

ILAEDA: An Imitation Learning Based Approach for Automatic Exploratory Data Analysis

Abhijit Manatkar

abhijit.manatkar@students.iiit.ac.in
Machine Learning Lab, KCIS, IIIT Hyderabad
Hyderabad, Telangana, India

Hima Patel

himapatel@in.ibm.com
IBM Research, India
Bangalore, Karnataka, India

Devarsh Patel

devarsh.patel@students.iiserpune.ac.in
IISER, Pune
Pune, Maharashtra, India

Naresh Manwani

naresh.manwani@iiit.ac.in
Machine Learning Lab, KCIS, IIIT Hyderabad
Hyderabad, Telangana, India

ABSTRACT

Automating end-to-end Exploratory Data Analysis (AutoEDA) is a challenging open problem, often tackled through Reinforcement Learning (RL) by learning to predict a sequence of analysis operations (FILTER, GROUP, etc). **Defining rewards for each operation is a challenging task and existing methods rely on various *interestingness measures* to craft reward functions to capture the importance of each operation.** In this work, we argue that not all of the essential features of what makes an operation important can be accurately captured mathematically using rewards. **We propose an AutoEDA model trained through imitation learning from expert EDA sessions, bypassing the need for manually defined interestingness measures.** Our method, based on generative adversarial imitation learning (GAIL), generalizes well across datasets, even with limited expert data. We also introduce a novel approach for generating synthetic EDA demonstrations for training. Our method outperforms the existing state-of-the-art end-to-end EDA approach on benchmarks by upto 3x, showing strong performance and generalization, while naturally capturing diverse interestingness measures in generated EDA sessions.

CCS CONCEPTS

• **Computing methodologies** → **Reinforcement learning; Learning from demonstrations.**

KEYWORDS

Exploratory Data Analysis, Imitation Learning, Interestingness

ACM Reference Format:

Abhijit Manatkar, Devarsh Patel, Hima Patel, and Naresh Manwani. 2024. ILAEDA: An Imitation Learning Based Approach for Automatic Exploratory Data Analysis. In *The Fourth International Conference on Artificial Intelligence and Machine Learning Systems (AIMLSys 2024)*, October 08–11, 2024.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).
AIML Systems 2024, October 08–11, 2024, Baton Rouge

© 2024 Association for Computing Machinery.
ACM ISBN 979-8-4007-1161-9/24/10...\$15.00
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

Louisiana State University, Baton Rouge, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Exploratory Data Analysis (EDA) stands as a pivotal step within any data science pipeline, offering an effective means to comprehend datasets by uncovering patterns, anomalies, and critical insights. Traditionally, this process has heavily relied on human intuition, leading to a trial-and-error exploration guided by the analyst's domain expertise. However, recent advancements have sought to automate this process through AutoEDA methods [3, 18, 23, 27, 30], aiming to provide analysts with initial insights into datasets automatically, which serve as a springboard for further exploration.

Some of these efforts have sought to automate the entire EDA process using deep reinforcement learning (RL)-based AutoEDA systems [3] [21]. These systems treat AutoEDA as a sequential decision-making problem, with RL algorithms trained to predict a sequence of EDA operations. The rewards for such RL systems are heuristics/rules. However, such approaches face significant challenges:

- (1) Defining appropriate reward functions that encapsulate all the relevant aspects of a dataset is inherently complex. Often, the reward function used is insufficient to capture all the interesting dimensions of the data.
- (2) For a given dataset, previous expert data analysis sessions provide major hints about the usefulness of certain reward functions. One has to analyze these expert sessions to get those hints.
- (3) Precise dataset-specific reward function definitions are required to learn meaningful policies in RL-based systems like the one presented by [3]. Defining detailed dataset-specific reward functions requires a good deal of data analysis to already have been done, a requirement which contradicts the motivations behind building an AutoEDA system.

To address these challenges, we propose **ILAEDA**, an imitation learning-based framework for AutoEDA. Unlike reward-based methods, ILAEDA learns directly from expert EDA sessions, eliminating the need for handcrafted reward functions and dataset-specific analyses. Our approach learns to imitate a human when shown a few expert sessions. To make sure that we don't overfit the model to the available expert sessions, we introduce a novel approach to automatically generate synthetic EDA demonstrations

that are used for training. This helps our system to offer better generalizability across datasets along with the ability to capture diverse aspects of effective EDA sessions, such as diversity and coherence. **Key contributions** of our work include:

- (1) Introducing ILAEDA, an imitation learning framework for AutoEDA, which removes the reliance on handcrafted rewards and dataset-specific analyses.
- (2) Showcasing ILAEDA’s enhanced generalizability across datasets and its capacity to capture various facets of effective EDA sessions. As our approach trains a model from expert demonstrations, a diverse set of demonstrations allows for generalization across different datasets.
- (3) Providing a comprehensive analysis of our results, demonstrating ILAEDA’s effectiveness even with minimal expert demonstrations. Experimental results show that the sessions generated by our model automatically capture multiple different notions of what makes for a good EDA session (diversity, coherence, readability, etc.) without using reward scores.

2 RELATED WORK

2.1 AutoEDA Methods

In recent years, several methods have attempted to automate different aspects of the EDA process. A set of next-step recommendation systems make use of pre-defined notions of interestingness to quantify the output of a specific EDA operation, e.g. for data-visualization [35], for finding interesting data tuple-subsets or data cube subsets [27] [26] [8] [23], OLAP drill-down [13], and data summaries [29]. Log-based EDA recommender systems, such as [1] [9] [36], use a collection of the previous exploratory sessions of the same or different users to generate recommendations for the next exploratory operation. Systems in [18] [19] utilize session logs to retrieve similar prefixes, generate candidate *next-actions*, and convert them into concrete recommendations of EDA operations based on interestingness measures. The approach in [30] recommends next action using a kNN-based classifier on user exploration logs. This classifier predicts the best interestingness measure that could capture user’s interest at every step of the EDA session. EDA can also be modeled as a meta-learning problem [4].

The approach in [6] generates comparison queries to highlight insights in a dataset. In [16], authors propose a novel formulation of basic data patterns to facilitate extraction of insights from a dataset. [17] present a Large Language Model [33] based system for extracting insights from a dataset and presenting them to the user through a natural language interface.

RL based AutoEDA systems are described in [3] [22] [15]. ATENA [3] automatically generates EDA sessions for a given tabular dataset. In this approach, EDA is formulated as a sequential decision-making problem. A deep reinforcement learning algorithm is used to learn a policy to generate a sequence of EDA operations which maximize the defined reward function. Authors of [15] describe a system built with ATENA as a base, augmenting it with a novel action space formulation and user-specified constraint compliance rewards.

2.2 Imitation Learning

In this setup, we do not have access to the reward structure. We want to learn the optimal policy without knowing the rewards.

Behavioral cloning [32] [25], Imitation learning [7][12] [24] [10], Inverse RL [37] are some of the techniques which are used to learn an optimal policy when we do not know the reward structure. All these approaches require expert trajectories to learn an optimal policy.

