

# Guiding GBFS through Learned Pairwise Rankings

Mingyu Hao<sup>1</sup>, Felipe Trevizan<sup>1</sup>, Sylvie Thiébaux<sup>1,2</sup>, Patrick Ferber<sup>3,4</sup>, Jörg Hoffmann<sup>4</sup>

<sup>1</sup>Australian National University, Australia

<sup>2</sup>LAAS-CNRS, Université de Toulouse, France

<sup>3</sup>University of Basel, Switzerland

<sup>4</sup>Saarland University, Germany

{mingyu.hao,felipe.trevizan,sylvie.thiebaux}@anu.edu.au, patrick.ferber@unibas.ch,  
hoffmann@cs.uni-saarland.de

## Abstract

We propose a new approach based on ranking to learn to guide Greedy Best-First Search (GBFS). As previous ranking approaches, ours is based on the observation that directly learning a heuristic function is overly restrictive, and that GBFS is capable of efficiently finding good plans for a much more flexible class of total quasi-orders over states. In order to learn an optimal ranking function, we introduce a new ranking framework capable of leveraging any neural network regression model and efficiently handling the training data through batching. Compared with previous ranking approaches for planning, ours does not require complex loss functions and allows training on states outside the optimal plan with minimal overhead. Our experiments on the domains of the latest planning competition learning track show that our approach substantially improves the coverage of the underlying neural network models without degrading plan quality.

## 1 Introduction

Greedy Best-First Search (GBFS) is one of the most widely used algorithms in satisfying planning. GBFS uses a heuristic function to estimate the cost to reach the goal from a given node. The nodes are put in a priority queue based on their heuristic values, and the node with the lowest heuristic value is expanded next. This strategy will return a plan if one exists, but may result in suboptimal solutions. Many researchers have made significant efforts to design informative heuristics to guide GBFS [Bonet *et al.*, 1997; Hoffmann and Nebel, 2001; Richter and Westphal, 2010], however the efficiency of the search depends on properties of the heuristic function which are not yet completely understood [Wilt and Ruml, 2016; Cohen and Beck, 2018].

Therefore, in recent years, the possibility of exploring the space of heuristic functions using machine learning has raised increasing attention [Arfaee *et al.*, 2011; Shen *et al.*, 2020; Ferber *et al.*, 2020; Karia and Srivastava, 2021; Ståhlberg *et al.*, 2022; Heller *et al.*, 2022; Chen *et al.*, 2024]. These approaches try to learn or improve a heuristic function for a planning domain or, occasionally, a single large problem

instance. They typically formulate the learning problem as one of regression, and train models using pre-computed state-value pairs extracted from (often optimal) plans generated by off-the-shelf planners for small problem instances.

However, labelling states with informative values and attempting to learn accurate values for the heuristic is expensive and unnecessary for the purpose of learning to guide GBFS. This is because the choice of node expansion by GBFS is not determined by the actual numerical values of the heuristic but by the relative ordering over nodes they induce [Röger and Helmert, 2010]. Therefore, regression models, which attempt to learn precise heuristic values, artificially restrict the target hypothesis space and the training dataset one can learn from. They also introduce issues such as the difficulty of distinguishing between states that have close heuristic values. In fact, for GBFS to find the optimal plan, it suffices for the heuristic to order the states on the optimal path as preferable to those off it. This is called the perfect ranking heuristic [Chrestien *et al.*, 2023] as it also minimises search effort.

These considerations motivate an alternative framing of the problem as one of learning to rank pairs of states, building on the active research on learning to rank in information retrieval and relevance analysis [Burges, 2010; Damke and Hüllermeier, 2021; Li, 2022]. Garrett *et al.* (2016) were the first to approach this problem for planning by using a statistical machine learning model, namely RankSVM [Joachims, 2002], hand-crafted domain-independent features, and a loss function measuring the normalised difference between the number of correctly and incorrectly classified pairs.

In contrast, this paper explores pairwise ranking with modern deep-learning methods. We propose a novel framework combining a Direct Ranker [Köppel *et al.*, 2020b] with any existing neural network regression model for learning heuristics, which are trained end-to-end in order to learn a total quasi-order, i.e. a total, transitive, and reflexive relation over pairs of states. This framework does not require complex loss functions, and exploits the transitivity of the ordering relation to learn a ranking just slightly stronger than the perfect ranking, from an order of magnitude less data than alternative approaches. Moreover, it supports efficient batched evaluation to speed up training.

We evaluate our approach on the learning track of the 2023 international planning competition, using the recent STRIPS-HGN [Shen *et al.*, 2020] and GOOSE [Chen *et al.*, 2024]

domain-independent graph neural network architectures, as the underlying regression model. We show that our ranking approach considerably improves the coverage of both these models and expands their ability to generalise to larger problems than those trained on, whilst preserving their plan quality and time to convergence. We also compare our approach with the recently and independently developed work from Chrestien *et al.* (2023). We show that our framework requires less training data and generalises better to unseen problems.

## 2 Background

We start by introducing the necessary background on planning, greedy best-first search, and learning to rank. We use  $\llbracket n \rrbracket$  to denote  $\{1, 2, \dots, n\}$ .

### 2.1 Planning

We consider classical planning problems described by a tuple  $\mathbb{P} = \langle S, s_0, G, A, C \rangle$  where  $S$  is a finite and discrete set of states,  $s_0$  is the initial state,  $G \subseteq S \setminus \{s_0\}$  is the set of goal states,  $A$  is the set of actions of which  $A_s$  is the subset applicable in state  $s$ , and an action  $a \in A_s$  is a deterministic transition function, which takes state  $s$  as input and outputs the successor state  $s'$ . In the following, we write  $N_s$  for the set of successors of a state  $s$ : it consists of all states  $s'$  such that there exists an applicable action  $a \in A_s$  with  $s' := a(s)$ . The cost function  $C(a)$  returns the non-negative cost of applying action  $a$ . A unit-cost problem is one for which all actions have cost 1. A plan for  $\mathbb{P}$  is a sequence of actions  $a_1, \dots, a_n$  that induces a sequence of states (a path)  $s_0, s_1, \dots, s_n$  where  $a_i \in A_{s_{i-1}}$ ,  $s_i := a_i(s_{i-1})$  for  $i \in \llbracket n \rrbracket$ , and  $s_n \in G$ . A plan is optimal if its cost  $\sum_{i \in \llbracket n \rrbracket} C(a_i)$  is minimal. In the unit-cost case, optimal plans are plans of minimal length.

Greedy Best-First Search (GBFS) is the most commonly used algorithm for satisficing planning. In order to guide the search towards the goal, GBFS relies on a heuristic function  $h : S \rightarrow \mathbb{R}$  returning an estimate  $h(s)$  of the cost to reach the goal from a given state  $s$ , and always expands the state with the lowest heuristic value on the search frontier. As shown in Algorithm 1, GBFS maintains two lists of nodes<sup>1</sup>: the open list ( $\mathcal{O}$ ) – a priority queue ordered by increasing  $h$  value, containing the nodes that have been generated but not yet expanded, and a closed list ( $\mathcal{C}$ ) containing previously expanded nodes which is used to avoid revisiting the same state twice and creating loops. At each iteration, the node at the front of the open list – the node with the lowest  $h$  value – is transferred to the closed list and expanded by generating its successors. The search stops if a successor is a goal state. Otherwise, any unvisited successor is added to the open list.

