AUTOMATED HYPOTHESES GENERATION VIA EVOLUTIONARY ABDUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Abduction is a powerful form of causal inference employed in many artificial intelligence tasks, such as medical diagnosis, criminology, root cause analysis, intent recognition. Given an effect, the abductive reasoning allows advancing a plausible set of explanatory hypotheses for its causes. This paper presents a new evolutionary strategy - called *Evolutionary Abduction* (EVA) - for automated abductive inference, aiming at effectively generating sets of hypotheses for explaining an occurred effect and/or predicting an effect that could occur in the future. EVA defines a set of abductive operators to repeatedly construct hypothetical cause-effect instances, and then automatically assesses their *plausibility* as well as their *novelty* with respect to already known instances - a mechanism mimicking the human reasoning employed whenever we need to select the best candidates from a set of hypotheses. Experiments with four datasets confirm that, given a background knowledge, EVA can construct new and realistic multiple-cause hypotheses for a given effect. EVA outperforms alternative strategies based on *causal structure discovery*, generating closer-to-real instances in most settings and datasets.

1 INTRODUCTION

The term *abduction* was coined by Charles S. Peirce, in his work on the logic of science, to denote a powerful form of causal reasoning frequently employed both in everyday common-sense reasoning and as first step of scientific reasoning (Peirce, 1931), (Douven, 2017). It refers to the creative generation of explanatory hypotheses for a set of facts or observations (Crowder & Carbone, 2017). With important exceptions ((Spohn, 2012); (Woods, 2013)), many view abduction as the strongest candidate for a third top-level inference type besides *deduction* and *induction* (Preyer & Mans, 1999), (Walton, 2001). Abductive reasoning infers *possible causes* for a given *effect*; it comes into play when we try to explain something that has no immediate explanation, by advancing *hypotheses* or generating new ideas outside the given facts (Crowder et al., 2020). As such, abduction is said to be an *ampliative* form of inference, as it is able to enlarge our knowledge, but *uncertain*, because its inferences are more susceptible to error than deductive and inductive ones, and need to be “validated”.

This article presents a new algorithm inspired to abduction, named *Evolutionary Abduction* (EVA). EVA mimics the process of actively searching for explanations for a given observation, by generating hypotheses for *plausible* causes of an effect, exploiting both an *experience-based* knowledge and the *ontological knowledge* a human has about the phenomena in explanation. It defines three operators to mimic and automate the most common *patterns* of abduction (Schurz, 2008), which ultimately lead to construct cause-effect combinations as solutions to causal problems.

The article first formulates a causal problem as a specific class of combinatorial optimization, named Combinatorial Causal Optimization Problems (CCOP) (Section 3). Then, EVA is presented in Section 4, wherein the algorithm, the operators to construct solutions and the mechanisms to assess their plausibility are introduced. Section 5 and 6 report the evaluation. EVA is experimented on four real-world datasets, two form the medical domain, one from hazard analysis in the avionic domain, and one from a decision-support system, having a number of variables (namely, co-occurring causes for a given effect) that ranges from 9 to 27. Results demonstrate the potential of using evolutionary computation for automated causal inference: using a small fraction (10%) of the datasets as knowledge base, EVA constructs cause-effect combinations very similar to the real occurred events. For three out of four problems, it produces several combinations (37% of the total, in the average) exactly matching the real occurred events. In the hardest problem (hazard analysis in the avionic domain), which calls for formulating hypotheses for sets of 27 co-occurring causes to predict a potential accident, it gives predictions with up to 24 (out of 27) matching causes. In all the cases, EVA outperforms alternative strategies based on causal structure discovery algorithms, generating closer-to-real cause-effect instances in most settings and datasets.

2 BACKGROUND

Consider the rule $B_j \leftarrow A_i$ (*if* A_i *then* B_j), where B_j are (sets of) observations and A_i their causes in a given domain U: **deduction** refers to deriving the conclusion (or explaining the *effect*) B_j given the premises (the *causes*) A and the rule; **induction** tries to learn the *"possible" rule* $B_j \leftarrow A_i$ from observing many A_i, B_j pairs; **abduction** refers to deriving the *possible "cause"* A_i (i.e., hypothesis) given B_j and the rule. Both induction and abduction are *uncertain* but *ampliative* inferences: the conclusion is a probable (not a necessary) derivation of the premises, but the information content in the conclusion is "wider" than its premises. However, while induction just determines a value by gradual modification of the actual conclusion, only abduction is a process able of introducing any new idea, allowing to form new explanatory hypotheses. There exist different types of abduction. A broad classification distinguishes abduction whose task is to choose the best candidate among a set of possible explanations, called *selective* , from abduction that introduces new concepts or models, called *creative* abduction (Magnani, 2001; 2010). *Factual abduction* is the common form of selective abduction, wherein the abduced causes are found by backward reasoning from the effect. Creative abduction is more rare in common-sense reasoning, but plays a decisive role in advanced scientific reasoning. It can construct something new, for example a new theoretical model or a new concept. Following (Schurz, 2008), creative abduction can be classified in *analogical* and *hypothetical cause abduction*, depending on whether the concept is merely partly or completely new. EVA considers:

Factual abduction, in which both the effect and the abduced causes are singular *facts*. The set of possible conjectures is finite and can be generated by backward-chaining inference. In this form, abductive inference has been studied in detail in artifical intelligence (Flach & Kakas, 2000), for instance in the context of medical and legal reasoning (e.g., (Pukancová & Homola, 2015), (Ciampolini et al., 2002),(Josephson & Josephson, 1994), (Bex & Verheij, 2012), (Bex & Verheij, 2013)).

Analogical abduction, in which one abduces a *partially new* hypothesis by projecting knowledge from previous situations in the domain under analysis – e.g., inferring the propagation and reflection law of sound based on the laws of water waves. The process involves a conceptual abstraction and a mapping between a *source* context (about which the agent has knowledge) and the *target* context.

Both factual and analogical abduction use a form of *experience-based knowledge* – the former about the problem to solve (i.e., the occurred *facts*), the latter also about an external (source) context from which the analogies are constructed. Such knowledge will be encoded by EVA.

Hypothetical cause abduction. In this case, one abduces that one or more intercorrelated phenomena are the effect(s) of a hypothetical (unobservable) cause or of a hypothetical common cause. In both cases, the abductive conjecture postulates a new unobservable entity together with new laws connecting it with the observable properties, without drawing on analogies to concepts with which one is already familiar. This also includes the "pure speculation" process that sometimes lead to find a solution serendipically. Importantly, unlike factual and analogical abduction, hypothetical cause abduction does not presuppose any experience-based knowledge, but just knowledge about the phenomenon in explanation (namely, about the entities that can be possibly involved in the inference – in EVA, it is called the *ontological knowledge* (simply *ontology*) of the domain under analysis).

3 CAUSAL OPTIMIZATION

The abductive causal inference is hereafter formulated as a combinatorial problem, that we call *Combinatorial Causal Optimization Problem* (CCOP). In a CCOP, the goal is to find a proper set of *causes* A_i (or *explanations* or *hypotheses*) for a given set of *effects* B_j (or *obervations* or *manifestations*) that minimize (maximize) one or more objectives. Causes and effects are abstractions of a *phenomenon* or *event* regarding any element of interest $i \in U$, where U is the domain of the problem to be solved (namely, the set of elements possibly involved in the inference). There are more ways of formalising abduction. Here, logic-based abduction is used, without loss of generality. In logic:

Definition 1 (***Cause and effect***)**.** Causes and effects are *literals*, namely atomic formulae or their negation (a.k.a. *atoms*). In particular, in first-order logic (and in abductive logic programming (Kakas et al., 1992)), atoms correspond to predicate symbols together with their arguments, and a cause-effect pair to infer is a *clause* conveniently represented as a *rule* $B_j \leftarrow A_i$.

The literals are the decision variables (DV) of a CCOP, x_i, with $i = 1, \ldots, n = |U|$. Each x_i takes values in a non-empty discrete set representing its respective domain, D_i. For instance, in medical diagnosis, a value from blood analysis can be modelled as a decision variable x_i taking the following values from a discrete set D_i=$\{M, C\}$: "*M: moderately over the threshold", "C: critically over the threshold*". In a CCOP, the DVs are related by a causality relation:

Definition 2 (***Causality relation***)**.** A causality relation ($\xrightarrow{c}$) between two decision variables holds if the values of one – *the cause* - can determine or contribute to the other – the *effect*.

Strictly speaking, this is a *potential* causality relation, since the relation holds if the values of a variable *can*, but not necessarily will, be a cause for the other. Let us denote as $X = \{x_1, \ldots x_n\}$ a set of DVs. In a CCOP, this is viewed as the union of two subsets of variables, representing the causes and effects. Given the above, a CCOP is characterized as follows:

Definition 3. A ***Combinatorial Causal Optimization Problem*** is an optimization problem where:
i) the (discrete) DVs encode *literals* referring to any element $i \in U$, where U is the set of elements possibly involved in the causal inference;
ii) the set of DVs X is the union of two non-empty disjoint subsets related by a *causality* relationship: X_s, that is the set of causes (named *sources*), and X_t, the set of effects to be explained (named *targets*): $X = X_s \cup X_t = \{x_{s_1}, \ldots, x_{s_j}; x_{t_{j+1}}, \ldots, x_{t_n}\}$, and: $\forall x_s \in X_s,\ \exists x_t \in X_t : x_s \xrightarrow{c} x_t$ (i.e., each variable in X_s has a potential causality relation with at least one variable in X_t). Each source (target) variable $\{x_{s_1}, \ldots, x_{s_j}\}$ ($\{x_{t_{j+1}}, \ldots, x_{t_n}\}$) takes values in the respective discrete set: $D_s = \{D_{s_1}, \ldots D_{s_j}\}$ ($D_t = \{D_{t_{j+1}}, \ldots D_{t_n}\}$);
iii) C is the set of constraints between DVs: $C = C_k \cup C_u = \{c_{k_1}, ..., c_{k_q}; c_{u_{q+1}}, ..., c_{u_l}\}$, with C_k and C_u being the *known* and *unknown* constraints. Not all constraints are necessarily known a priori, thus the limits of the admissible solution space are, in general, not known;
iv) KB is the knowledge base, which is a set of cause-effect combinations (i.e., $\{x_{s_1}, \ldots, x_{s_j}\}$, $\{x_{t_{j+1}}, \ldots, x_{t_n}\}$) already observed, representing the experience-based knowledge with respect to which plausibility and novelty of solutions are assessed.

Finding a *solution* of a CCOP means finding suitable combinations of causes and effects that meet the constraints. For condition ***iii)*** of Def. 3, there is no assumption about the knowledge of the constraints between DVs (e.g., causes that cannot occur together). Because of this, proposed solutions can be just **hypotheses** (like in abductive reasoning) and need to be assessed for their *plausibility*, as they could violate constraints not known a priori. Based on Def. 3, a CCOP is expressed as follows:

$$\begin{aligned} \text{Maximize} \quad & \pi(\mathbf{x}), \\ \text{s.t.} \quad & \nu(\mathbf{x}) > \nu_0, \qquad \mathbf{C} = (\mathbf{C}_k, \mathbf{C}_u) = (c_{k_1}, \ldots, c_{k_q}; c_{u_{q+1}}, \ldots, c_{u_l}) \\ & \mathbf{x} = (\mathbf{x}_s; \mathbf{x}_t) = (x_{s_1}, \ldots, x_{s_j}; x_{t_{j+1}}, \ldots, x_{t_n}) \in \Omega \end{aligned} \tag{1}$$

where:
i) $\Omega = \{D_s \cup D_t\}^n$ is the decision space, the set of all possible values that decision variables can take. In the abduction metaphor, Ω represents the ontological knowledge (or *ontology*) of the domain, namely all the causes and effects that can concur to the inference;
ii) $\mathbf{x}$ is a candidate solution – a solution proposes an explanation for the effect(s) in $\mathbf{x}_t$ by potential cause(s) in $\mathbf{x}_s$. Note that a variable *can* (not necessarily will) be part of a solution; $1 \leq |\mathbf{x}_s|, |\mathbf{x}_t| \leq n$;
iii) $\mathbf{C}$ is the set of constraints. Each c_j is a pair $\langle v_j, R_j \rangle$, where $v_j \subset \mathbf{x}$ is a subset of h DVs and R_j is a h-ary relation on the corresponding (source/target) subsets $D_{j_{s/t}}$. An *evaluation*, namely a function from a subset of DVs to a particular set of values in the corresponding subsets $D_{j_{s/t}}$, satisfies a constraint $\langle v_j, R_j \rangle$ if the values assigned to v_j satisfy the relation R_j. Constraints are split as known and unknown ($\mathbf{C}_k$, $\mathbf{C}_u$). The former are predefined, hence evaluated during the search by the algorithm. The satisfaction of the latter is assessed by evaluating the solution's *plausibility*;
iv) π, ν. A solution is characterized by a *plausibility* score, $\pi(\mathbf{x})$, which is the objective function ($\pi : \Omega \rightarrow \Pi \subseteq \mathbb{R}$). A solution is also assigned a *novelty* score, $\nu(\mathbf{x})$: $\nu : \Omega \rightarrow N \subseteq \mathbb{R}$, and ν_0 is a novelty threshold. $\pi(\mathbf{x})$ and $\nu(\mathbf{x})$ are assessed with respect to the KB (see Section 4.2).