Generative Adversarial Imitation Learning (GAIL) [12] treats imitation learning as a min-max adversarial problem. The method describes two modules - a discriminator D and a policy π with opposing objectives:

- (1) Given any state s , the objective of the policy is model the distribution $\pi(a|s)$ to mimic the expert demonstrations as closely as possible.
- (2) Given a pair of state and action taken in the state (s, a) , the objective of the discriminator is to differentiate between (s, a) pairs coming from an expert session and the policy. For any (s, a) , $D(s, a) \in [0, 1]$, where $D(s, a) = 1$ indicates that (s, a) comes from expert demonstrations and $D(s, a) = 0$ indicates that it is a generated pair.

The two modules are optimized in a procedure that alternates between optimizing D and optimizing π until D is no longer able to distinguish between state-action pairs coming from expert demonstrations and state-action pairs coming from π . GAIL has been shown to effectively imitate expert demonstrations even when the number of demonstrations is very low, which makes it an attractive candidate for the end-to-end AutoEDA problem as the availability of expert demonstrations is generally quite low.

Action	Measure
GROUP highest_layer AGGREGATE COUNT packet_number	A-INT, Diversity
BACK	-
GROUP eth_src AGGREGATE COUNT packet_number	A-INT, Readability
GROUP ip_src AGGREGATE COUNT packet_number	Peculiarity
FILTER info_line CONTAINS Echo (ping) reply	Diversity, Coherence
BACK	-
BACK	-
BACK	-
FILTER highest_layer NEQ ICMP	Coherence
GROUP tcp_srcport AGGREGATE COUNT packet_number	Diversity
GROUP ip_src AGGREGATE COUNT packet_number	Coherence, Peculiarity
FILTER ip_src NEQ 192.168.1.122	Coherence

Table 1: Expert EDA session and the interestingness measure maximized by each operation in the analysis sequence. Measures with relative score greater than 0.8 are mentioned

2.3 Interestingness Measures

Various interestingness measures have been studied in literature. Some of them have been used as rewards in RL based systems and they are also useful for analyzing EDA operations. We present a brief summary of some of the interestingness measures studied in literature, which we use in later sections.¹ (a) **A-INT**. This interestingness score [3] is used as a reward function that rewards display which are compact in case of GROUP operations, and displays with high deviation from the previous display in case of FILTER operations. (b) **Diversity score**. This metric favors a display that highlights parts of the dataset that are different from those seen in any of the previous displays in the session so far [3]. (c) **Readability score**. This metric rewards highly compact displays [5] with few rows as they are considered more readable and are given a higher reward. (d) **Peculiarity score**. This score helps quantify anomalous patterns by favoring display with a high difference from the initial display at the start of the analysis [35]. (e) **Coherence score**. This score is a highly detailed metric determined by a set of handcrafted rules which assign each display a penalty or reward based on whether the operation performed at the current step is coherent with previous operations [3].

3 PROBLEM SETUP

Our proposed approach attempts to tackle the end-to-end AutoEDA problem using imitation learning. In this section, we describe the problem setup and illustrate some of the challenges with reward-based approaches for this problem.

The input is a tabular dataset \mathcal{D} with a set of attributes (columns) \mathbf{A} and the desired output is a sequence of analysis operations which will form an *EDA session*. This problem can be modeled as a Markov Decision Process $\langle \mathcal{S}, \mathcal{A}, R, \mathcal{P} \rangle$ where \mathcal{S}, \mathcal{A} are the state and action space respectively, $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function, and $\mathcal{P} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ is a deterministic transition function. Let Π be the space of all stationary stochastic policies which return a distribution over actions in \mathcal{A} given a state in \mathcal{S} . The goal is to learn an optimal policy $\pi^* \in \Pi$ which maximizes a discounted cumulative sum of rewards for a fixed horizon T , i.e. $\pi^* = \arg \max_{\pi} \mathbb{E}_{\pi} [\sum_{t=0}^T \gamma^t R(s_t, a_t)]$, where $s_t \in \mathcal{S}$ is the state at time-step t given by $s_t = \mathcal{P}(s_{t-1}, a_{t-1})$, $a_t \sim \pi(\cdot | s_t)$ and $\gamma \in [0, 1]$ is the discounting factor.

3.1 State Space

State space [3] is described as follows. An EDA session begins with an initial display d_0 of the dataset \mathcal{D} . At each subsequent timestep t , an action $a_t \in \mathcal{A}$, which is an EDA operation, is applied to display d_t to obtain display d_{t+1} . For each display d_t , corresponding state $s_t \in \mathcal{S}$ is represented as $s_t = \text{Encode}(d_{t-2}) \oplus \text{Encode}(d_{t-1}) \oplus \text{Encode}(d_t)$, where \oplus is the vector concatenation operator and $\text{Encode}(d_t)$ encodes a display into a vector by extracting the following features. (a) For each attribute $A \in \mathbf{A}$ in the display, we include its entropy, number of distinct values, and the number of null values. (b) For each attribute $A \in \mathbf{A}$, we include a feature to determine whether that column is grouped, aggregated, or neither. (c) Three global features: number of groups, the mean

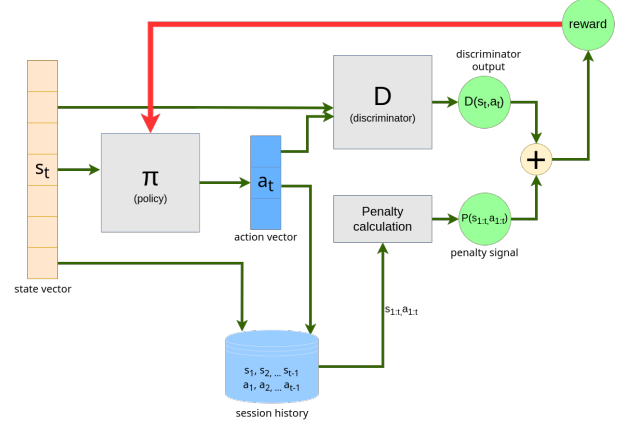


Figure 1: ILAEDA System Architecture

size of groups, and variance of size of groups. If less than three displays have been seen, $\text{Encode}(d_{t-2})$ and $\text{Encode}(d_{t-1})$ are replaced with 0 vectors of appropriate sizes as required.

3.2 Action Space

The action space \mathcal{A} , in addition to basic BACK and STOP actions for directing the flow of the analysis, consists of parameterized FILTER and GROUP operations with possible parameters drawn from the columns of the dataset \mathbf{A} and their distributions as described below:

- (1) **GROUP**(grp_col, agg_col, agg_func): This action groups on the grp_col $\in \mathbf{A}$ and aggregates the agg_col $\in \mathbf{A}$ using the agg_func, which can be one of {SUM, COUNT, MEAN, MIN, MAX}.
- (2) **FILTER**(filter_col, filter_func, filter_term): This action filters the dataset by using the comparison operator filter_func $\in \{=, \neq, \text{CONTAINS}, \text{STARTS_WITH}, \text{ENDS_WITH}\}$ and filter_term, which (numerical or textual) appears from within the filter_col $\in \mathbf{A}$.
- (3) **BACK**(): Allows to take a step backward and return to the previous display in the ongoing analysis session.
- (4) **STOP**(): Allows the signaling of the end of the current analysis session.