GBFS is guaranteed to find a plan if one exists, but the plan may not be optimal. In fact, even when guided with the  $h^*$  heuristic which returns the optimal cost to reach the goal, GBFS is only guaranteed to return the optimal plan for unit cost problems. In the following, we assume that problems have unit cost.

<sup>1</sup>In the interest of readability, Algorithm 1 identifies nodes with states, whereas nodes additionally record the  $h$  value of the state, the parent node, and the action applied at the parent.

---

### Algorithm 1 GBFS Algorithm

---

**Input:** problem  $\mathbb{P}$ , heuristic function  $h$

**Output:** Plan  $\pi$  or *unsolvable*

```

1:  $\mathcal{O} := \{s_0\}$ 
2:  $\mathcal{C} := \emptyset$ 
3: while  $\mathcal{O} \neq \emptyset$  do
4:    $s := \operatorname{argmin}_{s' \in \mathcal{O}} h(s')$ 
5:    $\mathcal{O} := \mathcal{O} \setminus \{s\}$ 
6:    $\mathcal{C} := \mathcal{C} \cup \{s\}$ 
7:   for all  $s' \in N_s$  do
8:     if  $s' \in G$  then
9:       return Extract-Solution( $s'$ )
10:    end if
11:    if  $s' \notin \mathcal{C} \cup \mathcal{O}$  then
12:       $\mathcal{O} := \mathcal{O} \cup \{s'\}$ 
13:    end if
14:  end for
15: end while
16: return unsolvable

```

---

### 2.2 Learning to Rank

Learning-to-rank is a well-studied field in machine learning [Liu and others, 2009; Li, 2022]. Given a set of objects, in our case a set of states,  $B = \{s_1, \dots, s_k\}$ , it aims to learn an order  $\preceq$  that results in the ranking  $s_{i_1} \preceq s_{i_2} \preceq \dots \preceq s_{i_k}$ , where  $i_1, \dots, i_k$  is a permutation of  $\llbracket k \rrbracket$ . In this work, we further require the order to have the following properties:

**Definition 1** (Total Quasi-Order  $\preceq$ ). *A total quasi-order  $\preceq$  over a set  $X$  is a binary relation with the following properties*

- *Totality:*  $\forall a, b \in X \ a \preceq b \vee b \preceq a$ ;
- *Transitivity:*  $\forall a, b, c \in X \ a \preceq b \wedge b \preceq c \Rightarrow a \preceq c$ .

*Moreover, totality implies*

- *Reflexivity:*  $\forall a \in X \ a \preceq a$ .

We define the strict order  $\prec$  in the obvious way:  $a \prec b$  iff  $a \preceq b$  and  $b \not\preceq a$ . Note that transitivity also holds for the strict order  $\prec$ :  $\forall a, b, c \in X \ a \prec b \wedge b \prec c \Rightarrow a \prec c$ .

There are several different approaches to solving the learning-to-rank problem, such as SVM [Shashua and Levin, 2002; Herbrich *et al.*, 1999], Boosting [Freund *et al.*, 2003; Wu *et al.*, 2010] and Neural Networks [Burges *et al.*, 2005; Köppel *et al.*, 2020b]. Based on how the order is represented, learning-to-rank approaches are divided into three groups: *pointwise*, *pairwise* and *listwise* [Li, 2022]. Listwise approaches learn to sort a list of objects to match a desired total order. For planning, this approach is infeasible since we do not know a priori the list of states we need to sort.

The pointwise approach represents the total quasi-order by a ranking function  $r : B \rightarrow \mathbb{R}$  which takes in one object and returns a numerical score [Li, 2022]. This numerical score is then used to sort  $B$ . Note that GBFS uses the heuristic  $h$  as a pointwise ranking function: the set of open states  $\mathcal{O}$  is sorted according to their  $h$  values (Line 4 of Algorithm 1). Similarly, learning a heuristic function can be seen as learning a pointwise ranking function with extra constraints imposed by the definition of a heuristic function, e.g.,  $h(s) \geq 0$  for all  $s$ ,  $h(g) = 0$  for all  $g \in G$ , or even admissibility or consistency.

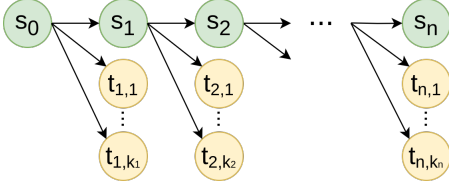


Figure 1: A example graph representation of part of the search tree. The optimal path is  $[s_0, s_1, \dots, s_n]$ . States  $t_{i,j}$  are siblings of  $s_i$  and have higher ranks, i.e.,  $s_i \preceq t_{i,j}$ .

Lastly, a pairwise ranking function takes a pair of elements as input and it outputs the ordering between the two objects. Formally, a pairwise ranking function  $r(s_i, s_j) \in \{-1, 0, 1\}$  represents whether  $s_i \prec s_j$ ,  $s_i = s_j$  or  $s_j \prec s_i$ , respectively [Köppel *et al.*, 2020b]. While learning a pointwise ranking function is a regression problem, learning a pairwise ranking is a classification problem, i.e., given two states  $s_i$  and  $s_j$ , we need to assign it either  $-1$ ,  $0$ , or  $1$ . As we show in the next section, this difference has a large impact on learning a ranking for planning in comparison to learning a heuristic.

### 3 Rank-based GBFS

In the context of GBFS, the heuristic function  $h$  is used solely for sorting states in the priority queue as shown in Line 4 of Algorithm 1. This means  $h$  is being used only as a total quasi-order over the states  $S$ . This fact has been exploited by others, for instance, by using functions other than the cost-to-goal to order states in GBFS [Ferber *et al.*, 2022] and to analyse what makes a good heuristic for GBFS [Wilt and Ruml, 2016]. For this reason, we extend GBFS to use a ranking instead of a heuristic. Formally, given a total quasi-order  $\preceq$ , rank-based GBFS is the algorithm obtained by replacing Line 4 in Algorithm 1 with: choose  $s$  such that  $s \preceq t$  for all  $t \in \mathcal{O}$ . We relate the two versions of GBFS with the lemma below.

**Definition 2** ( $\preceq_h$ ). *Given a heuristic  $h$ , the total quasi-order  $\preceq_h$  associated with it is so that  $s \preceq_h s' \Leftrightarrow h(s) \leq h(s')$ .*

**Lemma 1** (Rank-based GBFS equivalence). *Given a problem  $\mathbb{P}$ , a heuristic  $h$ , GBFS using  $h$  is equivalent to rank-based GBFS using  $\preceq_h$  for the same tie-breaking method.*

It is easy to see that Lemma 1 holds based on the definition of  $\preceq_h$ . Lemma 1 also explains why any monotonic non-decreasing function  $f$  applied to the heuristic  $h$ , i.e.,  $h'(s) = f(h(s))$ , does not change the solution found by GBFS:  $f$  preserves order, and therefore the total quasi-order  $\preceq_h$  and  $\preceq_{f(h)}$  are the same. We close this section by looking at what ranking function we should aim to learn.