Algorithm 1: Evolutionary Abduction

1 $P_{F/A/H_0} \leftarrow$ getRandomPop(); ▷ *Get initial random populations. $P_{F/A/H}$: short for P_F, P_A, P_H*

2 $S_{F/A/H_0}, T_{F/A/H_0} \leftarrow$ getAllSourcesTargets($P_{F/A/H_0}$); $t \leftarrow 0$;
▷ *Get all different source and target values from current population*

3 **while** *stopping conditions are not satisfied* **do**
  ▷ *Three sequential loops, cycling on P_F, P_A, and P_H*

4   **for** *i=1 to* $|P_{F/A/H_t}|$ **do**

5     $\mathbf{x}_{i,t} \leftarrow$ select_solution($P_{F/A/H_t}$, KB_A);

6     $\mathbf{y}_{i,t} \leftarrow$ apply_abd_operator($\mathbf{x}_{i,t}$, $P_{F/A/H_t}$, $S_{F/A/H_t}$, $T_{F/A/H_t}$);

7   evaluate($\mathbf{y}_{i,t}$); ▷ *Plausibility evaluation*

8   evaluateConstraints($\mathbf{y}_{i,t}$); ▷ *Novelty evaluation*

9     $P_{F/A/H_{t+1}} \leftarrow P_{F/A/H_{t+1}} \cup (\mathbf{y}_{i,t})$;

10   $Q_{F/A/H} \leftarrow nextPopulation(P_{F/A/H_t} \cup P_{F/A/H_{t+1}})$; ▷ *Merge by non-dominated sorting*

11   $t \leftarrow t+1$; $P_{F/A/H_t} \leftarrow Q_{F/A/H}$;

12   $S_{F/A/H_t}, T_{F/A/H_t} \leftarrow$getAllSourcesTargets($P_{F/A/H_t}$);

13 $P_t \leftarrow P_{F_t} \cup P_{A_t} \cup P_{H_t}$; $R \leftarrow$ getRankedSolutions(P_t); ▷ *Merge solutions*

4 EVA: THE EVOLUTIONARY ABDUCTION ALGORITHM

4.1 OVERVIEW

EVA builds solutions for the defined CCOPs. To this aim, it keeps an archive of known solutions representing the KB; besides, it keeps a second archive of solutions called *analogical* KB (KB_A), representing the further knowledge used in *analogical reasoning* to find new solutions starting from solutions in similar problems. The main steps EVA performs (Algorithm 1) follow:

Initialize. Initially, three sets of solutions (named *populations*) are created by selecting random values for each variable x_i in **s** and in **t** from their domains D_i and that satisfy the known constraints $\mathbf{C}_k$. These represent solutions from the *factual, analogical* and *hypothetical cause* abduction, and are merged to form the initial population P. Plausibility and novelty of P are evaluated (Sec. 4.2).

Apply operators. At every iteration, a *selection* and a specific *abduction operator* are applied to the corresponding sub-population (Section 4.3) – lines 4-9. *Selection* takes a solution from the population; the abduction operator operator creates a new solution starting from the selected one. The latter ones are: the *factual, analogical* or *hypothetical cause* operators, which mimic the three abduction patterns. The loops generate the *offpsring* sub-populations. The sub-populations evolve independently, and merged at the end of the algorithm (line 13).

Merge. The current and offspring populations are merged by a crowding distance criterion to form the new population, similarly to the *Deb et al.* (Deb et al., 2002) (line 10). All non-dominated fronts $\mathcal{F}_i$ are obtained from the union of current and offspring population by the fast non-dominated sort algorithm; then, until the population is filled, the crowding distance is calculated in each $\mathcal{F}_i$ and the corresponding solutions are included in the population.

4.2 PLAUSIBILITY AND NOVELTY EVALUATION

Plausibility. Whenever a solution is proposed by an operator, its plausibility needs to be assessed first. Plausibility is the degree to which a hypothesised solution is judged as possible to occur, namely how much it is *realistic*. The following definitions formalize the concepts that a human typically uses to judge a hypothesis as more or less plausible. Typically, our judgement depends on whether we recognize "parts" of the hypothesis in what we already know about the phenomenon under study. These "parts" are co-occurrences of (subsets of) causes and effects: whenever we recognize cause-effect patterns in the knowledge base (KB) at least once, we increase our belief about the plausibility of the entire solution, since at least part of the hypothetical solution has already been seen (hence, it can actually occur). Plausibility assessment is based on such a notion of pattern occurrence. Given the KB, the k-degree and m-degree of a solution $\mathbf{x}_j$={**s**, t} are defined as follows:

Definition 4 (***k- and m-degree of a solution*** (δ_k, δ_M)**).** The *k-degree* of $\mathbf{x}_j$ ($\delta_k(\boldsymbol{x}_j)$) is the number of distinct k-tuples of the source variables set **s** (with $k \leq |\mathbf{s}|$) that occurred at least once in KB along with the target t. The *m-degree* of $\mathbf{x}_j$ ($\delta_M(\boldsymbol{x}_j)$) is the maximum value of k such that $\delta_k(\boldsymbol{x}_j) > 0$.

Definition 5 (***Plausibility***)**.** The plausibility $\pi(\mathbf{x}_j)$ of a solution $\mathbf{x}_j$ = {**s**, t} with $p = |\mathbf{s}|$ is:

$$\pi(\mathbf{x}_j) = \sum_{k=1}^{p} \delta_k(\mathbf{x}_j) \tag{2}$$

which is the sum of all k-tuples of **s**, excluding the 0-tuple, that occurred at least once in KB along with t. The so-defined plausibility naturally prefers small solutions, since the denominator explodes with p. To account for this, $\pi(\mathbf{.})$ is multiplied by a scale factor, that is p/a (if $p \leq a$) or a/p (if $p > a$), where p is the solution size and a is the average size of the solutions in the KB.

Novelty. Hypothesised solutions need to be plausible but also *different* from the ones already observed, to avoid to converge towards solutions already present in KB. A novel solution is one that is not similar to existing ones. As similarity metric, the Jaccard similarity coefficient is used.

Definition 6 (***Novelty***)**.** The novelty $\nu(\mathbf{x}_j)$ of a solution $\mathbf{x}_j$ is given by the minimum dissimilarity of $\mathbf{x}_j$ with respect to all solutions $\mathbf{x}_h$ in KB, thus: $\nu(\mathbf{x}_j) = 1 - max_h(J(\mathbf{x}_j, \mathbf{x}_h))$, where: $\mathbf{x}_h \in KB$, with $h = 1$ to $|KB|$, $J(\cdot, \cdot)$ is the Jaccard similarity coefficient.

4.3 Operators

Selection. Selection is always done with the Deb's version of *Binary Tournament* (Deb et al., 2002).
Factual abduction. Factual abduction exploits the available experience to "select" the best hypothesis for solving a problem. In EVA, the interest is not in selecting an entire solution, but in creating a new, diverse, solutions by selecting parts of observed solutions and combining them. The agent mimics the factual abduction process where single *facts* are selected from the KB and combined in a new way. Specifically, it mimics a human that starts from an observed solution (i.e., a cause-effect occurrence) and comes up with a new solution by modifying the observed one. The agent uses solutions already observed or hypothesised: starting from them, the operator changes (*add*, *modify* or *delete*) the sources and targets using either solutions from the KB (*observed* facts) or from the current population (i.e., *hypothesed* facts). The algorithm is in (Appendix A, Algorithm 2).
Analogical abduction Analogical reasoning looks not (merely) for similar causes or effect, but for similar "*relations*" between the elements of the analogy across two problems, or two different instances of the same problem, P and P' (Schurz, 2008). Thus, the operator acts as follows: the starting solution **x** is the one selected from KB_A. To build the new solution **x'**, the operator first selects a target from the set of all different targets in P. Then, it builds the set of sources, coupled with the chosen target, by extracting and reproducing *structural* features of the sources in **x**. Three source-level constraints are defined currently in EVA, which require the new solution **x'** to have progressively stronger similarities with **x**: a *cardinality constraint*, requiring the number of sources of **x'** to be the same as **x**; *a group membership constraint*, requiring the new solution **x'** to have the same number of subsets with the same cardinalities; an *ordinal constraint*, requiring **x'** to have the same number of subsets with the same σ_M (see Def. 4) of the subsets of **x'** (Alg. 3, Appendix A).
Hypothetical cause abduction This operator mimics the creative abduction allowing a human to advance hypotheses exploiting just his knowledge about the domain of interest (i.e., the ontology Ω). The operator (Alg. 4, Appendix A) acts as the factual operator: it *add*s, *modif*ies or *delete*s the sources of the selected solution, but it considers Ω in lieu of KB as set from which a new source can be selected, thus opening to a wider range of novel solutions. A consequence is that these solutions are expected to have a higher novelty compared to the factual operator.

All the operators have a hyperparameter called *novelty index*, $\eta \in [0; 1]$, that regulates the novelty of a solution; *factual* and *hypothetical cause* have one more hyperparameter, the *change index* $\gamma > 0$: the number of *add/modify/delete* to apply is selected randomly in $[1; \gamma]$. Details are in Appendix A.

5 Experiments: Datasets, metrics, baseline, factors

Datasets. *Primary Tumor dataset*. This dataset is provided by the Ljubljana Oncology Institute (UCI, a). Here, the effect to predict is the *type of tumor*. The number of variables is 18 (17 causes plus the effect), with information potentially related to the type of tumor (e.g., age, sex, histological information). There are 339 entries; these are split (randomly at every repetition) in two sets: the former used as KB, the latter as test set.[1] A separate set of entries is used as analogical KB (KB_A).

[1]Note that this is not like a training set in machine learning, since EVA does not learn a "model" from the set, it just uses it as base for assessing plausibility and novelty of generated solutions

The sizes are: $|KB|$ = 10% of the dataset, thus $|KB| = 34$, and $|KB_A| = 2.5\% = |KB_A| = 8$. The size of the decision space is $|\Omega| = 59$ (in the average, 3 values per variable).

ASRS dataset. In safety engineering, a *hazard analysis* is the activity aimed at hypothesizing possible hazards that can occur in operation. It requires a systematic generation of new hypotheses. The ASRS (Aviation Safety Reporting System) maintains a DB with thousands of accident reports (ASR, a). The events regarding the aircraft components, the weather conditions, the personnel, and many other potential causes are recorded for each accident, which is the final effect. As this is a new dataset built from the ASRS DB, details about its structure are in Appendix G. A subset of 4,470 reports is used. As above, $|KB| = 10\% = 447$, and $|KB_A| = 2.5\% = 112$. The variables are 28 (27 causes, 1 effect); the decision space size is $|\Omega| = 676$ (24 values per variable, in the average).

Diabetes dataset. This is a known dataset with information pertaining to blood glucose levels, insulin dosage, eating and exercise patterns of 70 diabetes patients (Frank & Asuncion, 2010). The variables are formatted as in (Acharya, 2014); a record contains the events occurring in one day to one patient. The effect of interest is the presence of Hypoglycemic symptoms. The whole dataset consists of 28,265 events: the formatted dataset contains 3,640 entries. The sizes are: $|KB| = 364$, $|KB_A| = 91$, which are 10% and 2.5% of dataset size as above.The number of variables is 14 (13 causes and 1 effect); the size of Ω is 70 (5 values per variable, in the average).