3.3 Issues With Reward-Based Modeling

ATENA [3] uses hand-crafted rewards based on interestingness measures. The reward function used is given by:

$$R(s_t, a_t) = r_{int}(s_t, a_t) + \lambda_1 r_{div}(s_t, a_t) + \lambda_2 r_{coh}(s_t, a_t) \quad (1)$$

Here, r_{int} , r_{div} are the A-INT and Diversity score respectively (refer Section 2). r_{coh} is the output of a binary classifier trained using rules based on the Coherence score. λ_1, λ_2 are constants that are tuned to calibrate the reward. As described before, some of the rules defined for the Coherence reward include:

- (1) Filtering/Grouping on a specific set of filterable or groupable columns is rewarded, and specific non-filterable or non-groupable columns are penalized.
- (2) Using a specific set of operators to filter the info_line column is rewarded.

¹For a detailed description, refer to Appendix A

- (3) Grouping on column highest_layer on a filtered display is penalized.

Evidently, these rules are highly dataset-specific and require a great deal of familiarity with the dataset. This is one major drawback of reward based model.

To understand further shortcomings of reward-based modeling, we analyze an expert EDA session using the interestingness measures described in Section 2. Table 1 shows sequence of operations from an EDA session performed by an expert on the dataset mentioned above. At each step of the session, we calculate the scores of interestingness measures and examine which of them are optimized at each step. The scores are normalized [30] in the range of [0, 1] for comparison. At every step, measures with a normalized score exceeding 0.8 are shown. We observe that:

- (1) Within the same expert session, every step optimizes for different combinations of interestingness measures.
- (2) Some actions optimize the Peculiarity and Readability scores, which are not part of the reward definition in 1.

Consistent observations across multiple expert sessions reveal the notable limitation that handcrafted rewards often fail to capture the full spectrum of intriguing data aspects. Moreover, our analysis suggests that this limitation becomes more pronounced when considering a broader array of measures. These identified drawbacks serve as a compelling motivation for us to propose an alternative approach to learning an optimal policy for ILAEDA, which aims to mitigate these issues.

4 ILAEDA: PROPOSED METHODOLOGY

The overall architecture of the proposed approach ILAEDA is provided in Figure 1. Our main goal is to learn a policy directly from the expert EDA sessions (from here on, also referred to as *expert trajectories*) using Imitation Learning, allowing the system to learn to imitate a human expert. We have a set of expert trajectories $\mathcal{E} = \{\tau_1, \tau_2, \dots, \tau_N\}$ where each τ_i is a sequence of state-action pairs $\{(s_1, a_1), (s_2, a_2), \dots, (s_T, a_T)\}$ where each $s_i \in \mathcal{S}$ and $a_i \in \mathcal{A}$. We assume that expert actions come from an underlying expert policy π_E , i.e. $a_t \sim \pi_E(\cdot|s_t)$, $\forall (s_t, a_t) \in \tau$, $\forall \tau \in \mathcal{E}$. ILAEDA uses three neural networks: Policy π_θ , Value V_ϕ and Discriminator D_w parameterized by θ, ϕ, w respectively. The policy and discriminator networks undergo training at alternate steps. The discriminator is trained to minimize the loss:

$$L_D = \mathbb{E}_{\pi_\theta} [\log(D_w(s, a))] + \mathbb{E}_{\pi_E} [\log(1 - D_w(s, a))] \quad (2)$$

In the first term, the expectation is over a mini-batch of (s, a) pairs collected by environment interaction using the current policy π_θ . A set of trajectories $\mathcal{T} = \{\tau_1^\theta, \dots, \tau_N^\theta\}$ is obtained where each $\tau_i^\theta = \{(s_1, a_1), \dots, (s_T, a_T)\}$ consists of (s, a) pairs such that action $a \sim \pi_\theta(\cdot|s)$. In the second term, the expectation is over an equal sized mini-batch of (s, a) pairs randomly picked from $\tau \in \mathcal{E}$. The policy and value networks are trained using Proximal Policy Optimization (PPO) [28] on a mini-batch of 4-tuples (r_t, s_t, a_t, s_{t+1}) where $r_t = -\log(1 - D_w(s_t, a_t)) + P(s_{1:t}, a_{1:t})$. The first term is the imitation reward, based on the discriminator output as in GAIL [12]. The second term is our penalty term that we describe below.

Half of the mini-batch is composed of expert demonstrations (s_t, a_t, s_{t+1}) obtained from $\tau \in \mathcal{E}$, each of which is appended with

a reward term r_t derived from the output of the discriminator as described above. The other half is composed of (s_t, a_t, s_{t+1}) obtained from $\tau^\theta \in \mathcal{T}$ with r_t appended as above. The supplementary material provides a detailed algorithmic description of the overall method. Three key aspects of training of ILAEDA are as follows.

4.1 Penalties for Incoherent Action Sequences

An important feature of ILAEDA is that it augments the reward term with a penalty term for incoherent sequences of actions. Some such incoherent sequence of actions are as follows. (a) Model takes a BACK action at the beginning of the analysis, (b) model consecutively repeats the same FILTER/GROUP action, (c) model alternates between a FILTER/GROUP action and BACK action.

A sequence of operations is typically helpful when it consists of at least 2-3 continuous FILTER/GROUP operations before the steps are retraced using BACK. Taking a single FILTER/GROUP action and immediately retracing with a BACK action, and continuing to do so repeatedly signals a lack of confidence in the chosen path, and hence an uncertainty in the actions taken. To discourage the learning of these behaviours, we use the following reward function,

$$r_t = -\log(1 - D_w(s_t, a_t)) + P(s_{1:t}, a_{1:t}) \quad (3)$$

where $P(s_{1:t}, a_{1:t})$ is an incoherence penalty defined as follows:

$$P(s_{1:t}, a_{1:t}) = \begin{cases} -1.0 & \text{if } a_t = \text{BACK and } t = 1 \\ -1.0 & \text{if } a_t \neq \text{BACK and } a_t = a_{t-1} \\ -1.0 \times l & \text{if } a_t = a_{t-2} \dots = a_{t-2l} = \text{BACK} \neq a_{t-2(l+1)} \text{ and } \\ & a_{t-1}, a_{t-3}, \dots, a_{t-2l-1} \in \{\text{FILTER, GROUP}\} \text{ and } l > 1 \\ 0.0 & \text{otherwise} \end{cases} \quad (4)$$

4.2 Initialization of Policy with Behavioral Cloning

We initialize the policy network using a behavioral cloning objective. We observe that such initialization results in stronger final performance of the learnt model. Our reasoning for this finding is that behavioural cloning provides a *warm start* to the policy network compared to random initialization of parameters. As very few expert demonstrations are available for imitation, pre-training with behavioural cloning is valuable in initializing the policy at a point in parameter space which is likely to converge to the optimum faster. The pre-training stage seeks to minimize the following objective:

$$\underset{\theta}{\text{minimize}} \mathbb{E}_{(s,a) \in \mathcal{E}} [-\log \pi_\theta(a|s) + \|\theta\|_2] \quad (5)$$

where (s, a) pairs come from trajectories $\tau_E \in \mathcal{E}$. We conduct an ablation study to examine the effects of this initialization choice.