An intuitive ranking to learn is the ranking induced by  $h^*$ , i.e.,  $\preceq_{h^*}$  since GBFS is guaranteed to find the optimal solution using  $h^*$  under unit-cost actions. However, for usage with GBFS,  $\preceq_{h^*}$  orders more states than are necessary to find the optimal solution. To see this, consider the example in Figure 1. The total quasi-order  $\preceq_{h^*}$  imposes an ordering between any pair of states in the example, including between siblings of states on the optimal path, i.e.,  $t_{i,j} \preceq_{h^*} t_{i',j'} \Leftrightarrow h^*(t_{i,j}) \leq h^*(t_{i',j'})$ . However, the ordering among siblings is not needed, i.e., it is sufficient to know

that  $s_i \prec t_{j,m}$  for  $j \in \llbracket i \rrbracket$  and  $m \in \llbracket k_i \rrbracket$ . We exploit this fact to define an optimal ranking as:

**Definition 3** (Optimal Ranking  $\preceq^*$ ). *Given a problem  $\mathbb{P}$  and an optimal plan  $\pi^* = [s_0, s_1, \dots, s_n]$  for  $\mathbb{P}$ , the optimal ranking  $\preceq^*$  is a total quasi-order such that  $s_i \prec^* s_{i-1}$  for all  $i \in \llbracket n \rrbracket$  and  $s_i \prec^* t \forall t \in N_{s_{i-1}}$  and  $\forall i \in \llbracket n \rrbracket$ .*

Notice that Definition 3 only relates a state  $s_i$  in the optimal plan with its parent  $s_{i-1}$  and its siblings  $t$ . Since it does not enforce an ordering between the siblings  $t$  of  $s_i$ , the optimal ranking is still a total quasi-order. Moreover, due to the transitivity property of total quasi-orders, we have that  $s_i \prec^* t'$  for all  $t' \in \bigcup_{j=0}^{i-1} N_{s_j}$ , i.e.,  $s_i$  is ranked strictly better than any of its ancestors and their siblings. Next, we prove that GBFS finds the optimal solution when using an optimal ranking.

**Theorem 1.** *For a solvable problem  $\mathbb{P}$  with unit costs, given an optimal ranking  $\preceq^*$ , the plan returned by the rank-based GBFS is optimal.*

*Proof.* Consider a problem  $\mathbb{P}$  with unit costs, with an optimal solution  $\pi^* = [s_0, s_1, \dots, s_n]$  and optimal ranking  $\preceq^*$ . We prove by induction that  $s_{i-1}$  is expanded at the  $i$ th iteration of the algorithm and that after expansion, the open list is  $\mathcal{O}_i = \bigcup_{j=0}^{i-1} N_{s_j} \setminus \bigcup_{j=0}^{i-1} \{s_j\}$  for all  $i \in \llbracket n \rrbracket$ . For readability, given a state  $s$  and set of states  $X$ , we abbreviate  $\forall t \in X, s \prec^* t$  with  $s \prec^* X$ . For the **base case**: initially, the open list of GBFS only contains  $s_0$  (Line 1 of Algorithm 1), which is always expanded at the first iteration of the while loop in Line 3, leading to  $\mathcal{O}_1 = N_{s_0} \setminus \{s_0\}$ . For the **induction step**: assume the property holds, we show that  $s_i$  is expanded next and  $\mathcal{O}_{i+1} = \bigcup_{j=0}^i N_{s_j} \setminus \bigcup_{j=0}^i s_j$ . According to the definition of the optimal ranking we have:  $s_i \prec^* N_{s_{i-1}}$  (a) and  $s_i \prec^* s_{i-1}$  (b). Because  $s_{i-1}$  was expanded in the last iteration, we have  $s_{i-1} \prec^* \bigcup_{j=0}^{i-2} N_{s_j} \setminus \bigcup_{j=0}^{i-2} \{s_j\}$  (c). Since  $\prec^*$  is transitive, (b) and (c) lead to  $s_i \prec^* \bigcup_{j=0}^{i-2} N_{s_j} \setminus \bigcup_{j=0}^{i-2} \{s_j\}$  (d). Finally, the union of (a) and (d) gives  $s_i \prec^* \bigcup_{j=0}^{i-1} N_{s_j} \setminus \bigcup_{j=0}^{i-1} \{s_j\}$ . In other words,  $s_i$  is ranked at the top of the open list  $\mathcal{O}_i$  and will be expanded in iteration  $i+1$ . After expansion the queue will be  $\mathcal{O}_{i+1} = \mathcal{O}_i \cup (N_{s_i} \setminus \bigcup_{j=0}^i \{s_j\}) = \bigcup_{j=0}^i N_{s_j} \setminus \bigcup_{j=0}^i \{s_j\}$ .  $\square$

A key difference between optimal rankings and  $\preceq_{h^*}$  is that there are several total quasi-orders that are optimal rankings while  $\preceq_{h^*}$  is unique. This makes learning an optimal ranking potentially easier than learning  $\preceq_{h^*}$  or even  $h^*$  itself, since  $h^*$  is also unique. We exploit this in the next section where we introduce our framework to learn an optimal ranking.

## 4 Learning an Optimal Ranking

In this section, we introduce a new framework capable of leveraging any neural network regression model to learn an optimal ranking function. We also show how to efficiently integrate the ranking computations into GBFS and how to extract more data from the optimal plans for training while still efficiently handling this extra data through batching.

### 4.1 DirectRanker

Our approach for learning an optimal ranking through pairwise classification uses *DirectRanker* [Köppel *et al.*, 2020b].

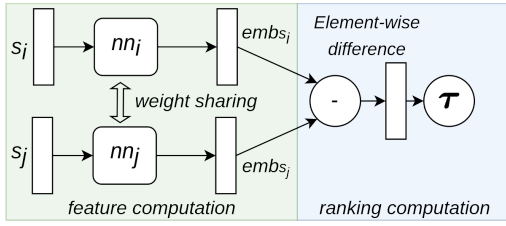


Figure 2: The structure of the DirectRanker [Köppel *et al.*, 2020b].

DirectRanker is a general model for pairwise ranking that is designed to learn total quasi-orders. Its structure is shown in Figure 2. The model consists of two parts: feature computation and ranking computation. The first part receives two objects (in our case, states)  $s_i$  and  $s_j$  to be ranked and generates the embedding vectors  $emb_{s_i}$  and  $emb_{s_j}$ , both in  $\mathbb{R}^n$ , through the neural networks  $nn_i$  and  $nn_j$ . The structure of the neural networks  $nn_i$  and  $nn_j$  is not restricted; the only requirement is that both  $nn_i$  and  $nn_j$  have the same structure and share weights, i.e., they must be identical.