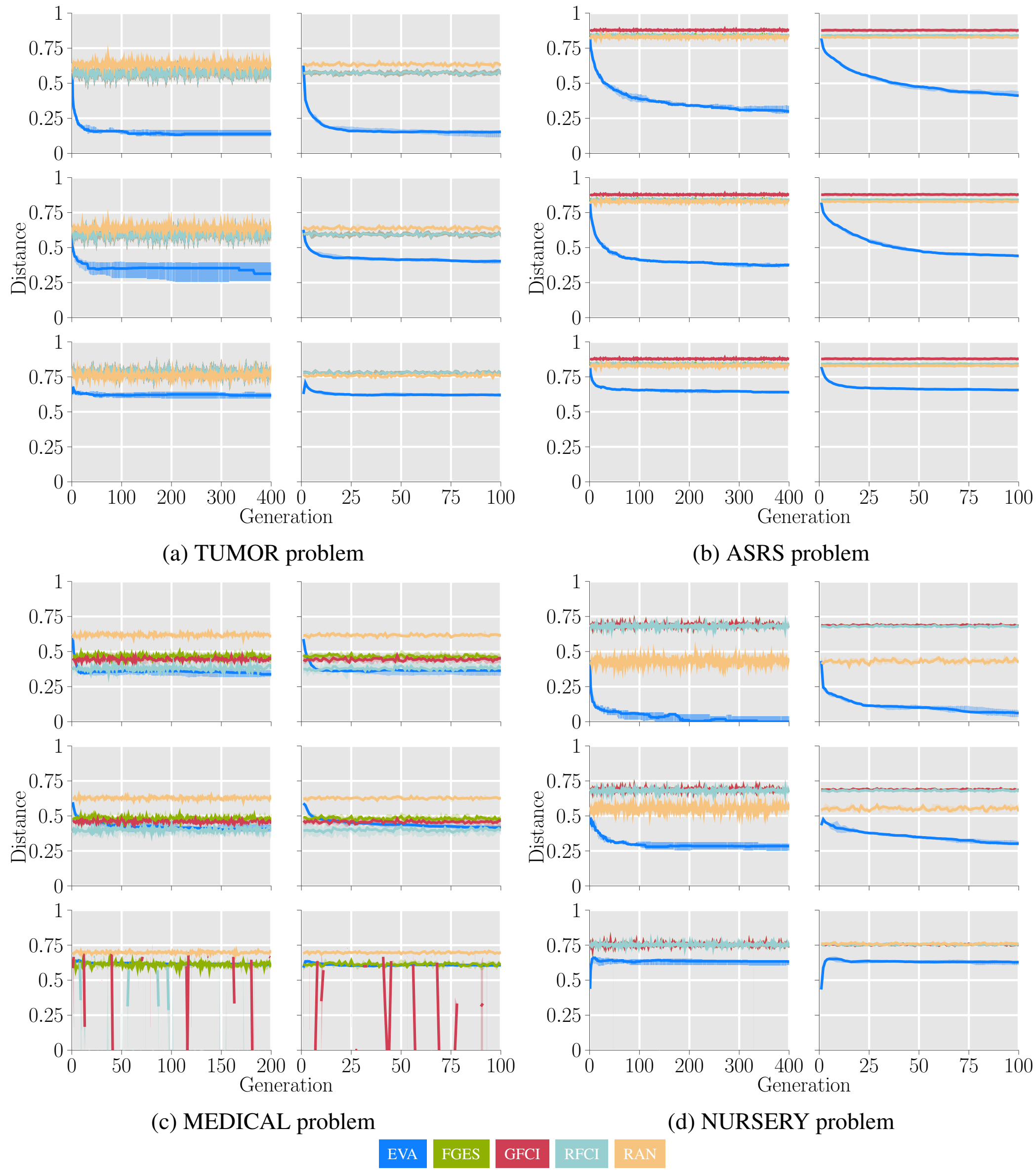
Nursery dataset. This was derived from a decision model to rank applications for nursery schools (UCI, b)(M. Olave, 1989). The effect variable is the decision on acceptance or rejection of an application (split in 5 classes, from "not recommended" to "special priority"); the potential explanatory causes are variables depending on the occupation of parents and child's nursery, the social and health picture of the family. The dataset has 12,960 entries (thus $|KB|$=$1,296$, $|KB_|A$=324) and 9 variables; $|\Omega| = 31$ (3 values per variable, in the average). The datasets, along with EVA implementation and results, are at **`http://github.com/eva-iclr-2021/EVA`**

Metrics. The generated solutions are compared against the test set Z. The aim for the agent is to generate hypotheses as close as possible to real ones, namely, plausible *with respect to its knowledge* represented by KB, but also sufficiently novel. Let us denote the set of generated solutions as Q. These are compared to each $\mathbf{z} \in Z$ to assess how much they are close to *at least one* real occurrence, according to a distance d. For each solution $q \in Q$, we take the *minimum distance*: $d_{min}(q) = \min_{z \in Z}(d(q, z))$. A small d_{min} suggests that the agent builds, based on its restricted knowledge, solutions that are similar to a real occurrence.[2] As distance metric, the Jaccard distance is taken between $q \in Q$ and $z \in Z$. Given a set of solutions Q, we take the average of $d_{min}(q)$, with $q \in Q$, and the minimum of $d_{min}(q)$, i.e., the smallest distance between $q \in Q$ and a solution $z \in Z$. The latter is the distance to a real solution of the *best* solution of the population.

Baselines. As baselines, we use a random generation strategy (RAN), for validation purpose, and three graph-based strategies (GB). RAN randomly selects values from the effect and cause variables to build hypotheses. GB strategies represent the probabilistic alternative to our plausibility-based approach. To implement them, we exploit *causal structure discovery* (CSD) algorithms to first learn a causal structure from the KB, represented as a directed acyclic graph (DAG)(Pearl, 2009). The learned structure is then used to generate hypotheses: fixing the effect, what are the most probable causes for it. Three state-of-the art CSD algorithms are used: FGES (Ramsey, 2015), which is a *score-based* strategy, RFCI (Colombo et al., 2012), a *constraint-based* one, and GFCI, a *hybrid* strategy (Ogarrio et al., 2016). These are run with bootsrtapping, giving weights $W = \{w_i\}$ on the arcs representing probabilities for the arc presence. Our generator then selects a variable v_i with probability w_i as part of the solution, and takes a value for v_i proportionally to the estimate of its probability of occurrence – estimated as its relative frequency within KB. Details in Appendix B.

Setting. We run a sensitivity analysis (via 3×3 grid search) on 3 configurations of EVA *hyperparameters* ($< \eta., \gamma. >= (< 0.1, 3 >, < 0.5, 5 >, < 0.9, 7 >)$, representing *Low, Medium* and *High* novelty, and 3 sizes of the *population*, $|P| = (15, 30, 60)$. A best (**B**) and worst (**W**) configuration for EVA are identified, for each datasets (results in Appendix C). The results obtained with these two settings (B/W) and varying the minimum novelty constraint as $\nu_0 = (0.1, 0.4, 0.7)$ (10 repetitions) are reported hereafter. All the experiments have a fixed budget: $y = 6,000$ evaluations (i.e., solution computations) for each technique. The number of generations g depends on the population size $|P|$, which in turn depends on the configuration (**B** or **W**), and derives from: $y = 6,000 = |P| \times g$.

[2]Note that this is a conservative metric: If d_{min} is large, the generated solution may either represent a non-plausible solution or it is a new plausible solution but far from the solutions in Z, namely it is new even with respect to the test set – in fact, Z is just a subset of all plausible solutions.



(a) TUMOR problem

(b) ASRS problem

(c) MEDICAL problem

(d) NURSERY problem

Figure 1: Distance by generation.

6 RESULTS

Results by generation. Figure 1 reports the average distance d of the population's solutions from the test set *vs.* generations (median and IQR over 10 repetitions). For all the datasets, EVA, has a decreasing trend, which is more pronounced in about the first 25% of the generations.

Impact of novelty constraint. The gain achieved by EVA is greater when $\nu > 0.1$, as there is more margin for improving the initial random solutions. Clearly, the stricter the constraint (i.e., bigger ν), the more difficult is to get solutions close to the test set. When $\nu > 0.7$ case, there are several generations in which the GB-based strategies do not manage to produce a population (a population requires at least one solution satisfying the constraint) – a problem that EVA does not have.

Impact of dataset. The different absolute values between the datasets is related to the features of the problem, i.e., to the decision space (number of variables and possible values per variable) and size of the dataset.The gain of EVA is more pronounced in more complex problems (ASRS, TUMOR) and with bigger datasets (NURSERY)while it is marginal with MEDICAL. Appendix D analyzes

Table 1: Best distance (d^*) of solutions of the best population – mean over 10 repetitions

		TUMOR		ASRS		MEDICAL		NURSERY	
		B	W	B	W	B	W	B	W
EVA	$\nu_0 = 0.1$	0.1097	0.0850	0.2271	0.2722	0.0	0.0	0.0	0.1466
	$\nu_0 = 0.4$	0.2414	0.1955	0.3313	0.3217	0.1166	0.1333	0.1377	0.2066
	$\nu_0 = 0.7$	0.4679	0.3254	0.6060	0.6132	0.4671	0.3683	0.1955	0.2066
FGES	$\nu_0 = 0.1$	0.1413		0.7749		0.0		0.5000	
	$\nu_0 = 0.4$	0.2210		0.7749		0.0		0.5000	
	$\nu_0 = 0.7$	0.5436		0.7749		0.4404		0.5999	
GFCI	$\nu_0 = 0.1$	0.1413		0.8342		0.0		0.5000	
	$\nu_0 = 0.4$	0.2210		0.8342		0.0		0.5000	
	$\nu_0 = 0.7$	0.5436		0.8342		0.4304		0.5999	
RFCI	$\nu_0 = 0.1$	0.1413		0.7181		0.0		0.3750	
	$\nu_0 = 0.4$	0.2210		0.6342		0.0		0.4166	
	$\nu_0 = 0.7$	0.5436		0.7181		0.4317		0.5999	
RANDOM	$\nu_0 = 0.1$	0.1499		0.6221		0.0		0.0	
	$\nu_0 = 0.4$	0.1782		0.6221		0.0		0.0555	
	$\nu_0 = 0.7$	0.4571		0.6431		0.3485		0.6066	

the results of EVA by operator, showing which operator is affected more by the problem's features. *Impact of **B/W** configuration*. The results show no noticeable difference between the two cases in terms of relative gain of EVA. Note that, for a given dataset, the single solutions of the baselines in the two cases (**B/W**) are the same: in fact, the different hyperparameters in **B** and **W** affect only EVA, and the different size of the population in **B** and **W** affects just the average (i.e., the same solutions are averaged every 15, 30 or 60 solutions). In all the cases, the evolution across generations almost cancels the impact of the configuration, as the final results are very close to each other.
Appendix E gives details about the solutions of the last generation (i.e., the final solutions).

Best solutions. If one is interested more in getting a single best hypothesis for an effect to explain rather than on a set of good hypothesis, s/he can look at the *best* solution (rather than the average) of the population. Table 1 reports the distances of the best solutions of the last generation.[3] Note that in some cases (e.g., TUMOR), the results of FGES, GFCI and RFCI are the same, because the weights inferred by these algorithms were exactly the same. The best achieved distances highlight the ability to find solutions very close to (and often exactly matching) real occurred events, even for complex problems with many multiple potential causes – see Appendix F, Figure 7. For instance, in the NURSERY problem (with $\nu_0 = 0.1$), which has 9 variables, 136/150 (B) and 399/600 (W) solutions have 0 distance, i.e., cause-effect events are predicted exactly (42/300 and 73/600 for MEDICAL; 12/150 and 41/600 for TUMOR). In ASRS, it is much more difficult: in this case, EVA does not manage to generate matching solutions, but very close ones – e.g., several severe accidents, with, 27 co-occurring causes are predicted almost exactly (up to 24 out of 27 co-occurring causes).

Relative distance. A further metric of interest is the *relative* distance, d_{rel}. In fact, for complex problems like ASRS, one could be interested in assessing to what extent the hypothesized solution of a certain size k is at least partly "contained" in the solutions $z \in Z$. Given **q**$= q_1, q_2, \dots q_k$, d_{rel} is the Jaccard distance $d(q, z')$ between **q** and the same subset of k variables of **z**. Table 2 reports the results. Again, the gain of EVA is more evident for more complex problems.[4]. In some cases, the solutions with $\nu_0 = 0.7$ generated by the baselines have better relative distance; the main reason is that, with such a strict constraint, only few solutions survive, and these are typically small solutions, with few causes. This determines a small relative distance, but not a small absolute distance. The relative distance distribution of final solutions is in Appendix E.