4.3 Proximal Policy Optimization Updates

In ILAEDA, we train the policy network using Proximal Policy Optimization (PPO) [28]. Specifically, we use the PPO-clip update given as:

$$\theta_{t+1} = \arg \max_{\theta} \mathbb{E}_{T \sim \mathcal{B}} \left[\min \left(\frac{\pi_\theta(a|s)}{\pi_{\theta_t}(a|s)} A^t(T), g(\epsilon, A^t(T)) \right) \right] \quad (6)$$

Dataset	Method	Precision	TBLEU-1	TBLEU-2	TBLEU-3	EDA-Sim
CS 1	BC	0.2117	0.1421	0.0826	0.0519	0.2496
	ATENA	0.1855	0.1855	0.1377	0.0625	0.2704
	ILAEDA	0.3750	0.3333	0.2041	0.0841	0.2950
CS 2	BC	0.3287	0.1497	0.0726	0.0497	0.3491
	ATENA	0.2340	0.2325	0.1873	0.1182	0.2682
	ILAEDA	0.4000	0.2857	0.2182	0.1060	0.3900
CS 3	BC	0.1614	0.1235	0.0456	0.0393	0.2728
	ATENA	0.1153	0.1122	0.0550	0.0320	0.2462
	ILAEDA	0.1429	0.1314	0.0449	0.0359	0.3497
CS 4	BC	0.1950	0.1474	0.0771	0.0472	0.3378
	ATENA	0.1929	0.1929	0.1451	0.0708	0.3017
	ILAEDA	0.7500	0.3333	0.2041	0.0841	0.2051
SD 6	BC	0.9167	0.1428	0.1092	0.0612	0.2929
	ATENA	0.1111	0.0515	0.0173	0.0132	0.1539
	ILAEDA	0.4286	0.4286	0.1816	0.0669	0.5647
SD 7	BC	0.7071	0.2190	0.1520	0.0932	0.3505
	ATENA	0.1111	0.0429	0.0143	0.0114	0.1265
	ILAEDA	0.8333	0.5333	0.4781	0.3852	0.5536

Table 2: Comparison of ILAEDA with baselines on Cyber Security (CS 1-4) and Synthetic Datasets (SD 6-7)

Here π_{θ_t} is the policy at the t th time step. We wish to update parameters θ_t to θ_{t+1} such that expected reward is maximized while making sure that $\pi_{\theta_{t+1}}$ is not too far from π_{θ_t} . \mathcal{B} is a set of transitions $T = (s, a, r, s')$ sampled from environment interaction with policy π_{θ_t} . A^t is the advantage function given by:

$$A^t(T) = A^t(s, a, r, s') = r + V_{\phi_t}(s') - V_{\phi_t}(s) \quad (7)$$

where $a \sim \pi_{\theta_t}(\cdot|s)$ and V_{ϕ} is the value network. g is the clipping function and ϵ is the clipping factor used by PPO to ensure that updates do not change the policy too much. It is defined as:

$$g(\epsilon, A) = \begin{cases} (1 + \epsilon)A & A \geq 0 \\ (1 - \epsilon)A & A < 0 \end{cases} \quad (8)$$

5 EXPERIMENTAL EVALUATION

5.1 Dataset Details

To evaluate the efficacy of our proposed method, we train and generate EDA sessions on two types of datasets:

5.1.1 Cyber Security Datasets.

This is a collection of four mutually exclusive datasets [31], which we will refer to as the *cyber security datasets*. Expert EDA sessions have been generated in two ways for each of the four datasets: (1) **REACT Dataset:** 56 cyber security analysts were

asked to explore the four cyber security datasets. Each dataset may reveal a unique security event, and analysts were asked to discover the specifics of the underlying security event for each

dataset using as many as necessary actions. The REACT dataset is a collection of EDA session traces obtained from these experienced analysts' data exploration on the cyber security datasets and was curated by authors of [19]. (2) **Gold Dataset:** The gold-standard EDA sessions are generated from walkthrough documents from cyber-security experts, which guide the EDA process for viewers and highlight essential insights in the dataset. Authors curate this from [3]. Table 3 shows the number of EDA trajectories found for each of these datasets. We discuss in the following sections how we use these different trajectories for training and testing purposes.

5.1.2 Synthetic Datasets. We algorithmically generate datasets with a given schema and inject interesting patterns into them. The injected patterns are in the form of a dependencies between columns which induce a graph structure. We traverse this graph of dependencies to generate *expert* EDA trajectories. Five synthetic datasets, each with three categorical columns, two text columns and three numeric columns and 1000 rows, and their corresponding expert trajectories are generated for our experiments. The generated trajectories are randomly split into train and eval sets. Further details on the generation process are provided in Appendix C.

5.2 Baselines

- (1) **Behavioural Cloning:** Behavioural Cloning (BC) is an imitation learning technique that uses expert demonstrations to learn a state-to-action mapping in a supervised way. We use it as a baseline to compare against ILAEDA. For the Cyber Security datasets, we train a BC model on three out of four datasets and report results on sessions generated on the fourth dataset. We train a BC model on the first five tables for the Synthetic datasets and report results for sessions generated on tables 6 and 7.

Dataset	REACT	Gold
1	104	7
2	79	7
3	76	7
4	59	7

Table 3: Trajectory distribution for training across cyber security datasets

- (2) **ATENA**: We use ATENA [3] as a baseline to compare against our approach. We use the open-source code provided by the authors [2]. Since ATENA is an RL method that does not use labeled training data, we train and test ATENA models’ performance by generating and evaluating trajectories on the same datasets they have been trained on. We train an ATENA model for each Cyber Security dataset and datasets 6 and 7 of the synthetic datasets using default parameters from [2] and report scores by comparing against gold trajectories from each training dataset.

To test the generalization capabilities of ILAEDA, we set up our training and evaluation so that evaluation is performed only on datasets **not seen at training time**. We train our model for the Cyber Security datasets using a *leave one out* strategy. The model is trained using REACT and gold trajectories from three out of four datasets while keeping the remaining one as a test dataset. This model is then evaluated by generating trajectories on the test dataset and comparing these with the gold-standard trajectories for that dataset. This is repeated four times for each of the four Cyber Security datasets, and we calculate scores to evaluate performance on each left-out dataset. For the synthetic datasets, we train on all the trajectories from the first 5 of the 7 datasets. For evaluation, we generate trajectories on the remaining two datasets and report results by comparing them with the evaluation trajectories of the left-out datasets.

5.3 Evaluation Metrics

We evaluate the performance of ILAEDA and the baselines using the following similarity metrics introduced in [3] for evaluation.

- (1) **Precision**: Precision measures the accuracy of the generated EDA sessions and is calculated as the number of times a view occurs in the gold standard divided by the total number of views.
- (2) **TBLEU score**: This measure is adopted from BLEU [20] score. It is stricter than precision as it compares subsequences of size n (1 to 3) and considers the order and prevalence of each view in the gold-standard set. The benchmark measures TBLEU-1, TBLEU-2, and TBLEU-3.
- (3) **EDA-Sim**: EDA-Sim [19], considers the order of views and enables a fine-grained comparison of EDA views. It considers nearly identical views as *hits*. The generated session is compared to each gold-standard session to get the final EDA-Sim score, using the highest score.