The second part of the model, ranking computation, computes the element-wise difference of the two embedding vectors and passes it through a single neuron  $\tau$  with zero bias. Formally, this second part computes  $p = \sigma(\mathbf{w}(emb_{s_i} - emb_{s_j}))$  where  $\mathbf{w} \in \mathbb{R}^{1 \times n}$  are the weights and  $\sigma$  is the activation function of  $\tau$ . The activation function  $\sigma$  must be odd and sign-preserving, i.e.,  $\sigma(-x) = -\sigma(x)$  and  $sign(\sigma(x)) = sign(x)$ . The input states  $s_i$  and  $s_j$  are classified as  $s_i \prec^* s_j$ ,  $s_i = s_j$ , or  $s_j \prec^* s_i$  if  $p$  is less than, equal to, or greater than 0, respectively. Under these assumptions, DirectRanker is guaranteed to return a total quasi-order [Köppel *et al.*, 2020b, Theorem 1]. Besides the activation function  $\sigma$ , the loss function used by DirectRanker is also a parameter of the framework and we use the same parametrisation as in Köppel *et al.* (2020b) experiments: activation function  $\sigma(x) = (1 + e^{-x})^{-1} - 0.5$  and mean squared error (MSE) as the loss function [Köppel *et al.*, 2020a].

## 4.2 Efficient Integration with GBFS

Once we learn the shared weights for  $nn_1$  and  $nn_2$  as well as  $\mathbf{w}$  for the ranking computation, we need to integrate it into GBFS. Let  $D(s, t) \in \{-1, 0, 1\}$  represent the result of the DirectRanker for  $s$  and  $t$ . In order to insert a new state  $s$  into the priority queue  $\mathcal{O}$ , we need to compare  $s$  with states already in  $\mathcal{O}$  since we have learned a pairwise ranking. However, the amortised computational complexity of insertion is  $O(k \log |\mathcal{O}|)$ , where  $k$  is the time needed to evaluate  $D(s, t)$  for a single comparison.

Since the ranking computation part of DirectRanker is a single-layer, unbiased neural network, we can transform our learned pairwise ranking into a point-wise representation to improve the insertion efficiency in the priority queue. This transformation is formalised in Lemma 2 and lets us represent the learned  $D(s, t)$  as the point-wise ranking  $r(s) \in \mathbb{R}$ . As a result, the priority queue in GBFS can now be sorted by  $r(s)$  resulting in an amortised insertion time of  $O(k' + \log |\mathcal{O}|)$  where  $k'$  is the time required to evaluate  $r(s)$ . With this transformation, only a single neural network evaluation is needed

instead of  $\log |\mathcal{O}|$  when using the pairwise ranking function  $D(s, t)$ . Although  $r(s)$  is being used as a heuristic in GBFS, it is important to note that it is not a heuristic in the traditional sense, for instance,  $r(s)$  can be negative. Moreover, since we are learning an optimal ranking, it is possible that  $r(t) \leq r(t')$  even though  $h^*(t) > h^*(t')$  for states  $t$  and  $t'$  outside the optimal plan.

**Lemma 2.** *Given a DirectRanker model with weights  $\mathbf{w}$  representing the total quasi-order  $\preceq$ , we can extract a point-wise ranking function  $r : \mathcal{S} \rightarrow \mathbb{R}$  representing  $\preceq$ .*

*Proof.* Given two states  $s_i, s_j \in \mathcal{S}$ , without loss of generality, we assume  $s_i \preceq s_j$ . Then from the definition of the DirectRanker, we have  $p = \sigma(\mathbf{w} \cdot (nn_1(s_i) - nn_2(s_j))) \leq 0$ . Since  $nn_1$  and  $nn_2$  are identical and  $\sigma$  is sign-preserving, we have that  $p \leq 0 \Leftrightarrow \mathbf{w} \cdot (nn(s_i) - nn(s_j)) \leq 0$ . Because the right-hand side is a linear transformation, it is equivalent to  $\mathbf{w} \cdot nn(s_i) \leq \mathbf{w} \cdot nn(s_j)$ . Defining  $r(s)$  as  $\mathbf{w} \cdot nn(s)$ , we have that  $p \leq 0 \Leftrightarrow r(s_i) \leq r(s_j)$  and therefore  $s_i \preceq s_j \Leftrightarrow r(s_i) \leq r(s_j)$ .  $\square$

## 4.3 Efficient Training

Another key difference of our framework lies in the training methodology. Unlike when learning a heuristic, where a specific value is required for each state, we train our DirectRanker  $D(s_i, s_j)$  to learn the optimal ranking through classification and then extract the point-wise ranking function  $r(s)$ . Thus, given a pair of states  $s_i$  and  $s_j$ , it is sufficient to determine their relative optimal ranking, i.e., whether  $s_i \preceq^* s_j$  or  $s_j \preceq^* s_i$ , without assigning specific values to each state. This allows us to use siblings of states in optimal plans in the training process without the need to compute any information about them. Thus, for a problem with a constant branching factor  $b$ , an optimal plan  $\pi^*$  of size  $n$  encodes  $O(bn)$  optimal ranking pairs for training. This contrasts with methods that directly learn a heuristic function (e.g., STRIPS-HGN [Shen *et al.*, 2020] and GOOSE [Chen *et al.*, 2024]), which can only extract  $n$  pairs in the form  $(s, h^*(s))$  from  $\pi^*$ . Note that it is computationally expensive to compute  $h^*(s')$  for any  $s' \notin \pi^*$  since it requires optimally solving the problem from  $s'$ .

Another advantage of our framework is that we can efficiently handle this extra training data by batching the pairs  $s_i \prec t$  for all siblings  $t$  of  $s_i$  and also  $t = s_{i-1}$  for each  $s_i \in \pi^*$  [Damke and Hüllermeier, 2021]. The performance gains come from the fact that the embedding  $emb_{s_i}$  is computed only once as opposed to  $|N_{s_{i-1}}|$  times, i.e., once for each pair involving  $s_i$ . Formally, let  $B_{s_i} = \{s_{i-1}\} \cup (N_{s_{i-1}} \setminus \{s_i\})$ , i.e., the set consisting of the siblings of  $s_i$  and its parent. We compute the embeddings for all states in  $B_{s_i}$  resulting in the matrix  $\mathbf{E}_r \in \mathbb{R}^{m \times n}$  where  $m = |B_{s_i}|$  and the  $j$ -th row of  $\mathbf{E}_r$  is the vector embedding of the  $j$ -th element in  $B_{s_i}$ . Next, we compute the embedding  $emb_{s_i}$  and stack  $m$  copies of it to create the matrix  $\mathbf{E}_l \in \mathbb{R}^{m \times n}$ . Then we compute  $\mathbf{p} = \sigma((\mathbf{E}_l - \mathbf{E}_r)\mathbf{w}) \in \mathbb{R}^m$  where the  $j$ -th row of  $\mathbf{p}$  is the output of  $\tau$  comparing  $s_i$  and the  $j$ -th element in  $B_{s_i}$ . Thus, we compute  $\mathbf{p}$  by generating only  $|B_{s_i}| + 1$  embeddings as opposed to  $2|B_{s_i}|$  embeddings. This difference is important since computing each embedding in our experiments requires evaluating an expensive neural network architecture.

## 5 Related Work

The closest work to ours is [Chrestien *et al.*, 2023] which learns a heuristic function using a pointwise ranking as the target, as opposed to  $h^*$ . As with other approaches that learn a heuristic function, e.g., STRIPS-HGN [Shen *et al.*, 2020] and GOOSE [Chen *et al.*, 2024], their learned heuristics can be used with  $A^*$ . However, all these approaches cannot guarantee that the solution found is optimal since the learned heuristic may not be admissible. For this reason, our work focuses on GBFS which allows us to remove the restriction that the learned ranking must also be a heuristic.