7 Related work

Causal structure discovery (CSD) Causal inference is usually supported by graphical models, like causal networks. In this regard, CSD solutions are the closest to our approach. They infer possible cause-effect relations between sets of variables, typically associating a weight to the relations.

[3]For the baselines, since there is no notion of "evolution" and the single generations are independent of each other, we select the generation with the best population (best average distance) rather than the last one.

[4]Note that the objective is not to generate solutions with small relative distance, in which case it would be enough to generate small solutions, e.g., with one single cause. The objective is to generate solutions with small distance; the Table shows how often the so-generated solutions have small relative distance.

Table 2: Average *relative distance* of solutions of the best population – mean over 10 repetitions

		TUMOR		ASRS		MEDICAL		NURSERY	
		B	W	B	W	B	W	B	W
EVA	$\nu_0 = 0.1$	0.1070	0.1206	0.1956	0.3079	0.0	0.0	0.0053	0.0517
	$\nu_0 = 0.4$	0.2711	0.2844	0.2590	0.3382	0.0	0.0	0.2627	0.2964
	$\nu_0 = 0.7$	0.3972	0.4438	0.5667	0.5781	0.01555	0.0343	0.4838	0.4862
FGES	$\nu_0 = 0.1$	0.2309		0.5363		0.0066		0.4754	
	$\nu_0 = 0.4$	0.2294		0.5363		0.0067		0.4754	
	$\nu_0 = 0.7$	0.1337		0.5363		0.0278		0.4641	
GFCI	$\nu_0 = 0.1$	0.2309		0.5413		0.0120		0.4754	
	$\nu_0 = 0.4$	0.2294		0.5413		0.0125		0.4754	
	$\nu_0 = 0.7$	0.1337		0.5413		0.1473		0.4641	
RFCI	$\nu_0 = 0.1$	0.2309		0.6151		0.0068		0.4664	
	$\nu_0 = 0.4$	0.2294		0.6151		0.0071		0.4665	
	$\nu_0 = 0.7$	0.1337		0.6151		0.1112		0.4556	
RANDOM	$\nu_0 = 0.1$	0.3520		0.6789		0.1047		0.0967	
	$\nu_0 = 0.4$	0.3528		0.6789		0.1066		0.1054	
	$\nu_0 = 0.7$	0.3042		0.6790		0.1324		0.1050	

Algorithms are *score-based*, *constraint-based* or *hybrid* ((Glymour et al., 2019) for details) – for comparison, we took one representative algorithm per strategy to create a hypothesis generator, hence exploiting them for abduction as described in Section 5. Other interesting CSD strategies can be investigated, such as NoTEARS (Zheng et al., 2018), or CSD with reinforcement learning (Zhu et al., 2020), which however need to be adapted to make them able to cope with a CCOP.

Logic-based strategies. Abduction is usually dealt with logic-based approaches, like in the works cited in the introduction (Douven, 2017), (Crowder & Carbone, 2017), (Spohn, 2012); (Woods, 2013), (Walton, 2001), (Crowder et al., 2020), (Eshghi, 1988), (Ma, 2012)). Abductive logic programming (ALP) is one way to automate abduction; it combines logic programming and abduction to solve problems declaratively, given a background knowledge (P, i.e., logic program) and integrity constraints (IC) Kakas et al. (1992). An abductive explanation is derived as a set of hypotheses on abducible predicates (A) that satisfy the IC. Other approaches combine logic with probabilistic frameworks, such as abduction reasoning combined with Markov Logic Networks Schoenfisch et al. (2016), or Bayesian networks Raghavan & Mooney (2010). These find little applications in the real world, as they suffer from severe limitations: *i)* an excessive effort (and knowledge of the problem at hand) is required to write a sufficiently detailed model (i.e., list of known cause-effect clauses) – in EVA, the model is the ontology, hence just the list of cause and effect variables (i.e., the literals) possibly involved in the inference; *ii)* the computational explosion caused by large or over-specified models makes them inapplicable for large (and with few constraints) problems – e.g., in ASRS, where the space of cause-effect combinations is $\approx 24^{27}$.

Evolutionary strategies. EVA can be viewed from the side of evolutionary algorithms (EA). EAs are inspired by biological evolution to search for solutions optimization problems. The closest ones are human-inspired algorithms. Examples include: the *human-inspired algorithm* (HIA) (Zhang et al., 2009), the *human strategy* algorithm (HS) (Soltani-Sarvestani et al., 2018), the *cultural algorithms* (CA) (Reynolds, 1994). Such algorithms use a metaphor of a human-like behaviour. Beside the specific metaphor, the most important difference is that, in order to solve a CCOP, one needs to cope with an unknown admissible solution space to search for, which makes EVA a strategy to *generate* (rather than *search*) solutions in those problems where solutions can only be hypothesised and assessed afterwards (like in uncertain causal inference).

8 Conclusion

We introduced EVA, a new algorithm for automated abductive inference based on evolutionary computation. EVA is a first step toward a better understanding of how the potential of abduction can be exploited for problem solving. Such a form of reasoning, if empowered by automated computation, can boost the human ability to generate explanations/predictions in complex causal problems. Improvements and variants are possible along different ways (e.g., the operators, the plausibility function). At higher level, a natural direction is to integrate EVA with ML algorithms, which are massively based on induction on multiple observations: abduction can complement it just like in human reasoning, namely with the ability to go well beyond what observed, enabling a form of learning that can be more fallible but much quicker.

9 ETHIC STATMENET

Not applicable

10 REPRODUCIBILITY STATEMENT

The material reported in this paper, including:

- The datasets used for the experimentation;
- The source code (and the executable `.jar`) of the proposed algorithm and of the implemented baselines; the experimental code;
- The results reported in the main text as well as in the appendixes;

is available at **`http://github.com/eva-iclr-2021/EVA`**.

Instructions are provided in the repository to *reproduce* the same results of the paper, as well as to *replicate* the study with other datasets. Textual configuration files allow to select the datasets, to set EVA hyperparameters and experimental parameters (e.g., population size, novelty constraint), to set the initial seed (leaving the default and specifying 10 runs will reproduce the same result of the paper), to set the split (knowledge base and test set, leaving the default will will reproduce the same result of the paper). Bash scripts named `run.sh` expedite the process of reproducing the results of the paper, with one script in each dataset's folder for both EVA and for the baselines.

REFERENCES

Nasa aviation safety reporting system, a. URL https://asrs.arc.nasa.gov/index.html.

Nasa aviation safety reporting system – abbreviations, b. URL https://asrs.arc.nasa.gov/docs/dbol/ASRS_Abbreviations.pdf.

URL https://github.com/bd2kccd/py-causal.

a. URL https://archive.ics.uci.edu/ml/datasets/primary+tumor.

b. URL https://archive.ics.uci.edu/ml/datasets/nursery.

URL https://zenodo.org/record/3592985#.YVh7bC2uaYU.

URL https://www.ccd.pitt.edu/tools/.

Saurav Acharya. Causal modeling and prediction over event streams. 2014.

Floris Bex and Bart Verheij. Solving a murder case by asking critical questions: An approach to fact-finding in terms of argumentation and story schemes. *Argumentation*, 26(3): 325–353, 2012. doi: 10.1007/s10503-011-9257-0. URL https://doi.org/10.1007/s10503-011-9257-0.

Floris Bex and Bart Verheij. Legal stories and the process of proof. *Artificial Intelligence and Law*, 21(3):253–278, 2013. doi: 10.1007/s10506-012-9137-4. URL https://doi.org/10.1007/s10506-012-9137-4.

A. Ciampolini, P. Mello, and S. Storari. Distributed medical diagnosis with abductive logic agents. In *In Proc. of ECAI2002 workshop on Agents applied in health care*, pp. 23–32, 2002.

Diego Colombo, Marloes H. Maathuis, Markus Kalisch, and Thomas S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 40(1):294–321, 2012. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/41713636.

James A. Crowder and John Carbone. Abductive artificial intelligence learning models. In *Int'l Conference on Artifical Intelligence*, pp. 90–96, 2017.

James A. Crowder, John Carbone, and Shelli Friess. *Artificial Creativity and Self-Evolution: Abductive Reasoning in Artificial Life Forms*, pp. 65–74. Springer International Publishing, Cham, 2020. ISBN 978-3-030-17081-3. doi: 10.1007/978-3-030-17081-3_6. URL https://doi.org/10.1007/978-3-030-17081-3_6.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

Igor Douven. Abduction. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition, 2017.

Kave Eshghi. Abductive planning with event calculus. In *Proc. of the fifth International Conference and. Symposium on Logic Programming (ILPS 1988)*, pp. 562–579, 01 1988.

P. Flach and A. Kakas. *Abduction and induction*. Dordrecht: Kluwer, 2000.

A. Frank and A. Asuncion. Uci machine learning repository., 2010. URL http://archive.ics.uci.edu/ml.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524, 2019. ISSN 1664-8021.

J. Josephson and S. Josephson. *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press, 1994.

Antonis Kakas, Robert Kowalski, and Francesca Toni. Abductive logic programming. *J. Log. Comput.*, 2:719–770, 12 1992. doi: 10.1093/logcom/2.6.719.

M. Bohanec M. Olave, V. Rajkovic. *Expert Systems in Public Administration*, chapter An application for admission in public school systems, pp. 145–160. Elsevier Science Publishers (North Holland), 1989.

Jiefei Ma. *Distributed abductive reasoning : theory, implementation and application*. PhD thesis, Imperial College London, UK, 2012. URL http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.544253.

L. Magnani. *Abduction, Reason and Science*. Springer, 2001.

Lorenzo Magnani. *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*. Springer, 2010.

Juan Miguel Ogarrio, Peter Spirtes, and Joe Ramsey. A hybrid causal search algorithm for latent variable models. In Alessandro Antonucci, Giorgio Corani, and Cassio Polpo Campos (eds.), *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, pp. 368–379, 2016.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.

Charles S. Peirce. *The Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, 1931.

G. Preyer and D. Mans. On contemporary developments in the theory of argumentation. *Protosociology*, 13:3–13, 1999.

Júlia Pukancová and Martin Homola. Abductive reasoning with description logics: Use case in medical diagnosis. *Description Logics*, 2015.

Sindhu Raghavan and Raymond Mooney. Bayesian abductive logic programs. In *Proceedings of the 6th AAAI Conference on Statistical Relational Artificial Intelligence*, AAAIWS'10-06, pp. 82–87. AAAI Press, 2010.

Joseph Ramsey. Scaling up greedy equivalence search for continuous variables. *ArXiv*, abs/1507.07749, 2015.

Joseph Ramsey, Kun Zhang, Madelyn Glymour, Ruben Sanchez Romero, Biwei Huang, Immé, Ebert-Uphoff, Savini M. Samarasinghe, Elizabeth A. Barnes, and Clark Glymour. Tetrad - a toolbox for causal discovery. 2018.

R. G. Reynolds. An introduction to cultural algorithms. In Anthony V. Sebald and Lawrence J. Fogel (eds.), *Evolutionary Programming — Proceedings of the Third Annual Conference*, pp. 131–139. World Scientific Press, 1994.

J. Schoenfisch, J. von Stulpnagel, J. Ortmann, C. Meilicke, and H. Stuckenschmidt. Root cause analysis through abduction in markov logic networks. In *2016 IEEE 20th International Enterprise Distributed Object Computing Conference (EDOC)*, pp. 1–8, 2016. doi: 10.1109/EDOC.2016.7579386.

G. Schurz. Patterns of abduction. *Synthese*, 164(2):201–234, 2008. doi: 10.1007/s11229-007-9223-4.

M. A. Soltani-Sarvestani, Zohreh Azimifar, and Ali Hamzeh. Human strategy (hs) optimization algorithm. *Soft Computing*, 22(3):715–735, 2018. doi: 10.1007/s00500-017-2484-z. URL `https://doi.org/10.1007/s00500-017-2484-z`.

W. Spohn. *The Laws of Belief: Ranking Functions and Their Applications*. Oxford University Press., 2012.

Matthew J. Vowels, Necati Cihan Camgoz, and Richard Bowden. D'ya like dags? a survey on structure learning and causal discovery, 2021.

Douglas Walton. Abductive, presumptive and plausible arguments. *Informal Logic*, 21(2), 2001.

J. Woods. *Errors of Reasoning: Naturalizing the Logic of Inference*. London: College Publications, 2013.

L. M. Zhang, C. Dahlmann, and Y. Zhang. Human-inspired algorithms for continuous function optimization. In *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*, volume 1, pp. 318–321, 2009.

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. Dags with no tears: Continuous optimization for structure learning, 2018.