5.4 Implementation Details

5.4.1 Architecture. ILAEDA’s policy and value networks are implemented as fully-connected neural networks with three hidden layers. There are 50 neurons in each layer with *tanh* non-linearities. The final output layer consists of multiple heads for predicting different components of the action. The design of this layer is identical to that of [3]. The discriminator network is a fully-connected neural network with two hidden layers of 32 neurons each, with ReLU non-linearities.

5.4.2 Training. The pre-training phase for ILAEDA is identical to the training for the Behavioural Cloning baseline. It consists of

training for 100 epochs with a learning rate of $1e-4$. A batch-size of 32 is used. The adversarial training phase of ILAEDA involves sampling a total of 100,000 interactions from the environment throughout the training, which corresponds to ~ 8000 episodes. A learning rate of $1e-6$ is used for all neural networks. A batch-size of 32 is used for training the policy and value networks, whereas the discriminator is trained with a batch-size of 192. We use a discounting factor of $\gamma = 0.99$. ATENA is trained using default parameters in [2].

6 RESULTS AND ANALYSIS

	Precision	TBLEU-1	TBLEU-2	TBLEU-3	EDA-sim
Without BC	0.259	0.103	0.038	0.029	0.195
Without penalty	0.413	0.239	0.144	0.058	0.258
ILAEDA (Ours)	0.417	0.271	0.168	0.078	0.310

Table 4: Ablation study scores for Cyber Security datasets. The low performance of ablated models validates our choice of penalties and pretraining using BC.

Table 2 shows a performance comparison between ILAEDA and the baselines on the evaluation metrics discussed in Section 5.3. We see that our method overall delivers a significantly better performance than the baselines. The results are all the more significant because, in each setting, ILAEDA is evaluated on datasets that it has not seen during training. In contrast, ATENA is evaluated on the same datasets it trains on. This indicates that ILAEDA generates better EDA sessions and has better generalization capability.

6.1 Ablation studies

6.1.1 Training without penalty scores: We trained a version of ILAEDA without the coherence penalties defined in equation 3 and kept all other configurations the same to study the impact of our design choice. We see that this model underperforms in comparison to ILAEDA trained with coherence penalties (Table 4). We explain the lower scores by inspecting the generated EDA session (Table 5). The EDA views 2 and 3 are identical because the same GROUP operators have been repeated consecutively. Such actions are penalized in our original model.

```

FILTER eth_src EQ 00:26:b9:2b:0b:59
GROUP ip_src AGGREGATE COUNT length
GROUP ip_src AGGREGATE COUNT length
BACK
GROUP tcp_srcport AGGREGATE SUM ip_dst
BACK
GROUP info_line AGGREGATE SUM eth_src
GROUP tcp_srcport AGGREGATE SUM length
GROUP info_line AGGREGATE SUM length
BACK
    
```

Table 5: EDA session without penalty score on Cyber Security dataset 1. Spurious repetition (red colored) of actions is seen.

6.1.2 Training without BC initialization: ILAEDA, which is pre-trained with Behavioural Cloning, performs better than a model without pretraining. This result suggests that pretraining the policy before training on the GAIL-based objective improves the overall model. We conjecture that this phenomenon is observed due to the pre-trained policy getting a *warm* start allowing it to start in a better position which gives it an advantage during the adversarial training.

6.2 Analysis of ILAEDA’s Ability to Capture A Diverse Set of Interestingness Measures

In Table 6, we show a session generated by our model on Cyber Security dataset 1 and normalized scores obtained on different interestingness metrics. We highlight scores greater than 0.7. Throughout this session, generated views obtained high readability, diversity, and peculiarity scores. This indicates that our model can effectively filter the dataset and present important tuples to users. Views presented are diverse, showing different insights from the dataset.

Table 7 shows a session generated using ATENA on the same dataset. We observe that views produced have high A-INT and diversity scores. This is because these scores were used as reward signals to train ATENA. In comparison, each action taken by our model shows different scores being highlighted, which shows that our model can capture a wider set of interestingness measures than ATENA.

6.3 Analysis of ILAEDA’s Ability to Capture Specific Interestingness Measures

To analyze ILAEDA’s ability to capture specific underlying interestingness measures we conducted an experiment where we trained ILAEDA on a subset of expert sessions filtered to optimize one specific measure on a source datasets. The goal is to examine if the sessions generated by the trained model on other datasets also reflect the domination of the same interestingness metric.

We filter the expert sessions from the Synthetic Datasets (refer to section 5.1.2) to obtain subsets where the sessions maximize one particular interestingness metric over others. We consider A-INT, Diversity, and Readability as the metrics for this analysis. To filter the expert sessions, we classify each session into one of three categories. (1) **Category 1:** Maximizing A-INT more than others. (2) **Category 2:** Maximizing Diversity more than others. (3) **Category 3:** Maximizing Readability more than others.

We compare the 75th percentile normalized score obtained for each metric over all the steps in that session’s trajectory to classify a session. If the 75th percentile normalized score over that trajectory is highest for the A-INT metric, then that session is classified as maximizing A-INT more than others. Similarly, we can classify a session as maximizing Diversity more than others or maximizing Readability more than others.

We do this classification for all Synthetic datasets 1-5 sessions and obtain three subsets. We then trained an ILAEDA model on these subsets using the default settings described before. We use the trained models to generate 100 EDA sessions each on Synthetic Dataset 7. For each generated session, we compare the median normalized score obtained for each metric across all the steps in the trajectory of that session. If the median normalized score is

highest for A-INT, we say this session maximizes A-INT. Similarly, a session can also be classified as maximizing Readability/Diversity more than other metrics. Figure 2 shows the distribution of most-maximized metrics for sessions generated using each of the three ILAEDA models. We see that in all three cases, most sessions maximize that interestingness metric more than others, for which the training sessions used to train the model generating those sessions were filtered. A model trained on sessions filtered for maximizing A-INT more than others tends to maximize A-INT more than other metrics when tested on an unseen dataset. This analysis strongly suggests that our method can capture specific underlying interestingness measures being optimized for in a set of expert demonstrations, and the trained model can also optimize for those measures on unseen datasets.

7 CONCLUSION AND FUTURE WORK

In this paper, we present ILAEDA, a novel imitation learning based end-to-end AutoEDA system which does not rely on handcrafted interestingness heuristics to learn a policy. Through automated benchmarking and manual analysis of generated EDA sessions, we demonstrate that ILAEDA can learn to perform meaningful EDA and outperform baselines across benchmarks. Further, we demonstrate the generalizability of our system on unseen datasets. We also show ablation studies to validate our design choices. For future work, some of the following investigations may be worthwhile: (a) training a system that generalizes to perform AutoEDA across datasets with different schemas, (b) exploring the effect of other penalties and (c) incorporating actions other than FILTER and GROUP into ILAEDA.