What makes the approach by Chrestien *et al.* (2023) different from all previous works on learning heuristics is that their target is a *perfect ranking*, a concept they introduced. A perfect ranking orders a pair of states  $s$  and  $t$  as  $s \prec t$  for all  $s$  on the optimal path and  $t \in \mathcal{O}$  when  $s$  is popped from the GBFS queue. It is equivalent to an optimal ranking without transitivity nor the requirement that  $s_i \prec s_{i-1}$  for two consecutive states  $s_{i-1}$  and  $s_i$  in  $\pi^*$ . To see the benefits of using an optimal ranking as the target instead of a perfect ranking, consider a problem with a constant branching factor  $b$  and an optimal plan  $\pi^* = [s_0, s_1, \dots, s_n]$ . The number of ordered pairs  $s \prec t$  encoded in the optimal plan is  $\sum_{i=1}^n ib = b(n^2 + n)/2$ . To learn a perfect ranking, Chrestien *et al.* (2023) use all pairs encoded in the optimal plan with the exception of the pairs  $s_i \prec s_j$  for  $s_i, s_j \in \pi^*$  and  $i < j$ , resulting in  $(b-1)(n^2 + n)/2$  total pairs. Alternatively, due to transitivity and the requirement that  $s_i \prec s_{i-1}$  for  $s_i \in \pi^*$  imposed by our optimal ranking, we can encode all the  $b(n^2 + n)/2$  pairs using only  $bn$  pairs. Therefore, our training data scales linearly instead of quadratically with  $|\pi|$  while representing all ordered pairs considered by [Chrestien *et al.*, 2023]. As our experiments will show, this has a large empirical effect on training efficiency and generalisation.

Our approach also differs in how the target ranking is learned. We learn a pairwise ranking function using classification through the DirectRanker architecture that is later converted to a pointwise ranking to be efficiently used with GBFS. The heuristic learned by [Chrestien *et al.*, 2023] follows an approach in which a pointwise ranking function is learned using the 0-1 loss for an implicit pairwise ranking function. Their loss function is later relaxed to a logistic loss in order to use gradient optimisation.

Another work based on ranking is [Garrett *et al.*, 2016]. They learn a pairwise ranking by training a Rank Support Vector Machine model [Joachims, 2002] using a convex relaxation of the *Kendall rank correlation coefficient*. This coefficient measures the degree of similarity between two rankings by computing the difference between the number of concordant and discordant pairs as a proportion of the total number of possible pairs. One main difference between our approaches is that we allow the use of any neural network for computing features while they use hand-crafted features extracted from the graph implicitly built by the FF heuristic [Hoffmann and Nebel, 2001]. Another key difference is that they use as data only the pairs within a suboptimal plan  $\pi$ , that is, they do not use the states outside of  $\pi$ .

| domain      | Training        |    | Testing           |    |
|-------------|-----------------|----|-------------------|----|
| blocksworld | $n \in [2, 21]$ | 56 | $n \in [5, 488]$  | 90 |
| childsnaek  | $c \in [1, 5]$  | 37 | $c \in [4, 292]$  | 90 |
| ferry       | $c \in [1, 13]$ | 66 | $c \in [2, 974]$  | 90 |
| floortile   | $t \in [2, 30]$ | 64 | $t \in [12, 980]$ | 90 |
| miconic     | $p \in [1, 10]$ | 99 | $p \in [1, 485]$  | 90 |
| rovers      | $r \in [1, 4]$  | 67 | $r \in [1, 30]$   | 90 |
| satellite   | $s \in [1, 10]$ | 90 | $n \in [3, 99]$   | 90 |
| sokoban     | $n \in [7, 13]$ | 99 | $b \in [8, 99]$   | 90 |
| spanner     | $s \in [1, 10]$ | 89 | $n \in [1, 487]$  | 90 |
| transport   | $v \in [1, 5]$  | 47 | $n \in [3, 50]$   | 90 |

Table 1: Number of problems per domain and problem size in terms of the most significant object type:  $n$  blocks in blocksworld,  $c$  children in childsnaek,  $c$  cars in ferry, etc. The full set of parameters governing problem size can be found in [Segovia and Seipp, 2023].

## 6 Experiments

In this section, we empirically compare our ranking framework to Chrestien *et al.* 2023’s framework using GBFS. For both frameworks, we consider two state-of-the-art neural network regression models, namely STRIPS-HGN [Shen *et al.*, 2020] and GOOSE-LLG [Chen *et al.*, 2024], for internal feature computation. To verify that learning a ranking is beneficial, we consider STRIPS-HGN and GOOSE by themselves, i.e., trained using  $h^*$  as the learning target. As a baseline, we also consider the  $h^{\text{ff}}$  heuristic [Hoffmann and Nebel, 2001].

### 6.1 Dataset

We use the 10 domains from the 2023 International Planning Competition Learning Track (IPC23-LT) [Segovia and Seipp, 2023]. Each domain provides 100 training problems and 90 testing problems. We first solve the training problems using the Scorpion planner [Seipp *et al.*, 2020] with a 30-minute time limit. Only problems solved optimally by Scorpion are used for training, but this still results in a dataset of sufficient size for our experiments. The training problems are randomly split between the training set (90%) and the validation set (10%). The problems in the validation set are used only to monitor convergence and control learning rates.

The testing set is divided into “easy”, “medium” and “hard” subsets per domain, based on problem size, with 30 problems per subset. In contrast the training set only contains “easy” problems. Details of the dataset are shown in Table 1.

### 6.2 Training Settings

The two graph neural network regression models we use are GOOSE Lifted Learning Graph (LLG) [Chen *et al.*, 2024] and STRIPS-HGN [Shen *et al.*, 2020]. The former encodes the full planning problem and domain in a lifted form, whereas the latter encodes the problem delete relaxation in a grounded form. Both models can learn domain-independent heuristics and our framework inherits this property. However, our experiments are restricted to domain-specific rankings.

For both regression models, we used the code provided by their authors with all ReLU activations replaced by LeakyReLUs, except in their last layer. The ReLU function is suitable for learning heuristics because its values are non-negative; however, this restriction is not needed for learning a ranking