Shengyu Zhu, Ignavier Ng, and Zhitang Chen. Causal discovery with reinforcement learning, 2020.

Supplementary material

The supplementary material for the paper entitled "Automated Hypotheses Generation via Evolutionary Abduction" includes *textual material* and *artifacts*. Textual material is in the following Appendixes A-G. Artifacts includes the source code (and the executable `.jar`) of the proposed algorithm and of all the implemented baselines, the experimental code, the datasets used for the experimentation, and the results reported in the main text and in the following appendixes. These are available at: **`http://github.com/eva-iclr-2021/EVA`**.

The following textual supplementary material is organized as follows. First, the algorithms of the three abduction operators are described (Appendix A). Appendix B describes the baseline strategies we have implemented to solve the causal problem by causal structure discovery algorithms followed by sampling. Appendix C reports the results of the tuning of the parameters used in the experimentation. These refer to both the EVA hyperparameters and to the size of the population used in the experimental study. A best and worst case for EVA are derived, then used in the final experimentation reported in the main text. Appendix D reports the results achieved by the three abductive operators of EVA, which together contribute to the overall performance of EVA. Appendix E reports the distribution of the distances of the last generation's solutions, namely of the final solutions. In particular, the distributions of the average and of the best distance (over the distances of the final population's solutions) are shown, as well as for the *relative distance*. Appendix F shows the best solution (namely, the solution with the best distance) of the populations at every generation, averaged over the 10 repetitions. Finally, Appendix G details the ASRS dataset, which, unlike the other datasets, is prepared from scratch starting from the ASRS database.

A The Evolutionary Abduction operators

Algorithms 2-4 are the *factual*, *analogical* and *hypothetical cause* operators of EVA described in Section 3.

Algorithm 2 is the *factual* operator. It takes, as input, a solution **x**, chosen by the selection operator (`select_factual`), all the different sources and targets (i.e., causes and effects) that are in the current population (S and T), and considers the KB to build a new solution. To build the new solution **x**', first, a target t is selected from the list of all the targets in the current population (line 1). Selection of the target is done taking two targets randomly, measuring their "support" (number of occurrences in KB) and taking the one with greater support or choosing randomly (with probability 0.5) one of the two if they have equal support. In essence, it is a binary tournament applied to single elements rather than to the whole solution. As for the sources, the same sources of **x** are used (line 2). The operator applies three types of modifications: `add`, `modify` or `delete` actions. It considers two parameters to regulate the extent of changes and the desired novelty: an integer called *factual change index* $\gamma_F > 0$ and a double called *factual novelty index*, $\eta_F \in [0; 1]$. The number of changes c to apply are selected randomly, with $c \in [1; \gamma_F]$ (line 3). The type of change (add, modify or delete) is also selected randomly with equal chance for the three actions (line 5). In case of *add* or *modify* (which is a replacement of a source), the new source is selected from the set S with probability η_F or from the KB with probability $1 - \eta_F$. Selection of the sources to add, replace or remove (lines 7, 11, 14, respectively) is done by a variable-level binary tournament like the above-mentioned target selection, so as to favour the sources/targets contributing more to plausibility.

Algorithm 3 presents the *analogical* operator. To build the new solution **x**', the operator first selects a target from T, just like the *factual* operator algorithm (i.e., via a variable-level binary tournament). Then, it builds the set of sources, coupled with the chosen target, by extracting and reproducing *structural* features of the sources in **x** (`extractConstraint`, line 2). Three source-level constraints are defined currently in EVA, which require the new solution **x**' to have progressively stronger similarities with **x**:

- *Cardinality constraint*: the number of sources of **x**' is required to be the same as **x**. The cardinality is a "proxy" indicator for the complexity of a solution, since more sources means co-occurrences of more causes together for an effect. The selection of the sources for **x**' is done by considering an analogical novelty index: $\eta_A \in [0; 1]$. The source to be added is selected from the set S with probability η_A, or from the ontology Ω with probability $1 - \eta_A$

Algorithm 2: factual_operator(x, S, T)

Input : **x**, the selected solution; S/T, all different sources/targets in the current population; η_F, Factual novelty index; γ_F, Factual change index

$t \leftarrow$ selectTarget(T);
x' = $\{$**x**, $t\}$; ▷ *initialize x' with the same sources as x, and target t*
$c \leftarrow$ Rand(1, η_F); ▷ *Number of changes*
for *i=1 to c* **do**
  $a \leftarrow$ Rand(*add, modify, delete*) ▷ *Action to apply*
  if *a=add* **then**
    $s \leftarrow$ selectSource(η_F); ▷ *η_F: Prob. to select from S or from M*
    addSource(**x**', s);
  if *a=modify* **then**
    removeSource(**x**');
    $s \leftarrow$ selectSource(η_F);
    addSource(**x**', s); ▷ *with s different from removed source*
  if *a=delete* **then**
    removeSource(**x**');
return **x**';

Algorithm 3: analogical_operator(x, P, S, T)

Input : **x**, the selected solution (from M_E), P, population; S/T, all different sources/targets in the current population; η_F, Analogical novelty index

; $t \leftarrow$ selectTarget(T);
[p, v_g, σ_{M_g}] = extractConstraints(); ▷ *Extract #sources (p), #sources per group (v_g), σ_{M_g} per group*
for *i=1 to p* **do**
  $s \leftarrow$ selectSource(η_A); ▷ *η_A: Prob. to select from S or from Ω*
  addSource(**x**', s);
while *(v_g and σ_{M_g} constraints are not satisfied)* **do**
  replaceSource(**x**'); ▷ *Adjust the solution to meet constraints*
return **x**';

Algorithm 4: hypothetical_operator(x, S, T)

Input : **x**, the selected solution; S/T, all different sources/targets in the current population; η_H, Hypothetical cause novelty index; γ_H, Hypothetical cause change index

$t \leftarrow$ selectTarget(T);
x' = $\{$**x**, $t\}$; ▷ *initialize x' with the same sources as x, and target t*
$c \leftarrow$ Rand(1, η_H); ▷ *Number of changes*
for *i=1 to c* **do**
  $a \leftarrow$ Rand(*add, modify, delete*) ▷ *Action to apply*
  if *a=add* **then**
    $s \leftarrow$ selectSource(η_H); ▷ *η_H: Prob. to select from S or from Ω*
    addSource(**x**', s);
  if *a=modify* **then**
    removeSource(**x**');
    $s \leftarrow$ selectSource(η_H);
    addSource(**x**', s); ▷ *with s different from removed source*
  if *a=delete* **then**
    removeSource(**x**');
return **x**';

(line 4). Selecting from the ontology rather than from the current population means that the agent intends to exploit new concepts not previously observed, taking from the entire knowledge about the domain of interest.

- *Group membership constraint*: in many causal problems, the joint causes available to explain an effect can be grouped in homogenous subsets (e.g.: all stress-related causes in medical diagnosis to explain a disease; or, in hazard analysis, all the environment-related events possibly causing an accident). In such cases, this constraint requires the new solution **x**' to have the same structure as **x**, namely the same number of subsets with the same cardinalities. For each group in **x**' there must be one distinct group in **x** with the same cardinality and viceversa.
- *Ordinal constraint*, uses the notion of support referred to subsets of **s** (rather than to entire solutions as in Def. 4):

  Definition 7 (***Support of order*** k ***and maximum support***)**.** The support of order k, $\sigma_k(\mathbf{q})$, of a subset of sources $\mathbf{q} \subseteq \mathbf{s}$, with $k \leq |\mathbf{q}|$, is the number of distinct k-tuples of **q** that occurr at least once in the solutions of the memory M. The *maximum degree of support*, $\sigma_M(\mathbf{q})$, of $\mathbf{q} \subseteq \mathbf{s}$ is the maximum value of k such that $\sigma_k(\mathbf{q}) > 0$.

  The *ordinal constraint* requires **x**' to have the same number of subsets with the same σ_M values of the subsets of **x**. For each pair of groups g'_i, g'_j in **x**' with $\sigma_M(g_i) > \sigma_M(g_j)$, there must be a distinct pair of group in **x**, g_i, g_j with the same relation, $\sigma_M(g_i) > \sigma_M(g_j)$ and viceversa. For implementing constraint 2 and 3, the sources added to match constraint 1 (line 4 of the Algorithm) are replaced in those (pairs of) groups that violate constraint 2 and/or 3 (line 6-7), until the constraint is met or a maximum number of attempts is reached.

Algorithm 4 is the *hypothetical-cause* abduction operator. This operator mimics the creative abduction allowing a human to advance hypotheses exploiting just his knowledge about the domain of interest (i.e., the ontology Ω). The initial solution **x**' is build like in the factual abduction. The operator also applies the same actions as the factual operator: *add*, *modify* or *delete*, again exploiting parameters to regulate the extent of changes and novelty (*hypothetical change index* γ_H and *factual novelty index*, $\eta_H \in [0; 1]$). The main difference lies in considering Ω in lieu of KB as set from which a source can be selected, thus opening to a wider range of novel solutions.

A consequence is that these solutions are expected to have higher novelty compared to the factual operator, contrasted by a lower plausibility. And this is what actually happens by adopting such two types of reasoning: while factual abduction supports more plausible but less original inference, hypothetical abduction, by its nature, is open to completely new scenarios but whose plausibility can be low. Analogical abduction lies in between; in fact, it is also called a *partially* ampliative inference. Although one can focus on just one of these operators in custom implementations of EVA, the suggestion is to exploit all the three operators for their complementarity.

B GRAPH-BASED BASELINE STRATEGIES

The graph-based (GB) strategies have been implemented as follows. A Causal Structure Discovery (CSD) algorithm is used to learn the causal structure from the knowledge base KB; the output is directed acyclic graph (DAG) with nodes being the variable and arcs being dependency relation between them (Pearl, 2009). This is exploited to generate solutions proportional to cause-effect strength as described hereafter.

The CSD algorithms, namely FGES (Ramsey, 2015), RFCI (Colombo et al., 2012), and GFCI (Ogarrio et al., 2016), are all present in the *py_causal* repository (Vowels et al., 2021)(PYC)ZEN, which exploits the `Tetrad` toolbox Ramsey et al. (2018)tet. The parameters setting to derive the DAG and the corresponding arc weights are in Table 3 – the default parameters are kept, except the number of bootstraps (i.e., number of resampling) raised to 50 to improve the accuracy. The data type is always “discrete”. The description of each field can be found at `http://cmu-phil.github.io/tetrad/manual/`:

As prior knowledge, we specified (by the `priorKnowledge` parameter) that arcs between causes should be forbidden, as we are interested in arcs between causes and effects. The weights between arcs from causes to the effect obtained for the four datasets (values in the repository,

Table 3: ...

	FGES	GFCI	RFCI
scoreId	bdeu-score	bdeu-score	–
testId	–	disc-bic-test	bdeu-test
maxDegree/depth	3	3	3
faithfulnessAssumed	True	True	–
numberResampling	50	50	50
resamplingEnsemble	1	1	1
maxPathLength	–	-1	-1
completeRuleSetUsed	–	False	False
addOriginalDataset	True	True	True

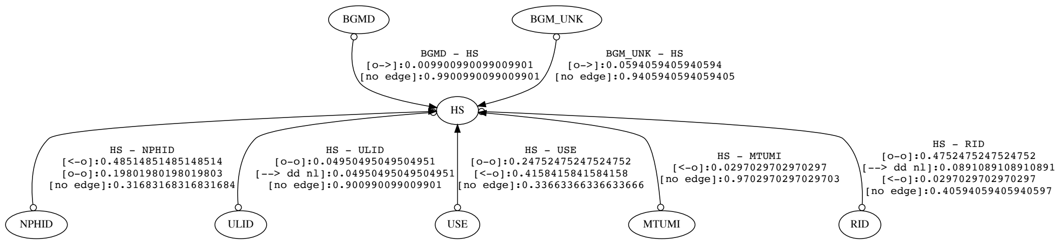


Figure 2: Example of DAG used for the RFCI GB strategy

`http://github.com/eva-iclr-2021/EVA`), which represent the probability that a potential *cause* node is causally related to the *effect* node, are used to generate the solution. An example of so-obtained DAG is in Figure 2, wherein *HS* is the effect (Hypoglycemic symptoms) and all the other variables are potential causes such as “More-than-usual meal ingestion” (MTUMI), “Blood Glucose Measurement Decrease” (BGMD). This is obtained by the RFCI algorithm.