REFERENCES

- [1] Julien Aligon, Enrico Gallinucci, Matteo Golfarelli, Patrick Marcel, and Stefano Rizzi. 2015. A collaborative filtering approach for recommending OLAP sessions. *Decision Support Systems* 69 (2015), 20–30.
- [2] Ori Bar El, Tova Milo, and Amit Somech. 2020. ATENA Basic Implementation. <https://github.com/TAU-DB/ATENA-A-EDA/tree/master/atena-basic>.
- [3] Ori Bar El, Tova Milo, and Amit Somech. 2020. Automatically generating data exploration sessions using deep reinforcement learning. In *ACM SIGMOD*. 1527–1537.
- [4] Yukun Cao, Xike Xie, and Kexin Huang. 2022. Learn to Explore: on Bootstrapping Interactive Data Exploration with Meta-learning. *IEEE ICDE* (2022), 1720–1733.
- [5] Varun Chandola and Vipin Kumar. 2007. Summarization—compressing data into an informative representation. *Knowledge and Information Systems* 12, 3 (2007), 355–378.
- [6] Alexandre Chanson. 2022. Automatic generation of comparison notebooks for interactive data exploration. In *EDBT*.
- [7] Yiming Ding, Carlos Florensa, Pieter Abbeel, and Mariano Phielipp. 2019. Goal-conditioned Imitation Learning. In *NeurIPS*, Vol. 32.
- [8] Marina Drosou and Evaggelia Pitoura. 2013. Ymaldb: exploring relational databases via result-driven recommendations. *The VLDB Journal* 22, 6 (2013), 849–874.
- [9] Magdalini Eirinaki, Suju Abraham, Neoklis Polyzotis, and Naushin Shaikh. 2013. Querie: Collaborative database exploration. *IEEE TKDE* 26, 7 (2013), 1778–1790.
- [10] Divyansh Garg, Shuvam Chakraborty, Chris Cundy, Jiaming Song, and Stefano Ermon. 2021. IQ-Learn: Inverse soft-Q Learning for Imitation. *ArXiv abs/2106.12142* (2021).
- [11] Liqiang Geng and Howard J Hamilton. 2006. Interestingness measures for data mining: A survey. *ACM Computing Surveys (CSUR)* 38, 3 (2006), 9–es.
- [12] Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. *NIPS* 29 (2016).
- [13] Manas Joglekar, Hector Garcia-Molina, and Aditya Parameswaran. 2017. Interactive data exploration with smart drill-down. *IEEE TKDE* 31, 1 (2017), 46–60.
- [14] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.
- [15] Tavor Lipman, Tova Milo, and Amit Somech. 2023. ATENA-PRO: Generating Personalized Exploration Notebooks with Constrained Reinforcement Learning.

Action #	ILAEDA Trajectory	A-INT	Diversity	Coherence	Readability	Peculiarity
1	FILTER eth_src EQ 00:26:b9:2b:0b:59	1.00	0.89	0.83	1.00	0.00
2	FILTER ip_src NEQ 82.108.69.238	0.05	0.62	0.90	0.95	0.00
3	FILTER highest_layer NEQ ICMP	0.49	0.73	0.00	0.99	0.97
4	FILTER highest_layer NEQ ARP	0.05	0.68	0.00	0.95	0.97
5	FILTER captured_length NEQ 62	0.05	0.68	0.00	0.95	1.00
6	GROUP captured_length AGGREGATE COUNT packet_number	0.00	1.00	0.00	0.00	1.00
7	BACK	0.00	0.00	0.80	0.85	1.00
8	GROUP highest_layer AGGREGATE COUNT length	0.00	0.81	0.00	0.00	1.00
9	GROUP ip_src AGGREGATE COUNT length	0.14	0.92	1.00	0.85	1.00
10	GROUP tcp_srcport AGGREGATE COUNT length	0.07	0.77	0.88	0.92	1.00

Table 6: Trajectory generated by the ILAEDA for Cyber Security dataset 1 and normalized values attained by different interestingness measures for each action.

Action #	ATENA Trajectory	A-INT	Diversity	Coherence	Readability	Peculiarity
1	GROUP highest_layer AGGREGATE COUNT packet_number	1.00	0.90	1.00	1.00	0.71
2	GROUP eth_src AGGREGATE COUNT packet_number	0.89	0.72	0.53	0.10	1.00
3	BACK	0.00	0.00	0.67	0.10	0.71
4	BACK	0.00	0.00	0.80	0.09	0.00
5	GROUP eth_dst AGGREGATE COUNT packet_number	1.00	0.72	0.53	1.00	0.61
6	FILTER tcp_srcport CONTAINS 139.0	0.79	0.96	0.00	0.10	0.08
7	GROUP length AGGREGATE COUNT packet_number	0.94	0.66	0.53	0.10	0.11
8	GROUP eth_src AGGREGATE COUNT packet_number	0.93	0.60	0.51	0.10	0.11
9	GROUP highest_layer AGGREGATE COUNT packet_number	0.91	0.59	0.54	0.10	0.11
10	GROUP ip_dst AGGREGATE COUNT packet_number	0.89	0.58	0.00	0.10	0.11
11	GROUP tcp_srcport AGGREGATE COUNT packet_number	0.87	0.58	0.44	0.10	0.11
12	FILTER ip_src CONTAINS 82.108.10.135	0.46	1.00	0.00	0.00	0.00

Table 7: Trajectory generated by ATENA model for Cyber Security Dataset 1 and normalized values attained by different interestingness measures for each action.

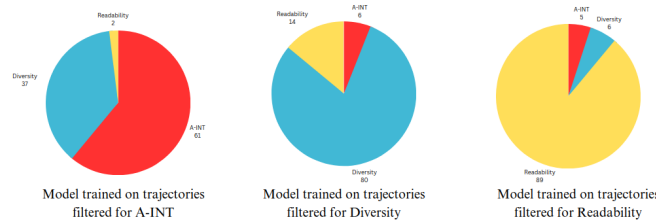


Figure 2: Distribution of metrics in sessions generated on Synthetic Dataset 7 using ILAEDA models. ILAEDA models were trained on sessions filtered to maximize A-INT, Diversity, or Readability more than other metrics.

- Companion of the 2023 International Conference on Management of Data (2023).*
- [16] Ping Ma, Rui Ding, Shi Han, and Dongmei Zhang. 2021. MetaInsight: Automatic Discovery of Structured Knowledge for Exploratory Data Analysis. *Proceedings of the 2021 International Conference on Management of Data* (2021).
 - [17] Pingchuan Ma, Rui Ding, Shuai Wang, Shi Han, and Dongmei Zhang. 2023. InsightPilot: An LLM-Empowered Automated Data Exploration System. In *EMNLP*.
 - [18] Tova Milo and Amit Somech. 2016. React: Context-sensitive recommendations for data analysis. In *Proceedings of the 2016 International Conference on Management of Data*. 2137–2140.
 - [19] Tova Milo and Amit Somech. 2018. Next-step suggestions for modern interactive data analysis platforms. In *KDD*. 576–585.
 - [20] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *ACL*. 311–318.
 - [21] Aurélien Personnaz, Sihem Amer-Yahia, Laure Berti-Equille, Maximilian Fabricius, and Srividya Subramanian. 2021. Balancing Familiarity and Curiosity in Data Exploration with Deep Reinforcement Learning. In *Workshop in Exploiting AI Techniques for Data Management*. 16–23.
 - [22] Aurélien Personnaz, Sihem Amer-Yahia, Laure Berti-Equille, Maximilian Fabricius, and Srividya Subramanian. 2021. DORA THE EXPLORER: Exploring Very Large Data With Interactive Deep Reinforcement Learning. *ACM CIKM* (2021).
 - [23] Kathy Razmadze, Yael Amsterdamer, Amit Somech, Susan B. Davidson, and Tova Milo. 2022. Selecting Sub-tables for Data Exploration. *IEEE ICDE* (2022), 2496–2509.
 - [24] Siddharth Reddy, Anca D. Dragan, and Sergey Levine. 2019. SQIL: Imitation Learning via Reinforcement Learning with Sparse Rewards. *arXiv: Learning* (2019). <https://api.semanticscholar.org/CorpusID:202888699>