| domain      | $h^{\text{ff}}$ | GOOSE |      |      |           | PerfRank(GOOSE) |      |      |           | OptRank(GOOSE) |      |      |            | HGN  |      |      |           | PerfRank(HGN) |      |      |          | OptRank(HGN) |      |      |            |
|-------------|-----------------|-------|------|------|-----------|-----------------|------|------|-----------|----------------|------|------|------------|------|------|------|-----------|---------------|------|------|----------|--------------|------|------|------------|
|             | sum             | easy  | med. | hard | sum       | easy            | med. | hard | sum       | easy           | med. | hard | sum        | easy | med. | hard | sum       | easy          | med. | hard | sum      | easy         | med. | hard | sum        |
| blocksworld | <b>28</b>       | 30    | 0    | 0    | <u>30</u> | 30              | 5    | 0    | <u>35</u> | 30             | 20   | 1    | <b>51</b>  | 22   | 0    | 0    | <u>22</u> | -             | -    | -    | -        | 23           | 0    | 0    | <u>23</u>  |
| childsnaek  | 26              | 19    | 0    | 0    | 19        | 21              | 4    | 0    | 25        | 26             | 4    | 0    | <b>30</b>  | 6    | 0    | 0    | 6         | 13            | 0    | 0    | 13       | 25           | 0    | 0    | <u>25</u>  |
| ferry       | <b>68</b>       | 30    | 30   | 1    | 61        | 30              | 30   | 2    | <u>62</u> | 30             | 30   | 1    | 61         | 26   | 2    | 0    | 28        | -             | -    | -    | -        | 30           | 8    | 0    | <u>38</u>  |
| floortile   | <b>12</b>       | 1     | 0    | 0    | 1         | 2               | 0    | 0    | <u>2</u>  | 1              | 0    | 0    | 1          | 0    | 0    | 0    | 0         | 1             | 0    | 0    | <u>1</u> | 1            | 0    | 0    | <u>1</u>   |
| miconic     | <b>90</b>       | 30    | 30   | 15   | 75        | 30              | 30   | 16   | <u>76</u> | 30             | 30   | 15   | 75         | 30   | 30   | 0    | <u>60</u> | 24            | 0    | 0    | 24       | 30           | 27   | 1    | 58         |
| rovers      | <b>34</b>       | 29    | 2    | 0    | <u>31</u> | 28              | 1    | 0    | 29        | 30             | 1    | 0    | <u>31</u>  | 25   | 0    | 0    | <u>25</u> | 10            | 0    | 0    | 10       | 28           | 0    | 0    | <u>28</u>  |
| satellite   | <b>65</b>       | 27    | 2    | 0    | 29        | 27              | 0    | 0    | 27        | 29             | 4    | 0    | <u>33</u>  | 13   | 0    | 0    | 13        | 6             | 0    | 0    | 6        | 30           | 1    | 0    | <u>31</u>  |
| sokoban     | <b>36</b>       | 26    | 1    | 0    | 27        | 30              | 2    | 0    | <u>32</u> | 30             | 2    | 0    | <u>32</u>  | 27   | 0    | 0    | 27        | 17            | 0    | 0    | 17       | 30           | 1    | 0    | <u>31</u>  |
| spanner     | 30              | 30    | 4    | 0    | 34        | 30              | 30   | 2    | <b>62</b> | 30             | 30   | 1    | 61         | 30   | 0    | 0    | 30        | 30            | 4    | 0    | 34       | 30           | 14   | 0    | <u>44</u>  |
| transport   | 41              | 30    | 8    | 0    | 38        | 30              | 9    | 0    | 39        | 30             | 16   | 0    | <b>46</b>  | 22   | 0    | 0    | 22        | 6             | 0    | 0    | 6        | 28           | 0    | 0    | <u>28</u>  |
| sum         | <b>430</b>      | 252   | 77   | 16   | 345       | 258             | 111  | 20   | 389       | 266            | 137  | 18   | <u>421</u> | 201  | 32   | 0    | 233       | 107           | 4    | 0    | 111      | 254          | 52   | 1    | <u>307</u> |

Table 2: Rounded average coverage per domain and problem difficulty for the different configurations. Best coverage per domain highlighted in bold. The best coverage for each regression model is underlined.

and we observed that it results in slower convergence. For consistency, the same model parameters are used across our experiments: 4 message-passing layers and hidden dimension  $m = 64$ . For our framework, we use the output from the second-last hidden layer of the underlying regression model as the embedding vector, thus  $emb_s \in \mathbb{R}^{64}$ .

For each domain, we compare GBFS using the following configurations for the heuristic/ranking:

**GOOSE and HGN.** The original GOOSE-LLG and STRIPS-HGN regression models trained using  $h^*$  as target and their original loss function. The training dataset consists of all state-value pairs  $(s, h^*(s))$  for  $s$  in the optimal plan.

**OptRank(GOOSE) and OptRank(HGN).** Our optimal ranking framework combined with the respective regression models. The training dataset is the set of all ordered pairs  $s \prec t$  for  $s$  in the optimal plan and  $t \in N_{s'} \cup \{s'\}$  where  $s'$  precedes  $s$  in the optimal plan.

**PerfRank(GOOSE) and PerfRank(HGN).** Chrestien *et al.* (2023)’s framework to learn a perfect ranking heuristic combined with each model. The loss function associated with the perfect ranking definition is also used. The training dataset consists of all ordered pairs  $s \prec t$  for  $s$  in the optimal plan and  $t \in \mathcal{O}$  when  $s$  is popped from the GBFS queue.

We use the Adam optimiser and an initial learning rate of  $10^{-3}$ . The learning rate is reduced by a factor of 10 if the accuracy on the validation set does not improve for 10 consecutive epochs. Training is stopped when the learning rate reaches  $10^{-6}$  or after 500 epochs.

For each configuration, we solve the problems in the testing set with Fast-Downward using eager-GBFS [Helmert, 2006], a time limit of 30 min per problem, and 8 GB of memory. All experiments are run on an Intel Xeon 2.1GHz CPU and NVIDIA A6000 GPU with 64GB of memory. The experiments were repeated three times and the average results are reported. The source code can be found at [Hao *et al.*, 2024a] and an extended set of results in [Hao *et al.*, 2024b].

### 6.3 Results

**Total Coverage.** Table 2 shows the coverage of each configuration on each domain, i.e., the number of problems it solves for this domain. Overall, using optimal rankings improved the coverage of the underlying regression model: 22%

and 32% increase for GOOSE and STRIPS-HGN, respectively. Our optimal ranking approach also obtained higher total coverage than the perfect ranking one: 8% and 176% increase for GOOSE and STRIPS-HGN, respectively. The large increase in coverage obtained by OptRank(HGN) w.r.t PerfRank(HGN) can be partially attributed to: (i) STRIPS-HGN being a slower model to train; and (ii) PerfRank(HGN) using quadratically more data than OptRank(HGN). The convergence time for all learned configurations is shown in Figure 3 and we observe that, for any framework, using STRIPS-HGN to generate features is slower than using GOOSE.

The impact of using STRIPS-HGN is less pronounced on OptRank(HGN) due to its compact dataset that exploits transitivity and is nearly equivalent to that of PerfRank(HGN). Moreover, despite processing  $O(b)$  more data than GOOSE where  $b$  is the average branching factor of a domain, OptRank(GOOSE) converges in a similar amount of time. Hence, it processes training data much faster.

**Domain-specific coverage.** When considering individual domains, OptRank(GOOSE) obtains the highest coverage in 3 domains, namely Blocksworld, Childsnack and Transport, with 46% (16 problems), 15% (4 problems) and 12% (5 problems) more coverage than the second-best planner, respectively. PerfRank(GOOSE) has the highest coverage only on Spanner, where it solves one more problem (2%) than OptRank(GOOSE). In Floortile, all configurations perform very poorly. This domain is known to be hard because it requires a large receptive field and it has many state space symmetries.