Given the graph and their cause-effect weights $W = \{w_i\}$, $i = 1, \ldots, n$ and n being the number of (source) variables, the implemented generator acts as follows: for each instance to generate ***i)*** includes variable v_i ($i = 1, \ldots, n$) with probability w_i as part of the solution, and then ***ii)*** selects a value j of the variable v_i, say $h_{i,j}$, proportionally to the estimate of its probability of occurrence $p_{i,j}$ within the KB – obtained as (normalized) relative frequency of that value within KB.

The random strategy just the variables v_i ($i = 1, \ldots, n$) with equal probability of selection, and then the values $h_{i,j}$ of v_i with equal probability of selection.

C Parameters tuning

A grid search approach is adopted for parameters tuning. The EVA hyperparameters are the novelty indexes, η_F, η_A and η_H, and the change indexes γ_F and γ_H of the abduction operators. Both regulate the extent to which solutions are required to be diverse (hence novel) with respect to the KB and to the current population: the higher the $\eta.$ values, the higher the probability of selecting new unseen sources, and the higher the $\gamma.$ the higher the number of modifications that are done to build a (factual or hypothetical-cause) solution. The following configurations are considered: $< \eta., \gamma. >= (< 0.1, 3 >, < 0.5, 5 >, < 0.9, 7 >$, representing, respectively, a *Low* novelty degree in the solution, a *Medium* novelty and a *High* novelty.

Additionally, due to its evolutionary nature, EVA exploits the notion of population of solutions, whose *size* can impact the final results. Three values are considered for the population size:$|P| = (15, 30, 60)$.

We ran 10 repetitions for each of the $3 \times 3 = 9$ configurations, each one for 600 evaluations, for the four datasets. Table 4 reports the average distance of the final population’s solution (averaged over the 10 repetitions) from the test set. The **best (B)** and **worst (W)** configurations for EVA are

Table 4: Average distance (standard deviation) of solutions of the best population – mean over 10 repetitions

		$\|P\| = 15$	$\|P\| = 30$	$\|P\| = 60$
TUMOR	Low	$0.1680_{0.0238}$	$0.1821_{0.0221}$	$0.2118_{0.0155}$
	Medium	$0.1648_{0.0240}$	$0.1755_{0.0154}$	$0.2099_{0.0205}$
	High	$0.1620_{0.0252}$	$0.1652_{0.0335}$	$0.2034_{0.0332}$
ASRS	Low	$0.5407_{0.0259}$	$0.5833_{0.0118}$	$0.6376_{0.0185}$
	Medium	$0.4803_{0.0203}$	$0.5204_{0.0181}$	$0.5961_{0.0191}$
	High	$0.5162_{0.0178}$	$0.4971_{0.0314}$	$0.5560_{0.0196}$
MEDICAL	Low	$0.3584_{0.0209}$	$0.3651_{0.0197}$	$0.3631_{0.0175}$
	Medium	$0.3629_{0.0212}$	$0.3421_{0.0169}$	$0.3677_{0.0134}$
	High	$0.3557_{0.0341}$	$0.3582_{0.0179}$	$0.3615_{0.0108}$
NURSERY	Low	$0.0742_{0.0479}$	$0.0806_{0.0306}$	$0.1292_{0.0193}$
	Medium	$0.0850_{0.0270}$	$0.1101_{0.0309}$	$0.1517_{0.0297}$
	High	$0.1253_{0.0187}$	$0.1336_{0.0300}$	$0.1723_{0.0196}$

highlighted (green and red, respectively). These two configurations are used to compare EVA with the baselines (over 6,000 evaluations) (cf. with Section 6), considering both the best and the worst case.

D Results by EVA operator

Figure 3 reports the average distance of the final solutions computed by each of the three operators of EVA (i.e.: *Factual, Analogical, Hypothetical-cause*), in every run and experimental scenario (10 runs per scenario) for every dataset.

Two main observations arise: *i)* the *Factual* and *Hypothetical-cause* abduction operators give similar distance values for all scenarios and datasets. These, in fact, have the same structure, the main difference is in the source of knowledge used (the former relies on the KB, while the latter on the ontology Ω); *ii)* the *Analogical* operator works better (i.e., small distances) than the others for small problems, namely when few multiple causes are involved, which is the case of the MEDICAL and NURSERY datasets; in contrast *Factual* and *Hypothetical-cause* outperform the *Analogical* operator for ASRS and TUMOR. The impact of the Best/Worst configuration is negligible, as it does not change the relative results. A higher novelty constraint up to $\nu_0 = 0.7$ causes the operators' results to flatten on values above 0.6, as it becomes difficult for all the operators to find close-to-real solutions that are also very different from the KB. The only exception is the case of NURSERY, where the analogical operator still manage to give solutions with distance around 0.5 even with such a strict constraint on the novelty.

Although in one specific problem one operator may provide better solutions, for EVA to work reasonably well with various problems of different size, the suggested strategy is to always exploit the contribution of all the three operators. This also ensures a better diversity of the obtained solutions.

E Distribution of solutions

Figure 4 reports the percentage of solutions of the final generation's population with average distance less than or equal to a given value – the average over 10 repetitions is reported. For the baseline strategies, since there is no notion of "evolution" and runs (i.e., generations) are independent of each other, we do not consider the final generation, but select the generation with the best population (i.e., having solutions with the best average distance). Results are broken down by novelty constraint and by configuration (best: **B**, worst: **W**).

EVA generates considerably more solutions in the left side of the histogram (i.e., closer to 0) for all the cases. In terms of datasets, the gain is more evident for more complex problems (ASRS, TUMOR), but also for problems with a may instances in the test set (NURSERY), while it becomes less evident for MEDICAL. Again, the Best/Worst configuration makes no relevant difference. With

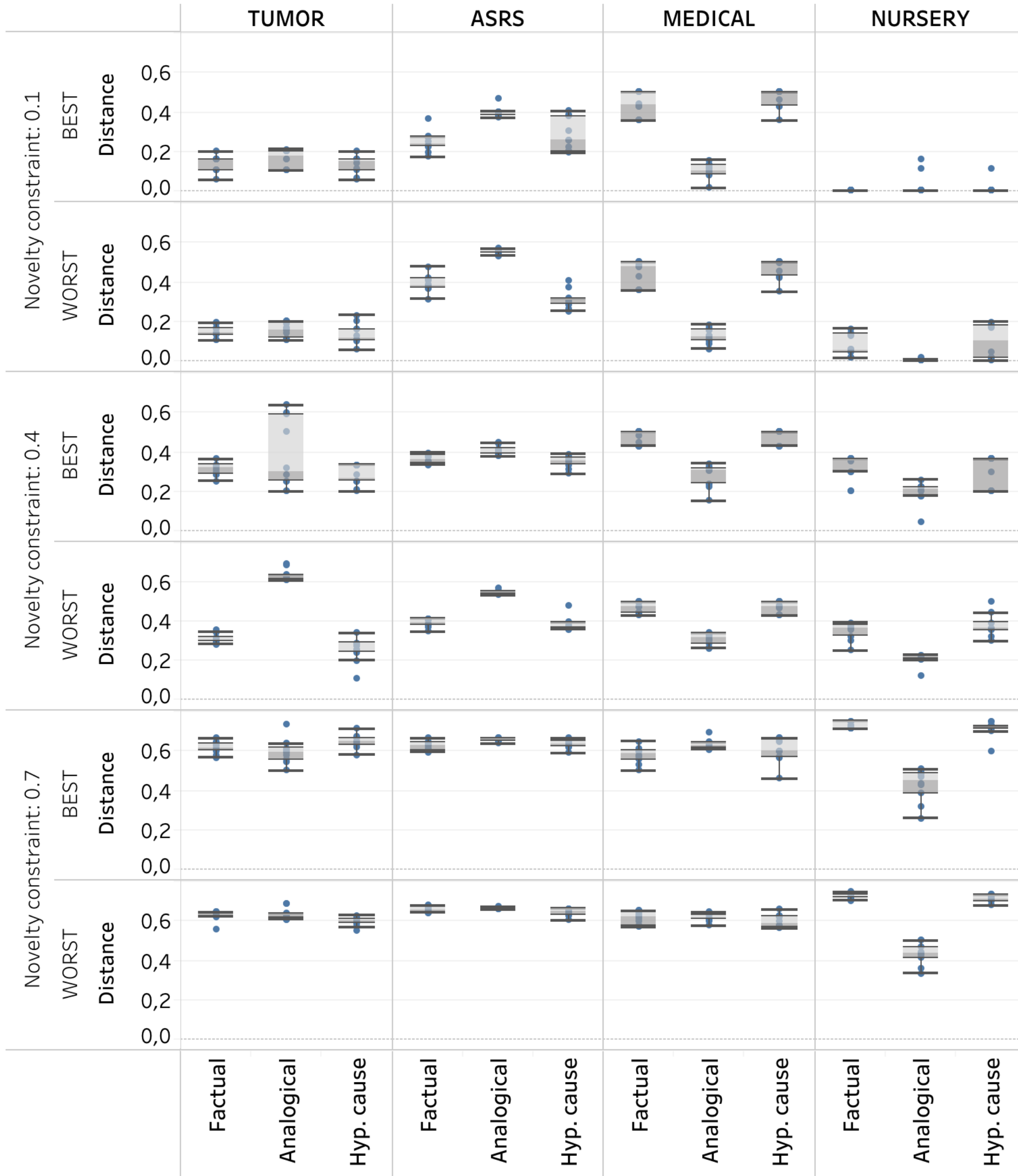
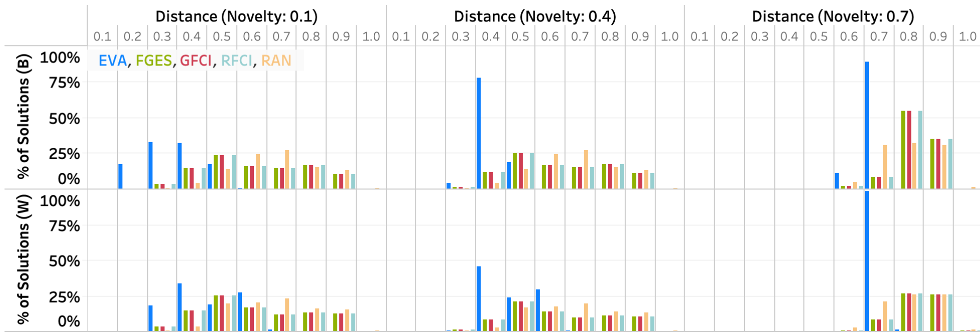


Figure 3: Results by operators

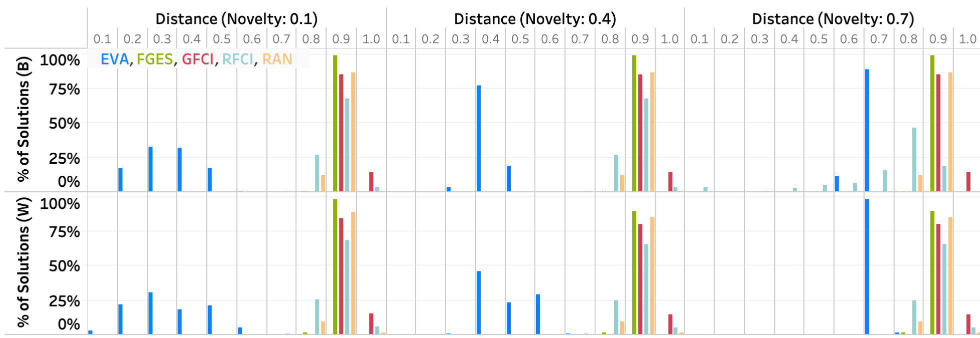
the increase of the novelty constraint the gain of course reduces, as there is less margin for improving over a random or graph-based strategy.

Figure 6 reports the same results but for the *relative* distance (cf. with Section 6). There are many cases in which solutions with relative distance equal to 0 are generated, namely solutions in which the set of causes is entirely contained in the se of causes of a real occurred event (an entry in the test set). For instance, in the MEDICAL dataset, many of the generated solutions (by all the techniques) have relative distance equal to 0[5]. For what said in Section 6, the gain of EVA in terms of relative distance is when the novelty constraint is at 0.1 and 0.4, not at 0.7.

