, Chao Yue, and YuanKai Zhang. 2022. Fr: Folded rationalization with a unified encoder. *Advances in Neural Information Processing Systems*, 35:6954–6966.

Wei Liu, Jun Wang, Haozhao Wang, Ruixuan Li, Zhiying Deng, YuanKai Zhang, and Yang Qiu. 2023c. D-separation for causal self-explanation. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Wei Liu, Jun Wang, Haozhao Wang, Ruixuan Li, Yang Qiu, Yuankai Zhang, Jie Han, and Yixiong Zou. 2023d. Decoupled rationalization with asymmetric learning rates: A flexible lipschitz restraint. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1535–1547.

Julian McAuley, Jure Leskovec, and Dan Jurafsky. 2012. Learning attitudes and attributes from multi-aspect reviews. *2012 IEEE 12th International Conference on Data Mining*, pages 1020–1025.

Bhargavi Paranjape, Mandar Joshi, John Thickstun, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2020. An information bottleneck approach for controlling conciseness in rationale extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 1938–1952, Online. Association for Computational Linguistics.

Judea Pearl. 2009. *Causality*. Cambridge university press.

Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. 1999. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12.

Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, page 783–792, New York, NY, USA. Association for Computing Machinery.

Mo Yu, Shiyu Chang, Yang Zhang, and Tommi S Jaakkola. 2019. Rethinking cooperative rationalization: Introspective extraction and complement control. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.

Mo Yu, Yang Zhang, Shiyu Chang, and Tommi Jaakkola. 2021. Understanding interlocking dynamics of cooperative rationalization. *Advances in Neural Information Processing Systems*, 34:12822–12835.

Hao Yuan, Lei Cai, Xia Hu, Jie Wang, and Shuiwang Ji. 2020. Interpreting image classifiers by generating discrete masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4).

Linan Yue, Qi Liu, Yichao Du, Yanqing An, Li Wang, and Enhong Chen. 2022. Dare: Disentanglement-augmented rationale extraction. *Advances in Neural Information Processing Systems*, 35:26603–26617.

Linan Yue, Qi Liu, Li Wang, Yanqing An, Yichao Du, and Zhenya Huang. 2023. Interventional rationalization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11404–11418, Singapore. Association for Computational Linguistics.

Wenbo Zhang, Tong Wu, Yunlong Wang, Yong Cai, and Hengrui Cai. 2023. Towards trustworthy explanation: on causal rationalization. In *Proceedings of the 40th International Conference on Machine Learning*. JMLR.org.

## A Proof of Lemma 1

Given random variables $X$, $Z$, $Y$, and $\mathcal{A}$, where $\mathcal{A}$ is drawn from the distribution of $X$. According to Section 2, to obtain a good predictor, we have

$$\min_{\theta_g,\theta_p} \mathcal{H}(Y,\hat{Y}) = \min_{\theta_g,\theta_p} \mathcal{H}(Y, f_P(Z)), \tag{15}$$

where $Z = f_G(X)$. It means that we need to minimize $H(Y,Z)$ (Liu et al., 2023b), i.e., to reduce more uncertainty and indicate the label $Y$. We assume that exist variable $\mathcal{A}$ could make to reduce the uncertainty of learning $Y$, then our goal is to make $H(Y,\mathcal{A}) \le H(Y,Z)$.

According to the mutual information formula, we can obtain:

$$H(Y) - H(Y,\mathcal{A}) \ge H(Y) - H(Y,Z), \tag{16}$$

so,

$$I(Y,\mathcal{A}) \ge I(Y,Z). \tag{17}$$

Next, since we have $X = \{Z, X\backslash Z\}$ where $X\backslash Z$ denotes the text derived from $X$ and unrelated to the rationale, so we can obtain mutual information between $X$ and $Y$,

$$\begin{aligned} I(Y;X) &= I(Y;\{Z, X\backslash Z\}) \\ &= I(Y;Z) + I(Y;X\backslash Z|Z) \end{aligned} \tag{18}$$

According to the non-negativity of mutual information, we have $I(Y;X\backslash Z|Z) \ge 0$, so

$$I(Y,X) \ge I(Y,Z) \tag{19}$$

Further, we denote $I(Y,X) = \varepsilon_0 \ge \varepsilon_1 \ge I(Y,Z) \ge \varepsilon_2$, where $\varepsilon_1$ and $\varepsilon_2$ indicate the upper and lower bounds of $I(Y,Z)$, respectively.