**GOOSE configurations.** OptRank(GOOSE) improves GOOSE’s coverage in 6 domains and obtains the same coverage in the other 4 domains, therefore using optimal rankings is highly beneficial for this regression model. E.g, in Blocksworld and Spanner, GOOSE solves none or hardly any of the medium-difficulty problems, whereas OptRank(GOOSE) solves a large fraction or all of them, demonstrating a greater ability to generalise to larger problem sizes. PerfRank(GOOSE) almost always improves the coverage of GOOSE and its coverage is marginally smaller (2 problems, representing about 9%) only in Rovers and Satellite. OptRank(GOOSE) has better coverage than PerfRank(GOOSE) in 5 domains, namely Blocksworld, Childsnack, Rovers, Satellite and Transport, with 46% (16 problems), 20% (5

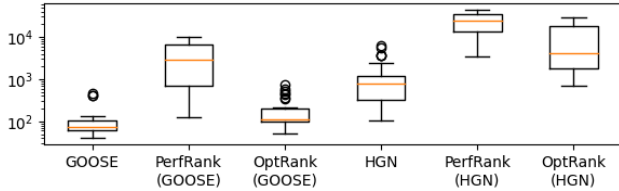


Figure 3: Convergence times (sec) grouped by configuration. Each box shows the training time of all 10 domains and 3 learned models for the configuration. PerfRank(HGN) Models whose training failed to converge are excluded.

problems), 7% (2 problems), 22% (6 problems) and 18% (5 problems) more coverage, respectively. Their coverage is the same for Sokoban and PerfRank(GOOSE) solves one more problem than OptRank(GOOSE) for Ferry, Floortile, Miconic and Spanner.

To better understand the differences between OptRank(GOOSE) and both GOOSE and PerfRank(GOOSE), Figure 4 compares them regarding the number of nodes expanded, the cost of the returned plan and the search time to find a plan. We observe that OptRank(GOOSE) generally expands fewer nodes than GOOSE while achieving similar plan quality. This means OptRank(GOOSE) spends less time exploring unpromising regions of the search space and shows that our learned ranking generalises better than the learned GOOSE heuristic to unseen problems. Although our framework guarantees to learn a total quasi-order, there is no guarantee that the learned quasi-order is an *optimal* ranking. This can be observed in the plan quality plots where other approaches find plans cheaper than OptRank(GOOSE) for some problems. We also observe that PerfRank(GOOSE) generally expands more nodes than OptRank(GOOSE) and provides plans of slightly lesser quality.

**STRIPS-HGN configurations.** Considering only configurations based on STRIPS-HGN, OptRank(HGN) has the highest coverage in all domains except in Miconic where it solves two fewer problems than HGN. As with GOOSE, the usage of optimal rankings is also highly beneficial for this regression model. Surprisingly, PerfRank(HGN)’s coverage is lower than HGN’s except on Childsnack, Floortile and Spanner. Moreover, PerfRank(HGN) did not converge for Blocksworld and Ferry. Note that our experiments use the original version of STRIPS-HGN [Shen *et al.*, 2020] whereas Chrestien *et al.* (2023) use an improved version of STRIPS-HGN. This and the different benchmark problems for each domain could explain part of the discrepancy.

**Comparison with  $h^{ff}$ .** Finally, none of the learned models were able to surpass the coverage of  $h^{ff}$ . However, OptRank(GOOSE) is not far off with just 9 fewer problems solved. This is an improvement on the competition results where the deep learning entries attempting to directly learn a heuristic or a policy did not exceed the score of HGN in our experiments. Moreover, OptRank(GOOSE) generally expands fewer nodes and finds plans of better quality than  $h^{ff}$ . Hence it is the inference time of GNN models remains an impediment to their competitiveness with special-purpose planning heuristics.

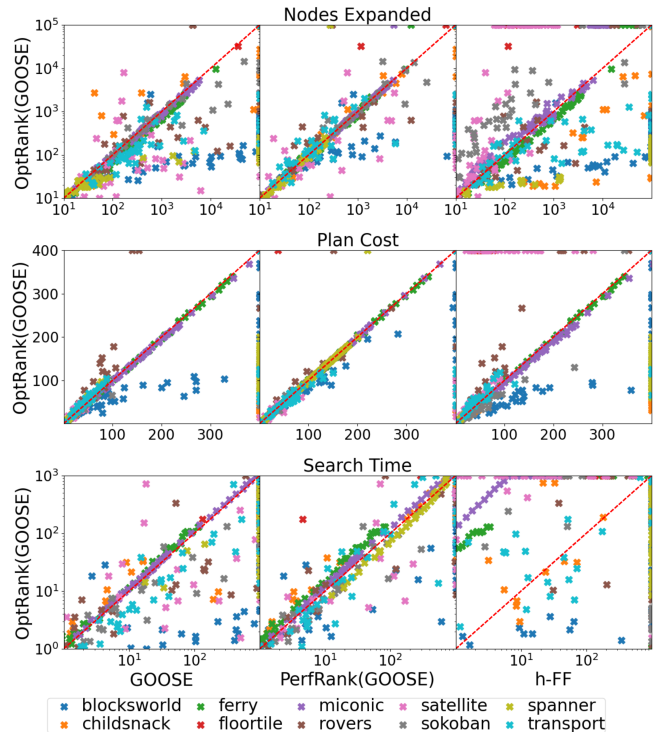


Figure 4: Comparison between OptRank(GOOSE) and GOOSE (first column), PerfRank(GOOSE) (second column), and h-FF (third column) w.r.t. the number of nodes expanded (first row), plan cost (second row) and search time (third row). Problems only solved by one of the planners are mapped to the plot’s boundaries. Points in the lower-triangle parts favor OptRank(GOOSE).

## 7 Conclusion

Heuristic search algorithms conventionally assume heuristic functions mapping states to goal-distance estimates. Yet, as was previously observed, for the popular greedy best-first search algorithm, the heuristic function merely serves to rank states, which is an overkill. To find a plan without search, as we show here, we only require an “optimal” ranking relating the states on an optimal plan to its neighbours. While this observation has yet to be used to design model-based heuristics, as we show here it is highly relevant for learning, and enables us to obtain much better performance than heuristic functions learned using the same neural network representations – even compared to a closely related ranking-learning framework.

It remains of course a bit disappointing that, at least in the benchmarks used here, the classic  $h^{ff}$  heuristic still performs best overall. This may change though with further research into representations and training algorithms for learning rankings. Apart from this, our results raise interesting questions regarding learning search guidance information for other settings and algorithms, for instance: Can our approach be extended to non-unit-cost problems? Are there other algorithms (e.g., hill-climbing or beam search) in which the heuristic function can be replaced with a simpler function that can be efficiently learned? What about variants of greedy best-first search for other forms of planning, such as top-K?

## Acknowledgements

We thank the reviewers for their comments which helped improving the paper. This work was supported by Australian Research Council grant DP220103815, by the Artificial and Natural Intelligence Toulouse Institute (ANITI) under the grant agreement ANR-19-PI3A-0004, and by the European Union’s Horizon Europe Research and Innovation program under the grant agreement TUPLES No. 101070149.