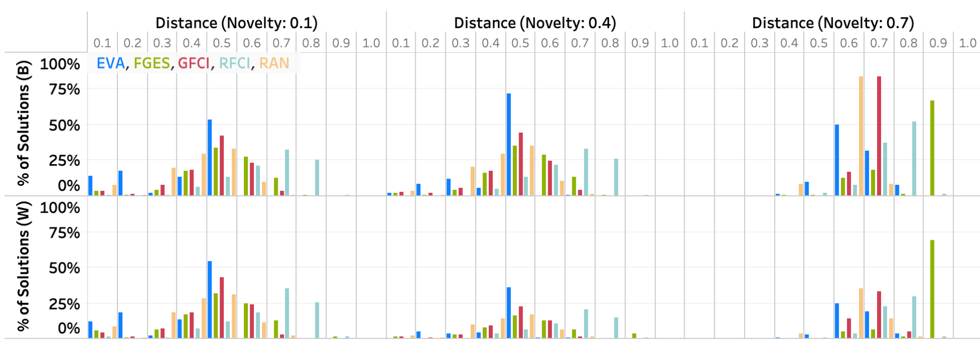
[5]Note that the objective is not to generate solutions with small relative distance, in which case would be enough to generate small solutions, e.g., with one single cause. The objective is to generate solutions with small absolute distance; this graph shows how often the so-generated solutions have small relative distance.



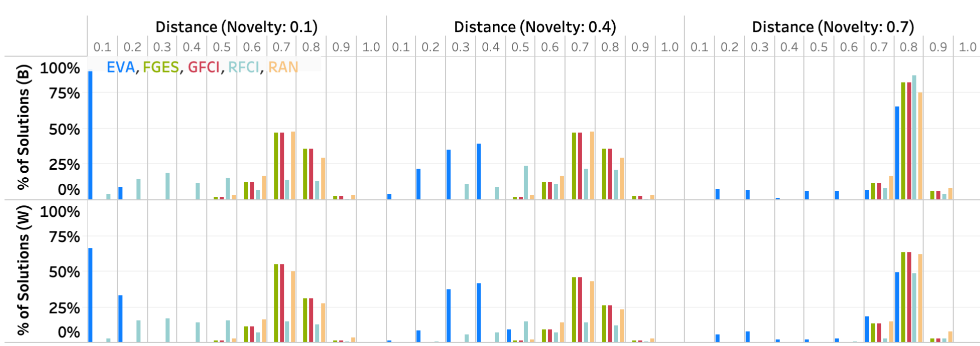
(a) TUMOR problem



(b) ASRS problem

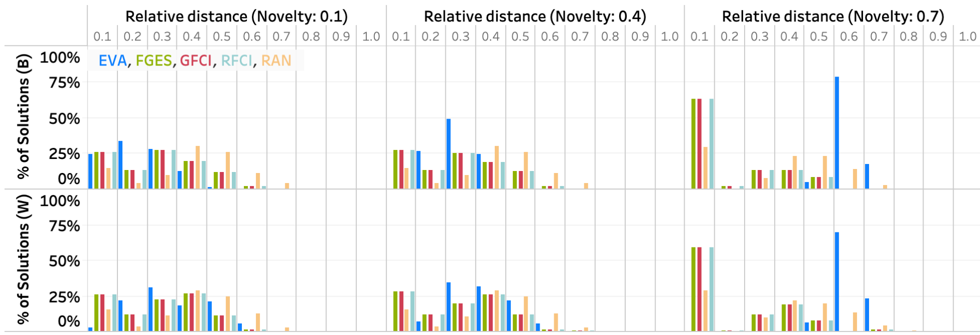


(c) MEDICAL problem

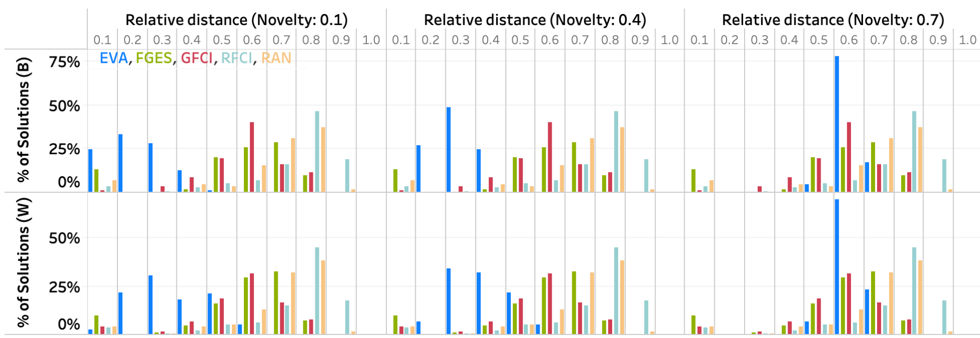


(d) NURSERY problem

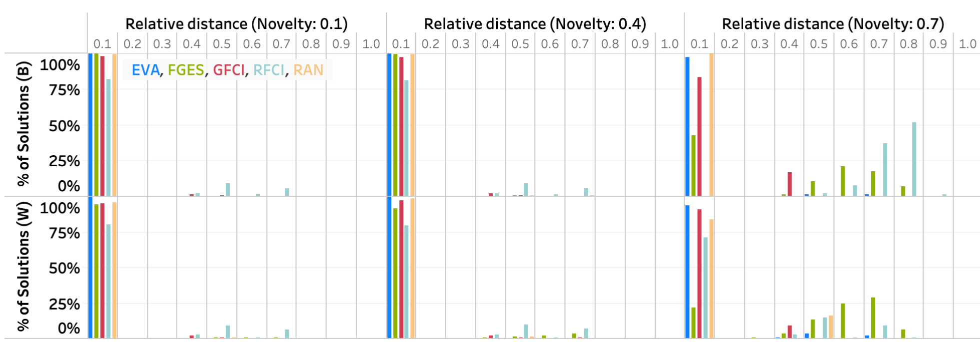
Figure 4: Distribution of solution's distance



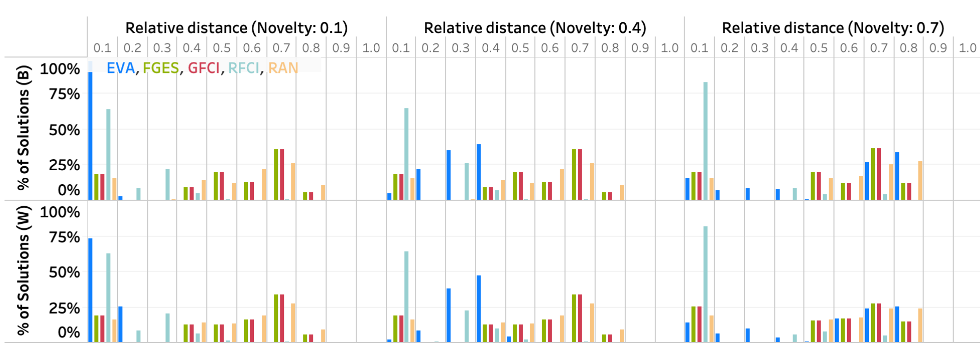
(a) TUMOR problem



(b) ASRS problem

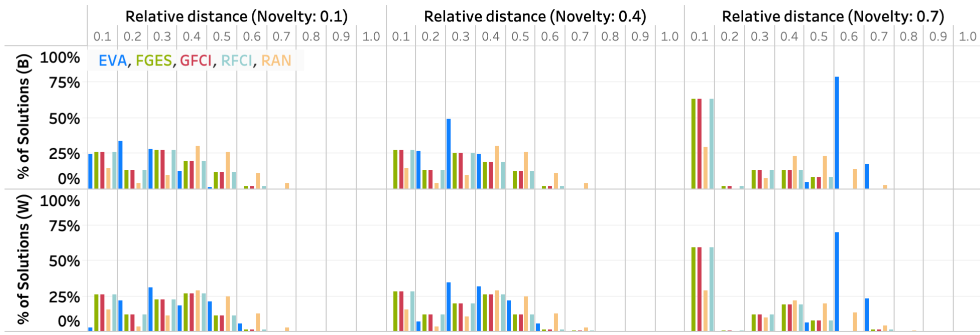


(c) MEDICAL problem

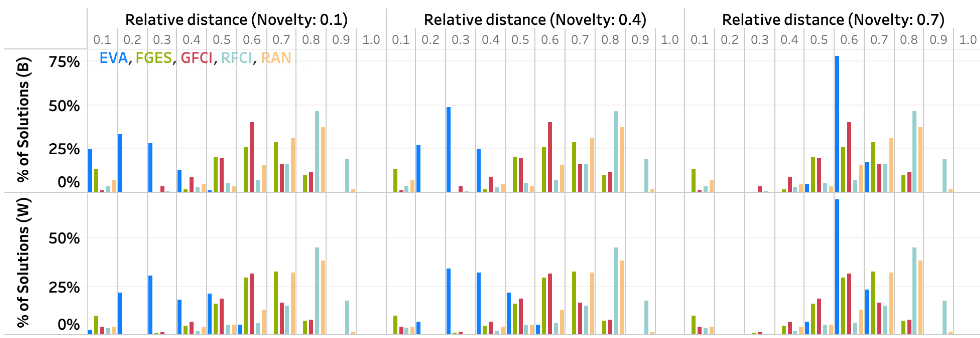


(d) NURSERY problem

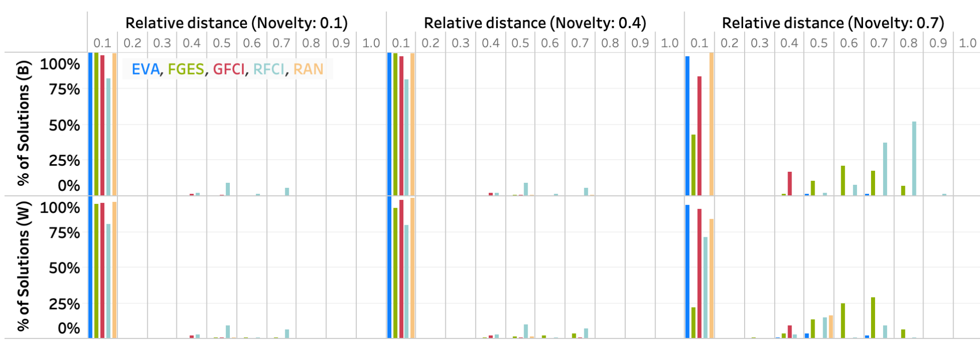
Figure 5: Distribution of solution's best distance