Therefore, we can obtain that when $\mathcal{A} = X$,the equation $I(Y,\mathcal{A}) = \varepsilon_0 \ge \varepsilon_1 \ge I(Y,Z)$ is satisfied. That is to say, a solution for $\mathcal{A}$ exists, and $X$ is a solution of $\mathcal{A}$.

The proof of Lemma 1 is completed.

## B Experiment Details

### B.1 Baselines

We compare AGR with the following baselines:
**RNP (2016)**, a original RNP sampling method.
**HardKuma (2019)**, a kumaraswamy-distribution-based sampling method.
**CAR (2019)**, a game theoretic-based approach to class-dependent rationalization.
**Information Bottleneck (IB) (2020)**, a model utilizing IB objective for balancing performance and rationale length.
**INVRAT (2020)**, a method that introduces an environment-agnostic predictor.

| Datasets | | Train | | Dev | | Annotation | |
|---|---|---|---|---|---|---|---|
| | | Pos | Neg | Pos | Neg | Pos | Neg |
| BeerAdvocate | Appearance | 202385 | 12897 | 28488 | 1318 | 923 | 13 |
| | Aroma | 172299 | 30564 | 24494 | 3396 | 848 | 29 |
| | Palate | 176038 | 27639 | 24837 | 3203 | 785 | 20 |
| HotelReview | Location | 7236 | 7236 | 906 | 906 | 104 | 96 |
| | Service | 50742 | 50742 | 6344 | 6344 | 101 | 99 |
| | Cleanliness | 75049 | 75049 | 9382 | 9382 | 99 | 101 |

Table 7: Statistics of datasets used in this paper.

**DMR (2021)**, which proposes a teacher-student distillation framework to align input distribution.
**A2R (2021)**, a method that introducing a soft rationale to predictor.
**DARE (2022)**, which introduces a guider into predictor to encapsulate more information from the input.
**FR (2022)**, a method using a unified encoder for generator and predictor.
**Inter-RAT (2023)**, which develops an interventional rationalization to discover the causal rationales.
**MGR (2023b)**, a method leveraging multiple generators to select rationales.

### B.2 Datasets

Following previous research (Huang et al., 2021; Yue et al., 2023; Liu et al., 2023b), we obtain BeerAdvocate and HotelReview datasets. BeerAdvocate (McAuley et al., 2012) and HotelReview (Wang et al., 2010) are publicly available from existing work. As shown in Table 7, the specific splitting details of the two datasets are presented.

### B.3 Implementation

To fairly compare with previous works and validate the effectiveness of the approach proposed, we utilize the 100-dimension Glove (Pennington et al., 2014) as the word embedding and the 200-dimension GRUs (Cho et al., 2014) encoder to build the generator $f_G(\cdot)$ in the AGR architecture. Further generator $f_G(\cdot)$ follows Equation 1 for cooperative optimization with predictor $f_P(\cdot)$. Meanwhile, we construct the policy network $q_\phi(\cdot)$ to collaborate with the generator $f_G(\cdot)$ and predictor $f_P(\cdot)$ to learn candidate actions in different training states, including the representation learning of action candidates and the sampling of actions. We use Adam (Kingma and Ba, 2015) as the optimizer.

## C Additional Examples

As shown in Table 8, we provide more examples of selected rationale from the *Beer-Aroma* and *Hotel-Location* two aspects, where their sparsity is set to be about 20% and 10%, respectively.

Table 8: Examples of generated rationales. Human-annotated rationales are underlined. Rationales from three models are highlighted in blue, respectively.