- [25] Stéphane Ross and Drew Bagnell. 2010. Efficient Reductions for Imitation Learning. In *AISTATS*.
- [26] Sunita Sarawagi. 2000. User-adaptive exploration of multidimensional data. In *VLDB*, Vol. 2000. Citeseer, 307–316.
- [27] Sunita Sarawagi, Rakesh Agrawal, and Nimrod Megiddo. 1998. Discovery-driven exploration of OLAP data cubes. In *International Conference on Extending Database Technology*. Springer, 168–182.
- [28] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
- [29] Manish Singh, Michael J Cafarella, and HV Jagadish. 2016. DBExplorer: Exploratory Search in Databases. In *EDBT*. 89–100.
- [30] Amit Somech, Tova Milo, and Chai Ozeri. 2019. Predicting "what is interesting" by mining interactive-data-analysis session logs. In *EDBT*.
- [31] Lance Spitzner. 2003. The honeynet project: Trapping the hackers. *IEEE Security & Privacy* 1, 2 (2003), 15–23.
- [32] Faraz Torabi, Garrett Warnell, and Peter Stone. 2018. Behavioral Cloning from Observation. In *IJCAI*. 4950–4957.
- [33] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *arXiv:2302.13971* [cs.CL]
- [34] Matthijs van Leeuwen. 2010. Maximal exceptions with minimal descriptions. *DMKD* 21, 2 (2010), 259–276.
- [35] Manasi Vartak, Sajjadur Rahman, Samuel Madden, Aditya Parameswaran, and Neoklis Polyzotis. 2015. Seedb: Efficient data-driven visualization recommendations to support visual analytics. In *VLDB*, Vol. 8. NIH Public Access, 2182.
- [36] Xiaoyan Yang, Cecilia M Procopiuc, and Divesh Srivastava. 2009. Recommending join queries via query log analysis. In *IEEE ICDE*. IEEE, 964–975.
- [37] Brian D. Ziebart, Andrew Maas, J. Andrew Bagnell, and Anind K. Dey. 2008. Maximum Entropy Inverse Reinforcement Learning. In *AAAI*.

A INTERESTINGNESS MEASURES

- (1) **A-INT.** This interestingness score [3] is used as a reward function. It is defined for FILTER and GROUP operations as follows:
 - (a) Interestingness Score for a GROUP operation: This score uses *conciseness measures* [5] [11] which rewards compact group-by views covering many rows as these views are considered informative and easy to understand. This measure considers the number of groups g , the number of grouped attributes a , and the number of tuples r . The score $Int(d_t)$ is calculated as

$$Int(d_t) = \frac{h_1(g \cdot a)}{h_2(r)} \quad (9)$$

where h_1 and h_2 are normalized sigmoid function with fixed width and center.

- (b) Interestingness Score for a FILTER operation: To quantify the interestingness of a FILTER operation, *exceptionality* [27] [34] [35] of filtered rows (generated after applying the operation) is compared with those of the unfiltered table. The Kullback-Leibler (KL) divergence [14] on each column is used to measure how much the filtered data view differs from the unfiltered view. This score favors filter operations whose resultant display deviates significantly from the previous display. Let P_t^A be the distribution of an attribute (column) A at time step t in the analysis. The score $Int(d_t)$ is given by:

$$Int(d_t) = h(\max_A D_{KL}(P_{t-1}^A, P_t^A)) \quad (10)$$

where h is the sigmoid function.

- (2) **Diversity score.** This metric favors a display that highlights parts of the dataset that are different from those seen in any of the previous displays in the session so far [3]. It is calculated as the minimum Euclidean distance between the resultant display

of an operation and all previous displays, i.e.

$$Div(d_t) = \min_{0 \leq t' < t} \delta(d_t, d_{t'}) \quad (11)$$

- (3) **Readability score.** This metric uses compaction gain from [5]. This score measures changes in the compactness of the display before and after performing the current EDA operation. Highly compact displays with few rows are considered more readable and are given a higher reward. Let compact display score, $C(d_t)$, for current display d_t with number of groups g be defined as

$$C(d_t) = h_1(g \cdot |d_t|) \quad (12)$$

where h_1 is normalized sigmoid function with fixed width and center. Readability score, $Read(d_t)$ is defined

$$Read(d_t) = 1 - \frac{C(d_{t-1})}{C(d_t)} \quad (13)$$

- (4) **Peculiarity score.** This score helps quantify anomalous patterns. This a deviation-based measure from [35], which favors display with a high difference from the initial display at the start of the analysis. It is defined in terms of the KL-divergence as:

$$Pec(d_t) = h(\max_A D_{KL}(P_0^A, P_t^A)) \quad (14)$$

- (5) **Coherence score.** This score is a highly detailed metric determined by a set of handcrafted rules which assign each display a penalty or reward based on whether the operation performed at the current step is coherent with previous operations. There are two types of rules: (1) Rules that apply to any generic EDA session. For example, an operation that results in an empty or unchanged display is considered incoherent. (2) Rules specific to the dataset's domain to be explored. [3] define many such rules for the Cyber security datasets [31] being analyzed. These rules are highly specific and are carefully hand-tuned to predict coherence for actions taken on the Cyber security dataset.

B ALGORITHMIC DETAILS OF ILAEDA

Algorithm 1 gives a detailed description of the ILAEDA procedure.

C SYNTHETIC DATA GENERATION

Our synthetic data generation algorithm consists of three steps:

- (1) **Pattern Injection and Correlation:** In this step, we take in a schema $S = \{C_1, C_2 \dots C_k, N_1, N_2 \dots N_l, T_1, T_2 \dots T_m\}$ where C_i, N_i, T_i represent Categorical, Numeric and Text columns respectively. For each column $c \in S$, we randomly generate a set of patterns $P(c) = \{p_1, p_2 \dots p_n\}$ and a random weighting $W(c) = \{w(p_1), w(p_2) \dots w(p_n)\}$ over these patterns ($\sum w(p_i) = 1$). These patterns are representative artifacts of the data appearing in column c , and the weights represent the likelihood that an element appearing in a column follows a pattern:

- For categorical columns, each p_i is a category that appears in the column.
- For a numeric column, each p_i is a tuple (μ_i, σ_i) representing a Gaussian distribution with mean and variance (μ_i, σ_i) respectively. Every element in the numeric column N_j will be sampled from one of the $p_i \in P(N_j)$.