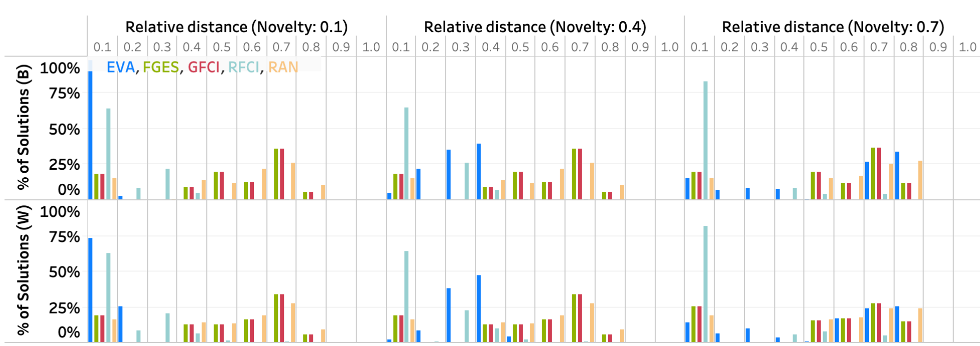
(a) TUMOR problem



(b) ASRS problem

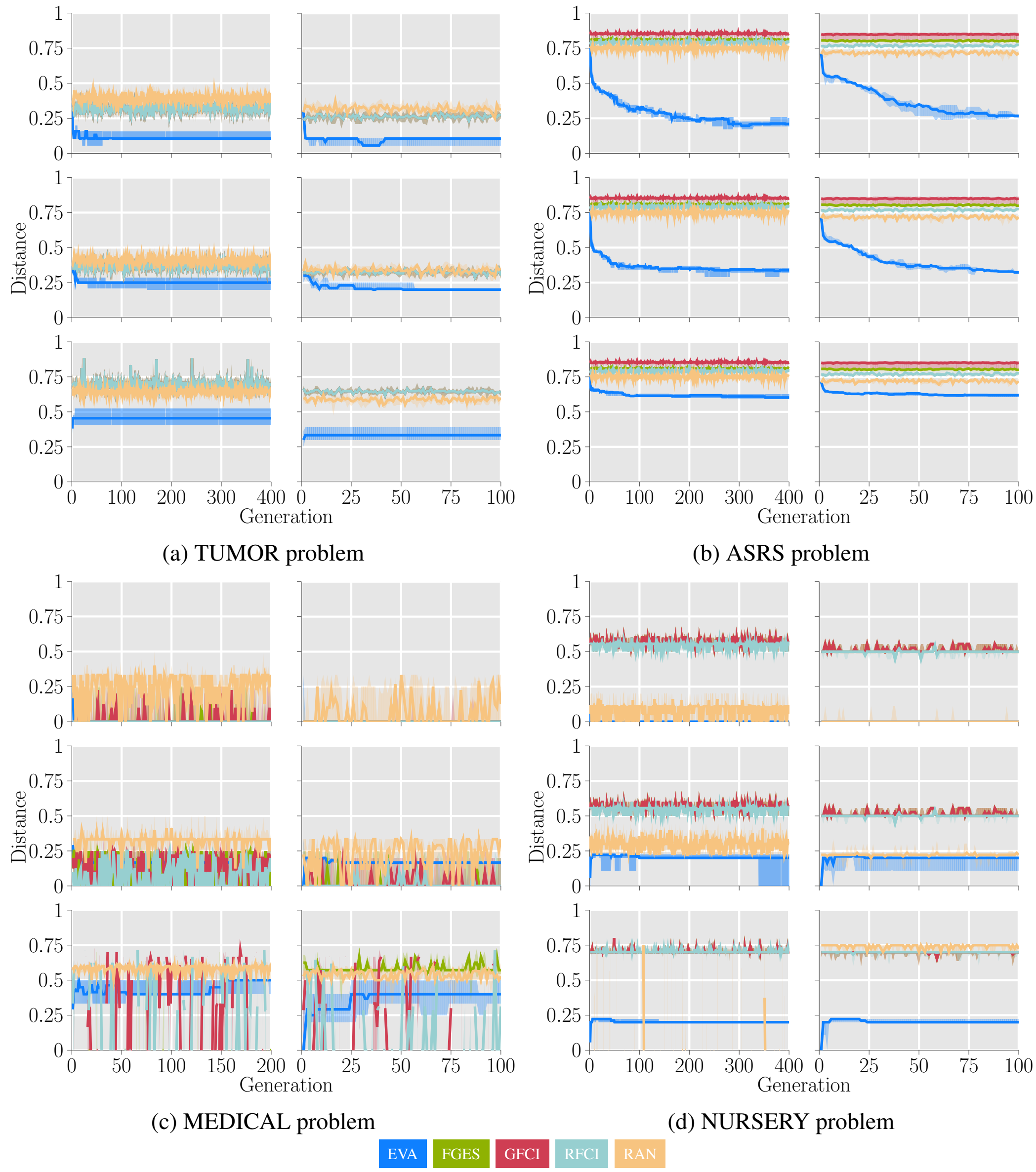


(c) MEDICAL problem



(d) NURSERY problem

Figure 6: Distribution of solution’s relative distance



(a) TUMOR problem

(b) ASRS problem

(c) MEDICAL problem

(d) NURSERY problem

Figure 7: *Best distance* by generation.

F Best solutions by generation

Figures 7 reports the *best* distance of the population's solutions *vs.* generations (median and IQR over 10). These are the same type of graph as Figure 1 in Section 6, but here the best solution of the population at every generation is considered. The evolution across generations leads to the final results summarized in Table 1 in Section 6. The distances are of course smaller than the average distances of Figure 1. In the case of ASRS, EVA gives distances that still decrease after 6,000 evaluations – it can still improve in that case, while in other cases it converged. When EVA is not visible in the graph (e.g., MEDICAL and NURSERY) it means the distances are 0. Finally, in the case of $\nu_0 = 0.7$, it often happens that the baselines do not provide solutions for some generations.

Table 5: ASRS. *Environment* entity.

Environment				
Flight conditions	Weather Elements/ Visibility	Work Env. Factors	Light	Ceiling
VMC	Cloudy	Poor lighting	Dawn	CLR
IMC	Fog	Glare	Daylight	Single value
Mixed	Hail	Temperature extreme	Dusk	
Marginal	Haze-Smoke	Excessive humidity	Night	
	Icing			
	Rain			
	Snow			
	Thunderstorm			
	Turbolence			
	Windshear			
	Other			

Table 6: ASRS. *Aircraft* entity.

Aircraft						
Flight plan	Flight Phase	Route in use	Navigation in use	Cabin Lighting	Maintenance status & items	Mission
VFR	Taxi	Direct	FMS/FMC	High	Deferred	Aerobatics
IFR	Parked	Oceanic	GPS	Medium	Records	Agricolture
SVFR	Takeoff	VFR Route	INS	Low	complete	Ambulance
DVFR	Initial	Vectors	Localizer/	Off	Released	Banner tow
None	climb	Visual appr.	Gideslop/ILS		for serv.	Ferry
	Climb	None	NDB		Required	Cargo/Freight
	Cruise	Airway	VOR/VORTAC		Scheduled	Passenger
	Descent	STAR			Unscheduled	Photo shoot
	Initial Appr.	SID				Personal
	Final Appr.	Other			*Maintenance items*	Refueling
	Landing				Inspection	Skydiving
	Other				Installation	Tactical
					Repair	Test Flight
					Testing	Traffic watch
					Work cards	Training
						Utility
						Other

G THE ASRS DATASET

While the TUMOR, MEDICAL and NURSERY datasets were already publicly available and explained, the ASRS dataset is new. Here we briefly describe the source of information from which the dataset is derived.

The Aviation Safety Reporting System (ASRS) database is the world's largest repository of voluntary, confidential safety information provided by aviation personnel, including pilots, controllers, mechanics, flight attendants and dispatchers ASR (a).

It contains more than 1 million of entries reported since 1988. It is a structured database used for data retrieval and analysis, with all the accidents stored in a cause-effect style: the events regarding the aircraft components, the weather conditions, the human personnel involved, the airport, and many other potential causes recorded for each accident as a categorised set of values (i.e., enumerative), along with the resulting accident (also categorised). The main entities are reported in the following:

- **Environment**, with information regarding the flight conditions when accident occurred, visibility, working environment factors such as lighting or temperature.

Table 7: ASRS. *Component* entity

Component		
Component		Problem
Weather Radar	Electrical Wiring & Connectors	Design
DC Battery	Autopilot	Failed
Turbine Engine	Landing Gear	Improperly operated
Indicating and Warning - Landing Gear		Malfunctioning
Nose Gear	Yaw Control	
Flap Vane	Brake System	
Powerplant Fire Extinguishing	Wheels/Tires/Brakes	
Cockpit Window	Aircraft Cooling System	
Turbine Assemb Blade	Landing Gear Indicating System	
Normal Brake System	Tires	
Gear Down Lock	Fuel System	
Engine Control	Fire/Overheat Warning	
Antiskid System	Piston	
Fuselage Skin	Powerplant Fuel Control	
External Power	Flap Control	
Supplemental Landing Gear	FCC (Flight Control Computer)	
Fuselage Panel	*(more than 350)*	
Engine	...	

- **Aircraft**-related elements, e.g., the flight plan, the route, the flight phase, the maintenance status, the mission.
- **Component**, with information about all the components of the aircraft and their status (e.g., design problem, failed, malfunctioning).
- **Person**, reporting the information about the persons involved, such as the flight crew, the air traffic control, or people working in maintenance, information about the human factors that could cause mistakes such as distraction, confusion, stress, etc.
- **Events**, including anomalies such as airspace violation, deviation of altitude, procedural errors, airbone or ground conflict, fire, as well as the event describing the final result, such as the type of accident and its consequences (which correspond to our target variables).

An excerpt of the main information is reported in the Tables 5-9. A glossary of terms is available on the website ASR (b). For illustrative purpose, a solution looks like follows:

```
Environment.Weather = Fog
Environment.Weather = Windshear
Environment.Weather = Turbulence
Environment.FlighConditions = IMC
Environment.Light = Night
Aircraft.Mission = Cargo/Freight
FlightAircraft.Phase =  Final Approach
Anomaly.Inflight Event = Object encountered
Result.Flight Crew = Landed in
Emergency Condition
Result.Aircraft = Aircraft Damaged
```

This describes an accident in which the pilot, while descending to approach for landing (`Final Approach`) during the night and under bad weather conditions (`IMC` stands for Instrument Meteorological Conditions as opposed to Visual Meteorological Conditions), struck a tree branch (`Object encountered`) and damaged the wing. Hence, he diverted to another airport, landing there in emergency conditions. This type combination is what EVA aims to construct by its operators as described in the main article. The dataset is made publicly available in our repository, `http://github.com/eva-iclr-2021/EVA`.

Table 8: ASRS. *Person* entity

Person			
Function	Qualification	Experience	Human Factors
Flight crew			
Captain	Student	Total	Communication breakdown
Check Pilot	Sport	Last 90 days	Confusion
First Officer	Private		Distraction
Flight Engineer	Commercial		Fatigue
Instructor	Air Transport Pilot		Human-Machine Interaction
Pilot Flying	Flight Instructor		Physiological
Pilot not Flying	Multiengine		Situational Awareness
Relief Pilot	Instrument		Time Pressure
Single Pilot	Flight Engineer		Training/Qualification
Trainee	Rotorcraft		Workload
Other	Lighter-Than-Air		Other
	Sea		
	Glider		*Location in aircraft*
Air Traffic Control			Flight deck
Approach	Fully certified	Radar	Cabin Jumpseat
Coordinator	Developmental	Non-radar	Crew Rest Area
Departure		Military	Dooe Area
Enroute		Supervisory	Galley
Flight data			General Searing Area
Flight service			Lavatory
Ground			Other
Handoff			
Instructor			
Trainee			
Local			
Oceanic			
Supervisor			
Traffic Management			
Other			
Maintenance			
Inspector	Airframe	Avionics	
Instructor	Powerplant	Inspector	
Lead Technician	Appentice	Lead Technician	
Parts/Stores Personnel	Avionics	Repairman	
Quality Assurance	Inspection Authority	Technician	
Technician	Nondestructive Testing		
Trainee	Repairman		
Other			

Table 9: ASRS. *Events* entity

Events		
Anomalies	Assessment Primary or Contributory factor	Results
Aircraft Equipment		*General*
Critical	Aircraft	Declared Emergency
Less severe	Airport	Evacuated
	Airspace structure	Flight Cancelled/Delayed
Airspace Violation	ATC Equip	Maintenance Action
All types	/Nav Facility/Buildings	Physical Injury/Incapacitation
ATC Issues	Chart or Publication	Police/Security Involved
All types	Company Policy	Release Refused/Aircraft not Accepted
Flight Deck/Cabin/Aircraft	Equipment/Tooling	Work Refused
Illness	Env. non-weather related	None
Passenger Electronic Device	Human Factors	*Flight crew*
Passenger Misconduct	Incorrect/Not Instal.	Reoriented
Smoke/Fire/Fumes/Odor	/Unav. Part	Diverted
Other	Logbook Entry	FLC Overrode Automation
Conflict	Manuals	FLC Complied
NMAC	MEL	Executed Go Around/Missed Approach
Airbone conflict	Procedure	Exited Penetrated Airspace
Ground Conflict, critical	Staffing	Inflight Shutdown
Ground Conflict, less severe	Weather	Landed as Precaution
Deviation - Altitude		Overcame Equipment Problem
Crossing Restriction Not Met		Regained Aircraft Control
Excursion from Assigned Altitude		Rejected Takeoff
Overshoot		Requested ATC Assistance/Clarification
Undershoot		Returned to Clearance
Deviation - Speed or Track/Healing		Returned to Departure Airport
All types		Returned to Gate
Deviation - Procedural		Took Evasive Action
Clearance		*Air Traffic Control*
FAR		Provided Assistance
Hazardous Material Violation		Issued Advisory/Alert
Landing without Clearance		Issued New Clearance
Maintenance		Separated Traffic
MEL		*Aircraft*
Published Material/Policy 5205 - Security		Aircraft Damaged
Weight and Balance		Automation Overrode Flight Crew
Other/Unknown		Equipment Problem Dissipated
Ground Excursion/Incursion		
Ramp		
Runaway		
Taxiway		
Ground Event/Encounter		
Aircraft		
FOD		
Gear Up Landing		
Ground Strike Aircraf		
Loss of Aircraft Control		
Object		
Person/Animal/Bird		
Vehicle		
Other		
Inflight Event/Encounter		
CFTT/CFIT		
Fuel Issue		
Loss of Aircraft Control 5215 - Object		
Bird/Animal		
Unstabilized Approach		
VFR in IMC		
Wake Vortex Encounter		
Weather/Turbulence		