## References

- [Arfae *et al.*, 2011] Shahab Jabbari Arfae, Sandra Zilles, and Robert C. Holte. Learning Heuristic Functions for Large State Spaces. *Artificial Intelligence*, 175(16-17):2075–2098, 2011.
- [Bonet *et al.*, 1997] Blai Bonet, Gábor Loerincs, and Héctor Geffner. A Robust and Fast Action Selection Mechanism for Planning. In *Proc. National Conference on Artificial Intelligence (AAAI)*, pages 714–719, 1997.
- [Burges *et al.*, 2005] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to Rank Using Gradient Descent. In *Proc. International Conference on Machine Learning (ICML)*, pages 89–96, 2005.
- [Burges, 2010] Chris J.C. Burges. From RankNet to LambdaRank to LambdaMART: An Overview. Technical Report MSR-TR-2010-82, Microsoft Research, June 2010.
- [Chen *et al.*, 2024] Dillon Z. Chen, Sylvie Thiébaux, and Felipe Trevizan. Learning Domain-Independent Heuristics for Grounded and Lifted Planning. *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, 38(18):20078–20086, March 2024.
- [Chrestien *et al.*, 2023] Leah Chrestien, Stefan Edelkamp, Antonin Komenda, and Tomas Pevny. Optimize Planning Heuristics to Rank, Not to Estimate Cost-to-Goal. In *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, page 25508–25527, 2023.
- [Cohen and Beck, 2018] Eldan Cohen and J. Christopher Beck. Local Minima, Heavy Tails, and Search Effort for GBFS. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4708–4714, 2018.
- [Damke and Hüllermeier, 2021] Clemens Damke and Eyke Hüllermeier. Ranking Structured Objects with Graph Neural Networks. In *Proc. International Conference on Discovery Science (DS)*, pages 166–180, 2021.
- [Ferber *et al.*, 2020] Patrick Ferber, Malte Helmert, and Jörg Hoffmann. Neural Network Heuristics for Classical Planning: A Study of Hyperparameter Space. In *Proc. European Conference on Artificial Intelligence (ECAI)*, pages 2346–2353, 2020.
- [Ferber *et al.*, 2022] Patrick Ferber, Florian Geiber, Felipe Trevizan, Malte Helmert, and Jörg Hoffmann. Neural Network Heuristic Functions for Classical Planning: Bootstrapping and Comparison to Other Methods. *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, 32(1):583–587, June 2022.
- [Freund *et al.*, 2003] Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An Efficient Boosting Algorithm for Combining Preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.
- [Garrett *et al.*, 2016] Caelan Reed Garrett, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Learning to Rank for Synthesizing Planning Heuristics. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3089–3095, 2016.
- [Hao *et al.*, 2024a] Mingyu Hao, Sylvie Thiébaux, and Felipe Trevizan. Source Code for “Guiding GBFS Through Learned Pairwise Rankings”. <https://doi.org/10.5281/zenodo.11107790>, 2024.
- [Hao *et al.*, 2024b] Mingyu Hao, Felipe Trevizan, Sylvie Thiébaux, Parick Ferber, and Jörg Hoffmann. Learned Pairwise Rankings for Greedy Best-First Search. In *Proc. ICAPS Workshop on Reliable Data-Driven Planning and Scheduling*, 2024.
- [Heller *et al.*, 2022] Daniel Heller, Patrick Ferber, Julian Bitterwolf, Matthias Hein, and Jörg Hoffmann. Neural Network Heuristic Functions: Taking Confidence Into Account. In *Proc. International Symposium on Combinatorial Search (SOCS)*, pages 223–228, 2022.
- [Helmert, 2006] Malte Helmert. The Fast Downward Planning System. *Journal of Artificial Intelligence Research*, 26:191–246, 2006.
- [Herbrich *et al.*, 1999] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Large Margin Rank Boundaries for Ordinal Regression. In *Advances in Large Margin Classifiers*. The MIT Press, 1999.
- [Hoffmann and Nebel, 2001] Jörg Hoffmann and Bernhard Nebel. The FF Planning System: Fast Plan Generation Through Heuristic Search. *Journal of Artificial Intelligence Research*, 14:253–302, 2001.
- [Joachims, 2002] Thorsten Joachims. Optimizing Search Engines Using Clickthrough Data. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 133–142, 2002.
- [Karia and Srivastava, 2021] Rushang Karia and Siddharth Srivastava. Learning Generalized Relational Heuristic Networks for Model-Agnostic planning. In *Proc. AAAI Conference on Artificial Intelligence (AAAI)*, pages 8064–8073, 2021.
- [Köppel *et al.*, 2020a] Marius Köppel, Alexander Segner, Martin Wagener, Lukas Pensel, Andreas Karwath, and Stefan Kramer. Code and Supplementary Material to the Paper: Pairwise Learning to Rank by Neural Networks Revisited: Reconstruction, Theoretical Analysis and Practical Performance. <https://doi.org/10.5281/zenodo.3700819>, 2020.
- [Köppel *et al.*, 2020b] Marius Köppel, Alexander Segner, Martin Wagener, Lukas Pensel, Andreas Karwath, and Stefan Kramer. Pairwise Learning to Rank by Neural Networks Revisited: Reconstruction, Theoretical Analysis and Practical Performance. In *Proc. European Confer-*



- ence on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, pages 237–252, 2020.
- [Li, 2022] Hang Li. *Learning to Rank for Information Retrieval and Natural Language Processing*. Springer, 2022.
- [Liu and others, 2009] Tie-Yan Liu et al. Learning to Rank for Information Retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225–331, 2009.
- [Richter and Westphal, 2010] Silvia Richter and Matthias Westphal. The LAMA Planner: Guiding Cost-Based Anytime Planning with Landmarks. *Journal of Artificial Intelligence Research*, 39:127–177, 2010.
- [Röger and Helmert, 2010] Gabriele Röger and Malte Helmert. The More, the Merrier: Combining Heuristic Estimators for Satisficing planning. In *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, pages 246–249, 2010.
- [Segovia and Seipp, 2023] Javier Segovia and Jendrik Seipp. Benchmarking Repository of IPC 2023 - Learning Track. <https://github.com/ipc2023-learning/benchmarks>, 2023.
- [Seipp et al., 2020] Jendrik Seipp, Thomas Keller, and Malte Helmert. Saturated Cost Partitioning for Optimal Classical Planning. *Journal of Artificial Intelligence Research*, 67:129–167, 2020.
- [Shashua and Levin, 2002] Amnon Shashua and Anat Levin. Ranking with Large Margin Principle: Two Approaches. In *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, pages 937–944, 2002.
- [Shen et al., 2020] William Shen, Felipe Trevizan, and Sylvie Thiébaux. Learning Domain-Independent Planning Heuristics with Hypergraph Networks. In *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, volume 30, pages 574–584, 2020.
- [Ståhlberg et al., 2022] Simon Ståhlberg, Blai Bonet, and Hector Geffner. Learning General Optimal Policies with Graph Neural Networks: Expressive Power, Transparency, and Limits. In *Proc. International Conference on Automated Planning and Scheduling (ICAPS)*, volume 32, pages 629–637, 2022.
- [Wilt and Ruml, 2016] Christopher Makoto Wilt and Wheeler Ruml. Effective Heuristics for Suboptimal Best-First Search. *Journal of Artificial Intelligence Research*, 57:273–306, 2016.
- [Wu et al., 2010] Qiang Wu, Christopher J.C. Burges, Krysta M. Svore, and Jianfeng Gao. Adapting Boosting for Information Retrieval Measures. *Information Retrieval*, 13:254–270, 2010.