Algorithm 1 ILAEDA

Section 1: Pre-training with Behavioral Cloning

- 1: **Input:** Expert trajectories \mathcal{E} , initial policy parameters θ_0 , number of steps BC , α
- 2: **for** $i=0, 1, 2, \dots, BC$ **do**
- 3: Sample $\mathcal{T}_E = \{\tau_1, \tau_2, \dots\} \sim \mathcal{E}$
- 4: Update θ_i to θ_{i+1} by

$$\theta_{i+1} = \theta_i + \alpha \nabla_{\theta_i} \left[\frac{1}{|\mathcal{T}_E|} \frac{1}{T} \sum_{\tau \in \mathcal{T}_E} \sum_{t=0}^{T-1} -\log(\pi_{\theta_i}(a_t | s_t) + \|\theta_i\|_2) \right]$$

- 5: **end for**
- 6: **Output:** Pre-trained policy parameters θ_{BC}

Section 2: Adversarial Training

- 1: **Input:** Environment, Max Episode length T , Expert trajectories \mathcal{E} , Pretrained policy parameters $\theta_0 = \theta_{BC}$, randomly initiated value and discriminator parameters ϕ_0 and w_0 respectively, train interval T_{train} , learning rate α , clipping parameter ϵ
- 2: Initialize buffer \mathcal{B} for generated trajectories.
- 3: **for** $i=1, 2, \dots$ **do**
- 4: Sample trajectories $\mathcal{T}_i = \{\tau_1^{\theta_i}, \tau_2^{\theta_i}, \dots\}$ by interacting with the environment using π_{θ_i} , where each $\tau^{\theta_i} = \{(s_1, a_1), \dots, (s_T, a_T)\}$
- 5: Obtain (s_t, a_t, r_t, s_{t+1}) tuples from \mathcal{T}_i with $r_t = -\log(1 - D_{w_i}(s_t, a_t)) + P(s_{1:t}, a_{1:t})$. Add these tuples to \mathcal{B}
- 6: **if** $i \equiv 0 \pmod{T_{train}}$ **then**
- 7: Sample generated trajectories $\mathcal{T}_G \sim \mathcal{B}$ and expert trajectories $\mathcal{T}_E \sim \mathcal{E}$ such that $|\mathcal{T}_E| = |\mathcal{T}_G|$
- 8: Update discriminator parameters w_i to w_{i+1} as

$$w_{i+1} = w_i + \alpha \nabla_{w_i} \left[\frac{1}{|\mathcal{T}_G|} \frac{1}{T} \sum_{\tau \in \mathcal{T}_G} \sum_{t=0}^{T-1} \log(1 - D_{w_i}(s_t, a_t)) + \frac{1}{|\mathcal{T}_E|} \frac{1}{T} \sum_{\tau \in \mathcal{T}_E} \sum_{t=0}^{T-1} \log(D_{w_i}(s_t, a_t)) \right]$$

- 9: Update policy parameters θ_i to θ_{i+1} as

$$\theta_{i+1} = \arg \max_{\theta} \left[\frac{1}{|\mathcal{T}_G|} \frac{1}{T} \sum_{\tau \in \mathcal{T}_G} \sum_{t=0}^{T-1} \min \left(\frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_i}(a_t | s_t)} A^i(s_t, a_t, r_t, s_{t+1}), g \left(\epsilon, A^i(s_t, a_t, r_t, s_{t+1}) \right) \right) \right]$$

- 10: Update value parameters ϕ_i to ϕ_{i+1} as

$$\phi_{i+1} = \phi_i - \alpha \nabla_{\phi_i} \left[\frac{1}{|\mathcal{T}_G|} \frac{1}{T} \sum_{\tau \in \mathcal{T}_G} \sum_{t=0}^{T-1} (r_t + V_{\phi_i}(s_{t+1}) - V_{\phi_i}(s_t))^2 \right]$$

- 11: **end if**
- 12: **end for**

- For a text column, each p_i is a tuple (s_i, POS_i) where s_i is a string and $POS_i \in \{START, MIDDLE, END\}$. If an element in a string column follows pattern (s_i, POS_i) , that element contains the string s_i at the position specified by POS_i . The rest of it is padded with random strings.

Dataset	Train	Evaluation
1	614	154
2	3686	922
3	614	154
4	921	231
5	1638	410
6	819	205
7	1228	308

Table 8: Trajectory distribution for synthetic datasets

We then randomly generate a list of dependencies of *correlations* $C = \{C_1, C_2 \dots C_n\}$ where each $C_i = (c_{i_1}, c_{i_2}, \{r_{i_1}, r_{i_2} \dots r_{i_n}\})$ where every $c_{i_1}, c_{i_2} \in \mathcal{S}$ and $r_{i_k} = (p_{i_k}^1, p_{i_k}^2)$ such that $p_{i_k}^1 \in P(c_{i_1})$ and $p_{i_k}^2 \in P(c_{i_2})$. Note that the term correlation here is not confused

with mathematical correlations. Here, the term correlation between two columns indicates that the likelihood of a pattern appearing in one is influenced by the likelihood of patterns appearing in the other. Each $C_i = (c_{i_1}, c_{i_2}, \{r_{i_1}, r_{i_2} \dots r_{i_n}\})$ is to be understood to mean that column c_{i_1} is correlated with column c_{i_2} in the following way: each $r_{i_k} = (p_{i_k}^1, p_{i_k}^2)$ indicates that likelihood of $p_{i_k}^2$ appearing in column c_{i_2} is greater every time $p_{i_k}^1$ appears in column c_{i_1} . We ensure that there are no cyclic chains formed in the list of correlations (a set $\{C_1, C_2 \dots C_k\}$ such that $c_{1_2} = c_{2_1}, c_{2_2} = c_{3_1} \dots c_{n_2} = c_{1_1}$) and limit the number of correlations that a column can appear in to make sure the set of correlations forms a Directed Acyclic Graph (DAG) over the set of columns.

(2) Row population: In this step, we take the schema \mathcal{S} and generated list of correlations C and generate the rows of the dataset. We generate each row one by one. To generate the element at column c in a row, we do the following:

- Set base distribution $W = W(c)$ and get the relevant set of correlations $C_R = \{C_i \in C \text{ such that } c_{i_2} = c\}$.
- For each correlation $C_i \in C_R$, for each $r_{i_k} = (p_{i_k}^1, p_{i_k}^2)$, multiply weight of $p_{i_k}^2$ in W with pre-defined multiplier m .

- Normalize W and sample pattern p according to updated weights. Generate elements according to p as described above.

(3) Trajectory Generation: The correlations generated in step 1 form a Directed Acyclic Graph (DAG) where the nodes are columns and the edges correspond to the correlation between them. We model an example of an EDA trajectory to be a traversal of this DAG from the root node (corresponding to the base view) to each leaf node and back. The path taken determines the correlations uncovered in the analysis session with each edge traversal corresponding to a particular analysis action. The column type and its patterns determine the operations performed during this process.

For instance, if there is an edge between two categorical columns (c_i and c_j) with related patterns values p_i and p_j , the possible order of operation would be either `[FILTER c_i EQ p_i , FILTER c_j EQ p_j]` or `[FILTER p_i EQ v_j , GROUP c_i AGGREGATE COUNT c_j]`. Numerical columns are filtered based on the mean of the distribution from which the values in each column are sampled, and text columns are filtered based on the pattern present in the column and its position. The algorithm returns to the root node of the subtree using BACK actions, and the trajectory ends when each node has been visited in topological order. We traverse the graph in a depth-first search approach to generate expert trajectories over the synthetically generated dataset. Each trajectory forms an expert EDA session.