| FR (2022) | MGR (2023b) | AGR (Ours) |
|---|---|---|
| **Aspect:** Beer-Aroma<br>**Label:** Positive, **Pred:** Positive<br>**Text:** had this at bocktown with wvbeergeek and jasonm , came in a 750ml caged and corked the corked banged out of sight as soon as the cage was undone .seved into a tulip glass between the 3 of us hazy , deep copper , mahagony , hard to get a really good look at the color at bocktown . off white head hard to pour without a glass full of fluffy everlasting head . left lot of thick webbing all over the inside of the glass , sticky looking . great aroma ca n't seem to keep it away from the nose . sweet , dark , tart fruit notes , some sour cherry , earthy , spicy , with hints of currants , clove , allspice also nutty , with some belgium yeast . lots of sweet booziness from the start , vinious , dark fruityness with plum notes . the fruittyness was remisent of dried fruit.lots of spicyness lots of clove.also nutty and earthy . finished clean , spicy and very sugary . syrupy , big full mouthfeel , smooth and very creamy with lots of juicyness . a beer to sip , but very enjoyable , wish i had the whole bottle to drink would be no problem . a must try beer if you like this style . seems like a beer that would age very well . | **Aspect:** Beer-Aroma<br>**Label:** Positive, **Pred:** Positive<br>**Text:** had this at bocktown with wvbeergeek and jasonm , came in a 750ml caged and corked the corked banged out of sight as soon as the cage was undone . seved into a tulip glass between the 3 of us hazy , deep copper , mahagony , hard to get a really good look at the color at bocktown . off white head hard to pour without a glass full of fluffy everlasting head . left lot of thick webbing all over the inside of the glass , sticky looking . great aroma ca n't seem to keep it away from the nose . sweet , dark , tart fruit notes , some sour cherry , earthy , spicy , with hints of currants , clove , allspice also nutty , with some belgium yeast . lots of sweet booziness from the start , vinious , dark fruityness with plum notes . the fruittyness was remisent of dried fruit.lots of spicyness lots of clove.also nutty and earthy . finished clean , spicy and very sugary . syrupy , big full mouthfeel , smooth and very creamy with lots of juicyness . a beer to sip , but very enjoyable , wish i had the whole bottle to drink would be no problem . a must try beer if you like this style . seems like a beer that would age very well . | **Aspect:** Beer-Aroma<br>**Label:** Positive, **Pred:** Positive<br>**Text:** had this at bocktown with wvbeergeek and jasonm , came in a 750ml caged and corked the corked banged out of sight as soon as the cage was undone . .seved into a tulip glass between the 3 of us hazy , deep copper , mahagony , hard to get a really good look at the color at bocktown . off white head hard to pour without a glass full of fluffy everlasting head . left lot of thick webbing all over the inside of the glass , sticky looking . great aroma ca n't seem to keep it away from the nose . sweet , dark , tart fruit notes , some sour cherry , earthy , spicy , with hints of currants , clove , allspice also nutty , with some belgium yeast . lots of sweet booziness from the start , vinious , dark fruityness with plum notes . the fruittyness was remisent of dried fruit.lots of spicyness lots of clove.also nutty and earthy . finished clean , spicy and very sugary . syrupy , big full mouthfeel , smooth and very creamy with lots of juicyness . a beer to sip , but very enjoyable , wish i had the whole bottle to drink would be no problem . a must try beer if you like this style . seems like a beer that would age very well . |
| **Aspect:** Hotel-Location<br>**Label:** Negative, **Pred:** Negative<br>**Text:** we stayed at the dona palace for 3 nights and while the location is central , it is also more crowded and noisy . the windows of the room we stayed in did not have adequate sound proofing , noise from the canal and outside would wake us up early in the morning . the breakfast was a nice bonus though , the two waitresses serving the room were always gracious and helpful . the front desk personnel however were rude and abrupt , so that was n't pleasant to deal with . the rooms are dated and had a musty smell . the bed was uncomfortable , blankets were rough , and the shower drain did not work very well . overall , i probably wound not stay here again . | **Aspect:** Hotel-Location<br>**Label:** Negative, **Pred:** Negative<br>**Text:** we stayed at the dona palace for 3 nights and while the location is central , it is also more crowded and noisy . the windows of the room we stayed in did not have adequate sound proofing , noise from the canal and outside would wake us up early in the morning . the breakfast was a nice bonus though , the two waitresses serving the room were always gracious and helpful . the front desk personnel however were rude and abrupt , so that was n't pleasant to deal with . the rooms are dated and had a musty smell . the bed was uncomfortable , blankets were rough , and the shower drain did not work very well . overall , i probably wound not stay here again . | **Aspect:** Hotel-Location<br>**Label:** Negative, **Pred:** Negative<br>**Text:** we stayed at the dona palace for 3 nights and while the location is central , it is also more crowded and noisy . the windows of the room we stayed in did not have adequate sound proofing , noise from the canal and outside would wake us up early in the morning . the breakfast was a nice bonus though , the two waitresses serving the room were always gracious and helpful . the front desk personnel however were rude and abrupt , so that was n't pleasant to deal with . the rooms are dated and had a musty smell . the bed was uncomfortable , blankets were rough , and the shower drain did not work very well . overall , i probably wound not stay here again . |