On Constructing Ensembles for Combinatorial Optimisation

Emma Hart
School of Computing
Edinburgh Napier University
Edinburgh, Scotland, UK
e.hart@napier.ac.uk

Kevin Sim
School of Computing
Edinburgh Napier University
Edinburgh, Scotland, UK
k.sim@napier.ac.uk

ABSTRACT

Although the use of ensemble methods in machine-learning is ubiquitous, ensembles of optimisation algorithms have received relatively little attention. In [2] we address fundamental questions regarding ensemble composition in optimisation using the domain of bin-packing as a example. We first show that ensembles constructed by *random* selection from a large pool of heuristics can outperform ensembles composed from individually *high-performing* heuristics under some conditions. We propose that this is due to the *diverse* nature of the randomly formed ensembles. Ensembles are then constructed using diversity as a criteria for inclusion. Experiments reveal that judicious choice of diversity metric is required to construct good ensembles. The results provide new insights into the how to undertake principled ensemble design.

CCS CONCEPTS

•Computing methodologies → Search methodologies;

ACM Reference format:
Emma Hart and Kevin Sim. 2017. On Constructing Ensembles for Combinatorial Optimisation. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3084379

1 INTRODUCTION

In the field of machine-learning, *ensemble-methods* that combine decisions from multiple learning algorithms to make accurate predictions have been the focus of intense research for well over a decade [4], with a broad range of empirical results underpinned by a sound theoretical understanding of the factors that influence performance.

Diversity in terms of the behaviour of ensemble members is well-known to be crucial in producing successful ensemble, as is ensemble *size* — classification ensembles often contain hundreds of classifiers [1]. In contrast, optimisation ensembles typically contain around 5 algorithms [3] and members are chosen from a pool of elite algorithms. Although some effort has been directed towards selecting algorithms that exhibit *complementarity* [3, 5] choice is still restricted to a small set of well-known methods.

In order to move the use of ensembles in optimisation forward, some basic questions need to be answered to gain the necessary insights required to construct useful ensembles. These include what algorithms to consider for inclusion, and how better to select an optimal mix for the ensemble. Specifically, we consider these questions:

- Is there a trade-off between the diversity and accuracy of the component algorithms in an ensemble? (i.e. do we have to sacrifice accuracy in order to increase diversity?)
- Under what conditions (if any) does an ensemble composed of functionally diverse algorithms outperform an ensemble of high-quality algorithms?
- Is diversity an appropriate proxy for constructing an ensemble: if so, what diversity measure is best?

2 METHOD

We conduct experiments in the 1d bin-packing domain using a set of 1370 benchmark instances. An algorithm is defined as a sequence of n low-level heuristics. Each heuristic packs a single item, and heuristics are applied in turn (cycling as necessary) until all bins are packed. Heuristics are chosen from 9 well known packing heuristics from the literature, and an algorithm is defined as sequence of length 3, enabling $9^3 = 729$ algorithms to be generated. Each algorithm is assigned a fitness defined as its average performance across the set of 1370 algorithms. An ensemble of size n is formed by either:

- **[Random Ensembles]** selecting n algorithms at random
- **[Elite Ensembles]** selecting the n best algorithms, from the ranked list

Given an instance i and an ensemble E, let V_i be a set of size $|E|$ containing the fitness of each heuristic in the ensemble on the instance. If $f_{h,i}$ is the fitness of heuristic h on instance i as specified by the objective function for the problem, then the ensemble has fitness $f_i^* = max(V_i)$, i.e. a greedy selection method assigns the fitness of an instance as the best fitness obtained from applying each heuristic in the ensemble. The collective fitness of the ensemble f_E over the complete dataset is given by equation 1:

$$f_E = \sum_{i=1}^{i=p} f_i^*, \ where \ \ f_i^* = max\{f_{h,i} : h \in E\} \tag{1}$$

GECCO '17 Companion, Berlin, Germany
 978-1-4503-4939-0/17/07...$15.00
DOI: http://dx.doi.org/10.1145/3067695.3084379

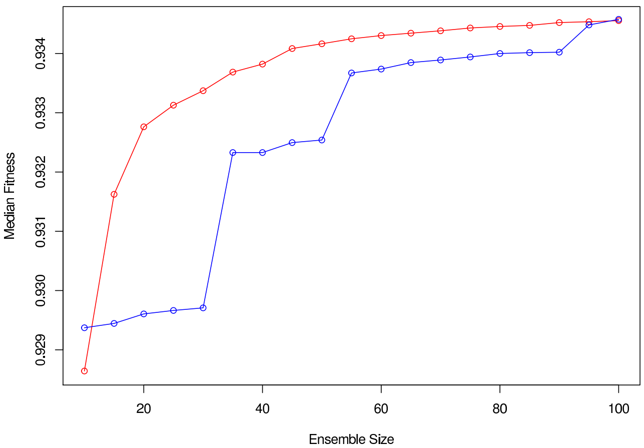

Figure 1: Median ensemble fitness for ensembles of size E. The figure shows ensembles constructed randomly (red) and the elite ensembles constructing using the best E heuristics (blue)

.

3 RESULTS

The performance of ensembles of size 10 to 100 is evaluated on the full set of 1370 instances according to equation 1. Figure 1 compares the performance of the randomly composed ensembles $\mathcal{E}_r$ and the elite ensembles $\mathcal{E}_e$. The figure shows the median fitness value obtained for the randomly composed ensembles from the 50 randomly selected ensembles for size 10-100 (after this point the results from the two approaches converge).

The main result is clear: *randomly composed ensembles outperform those composed of the best ranked heuristics for* $n < 95$. A Mann-Whitney Wilcoxon Test is used to test for significance at the 95% confidence level. No significant differences are observed between the ensembles $\mathcal{E}_r$ and $\mathcal{E}_e$ at $e = 10$, and when $e \geq 95$. At all other values of e, significant differences are observed with $p << 0.05$ in each case.

Next we ask the question *"Does explicitly seeking diversity while constructing an ensemble result in consistently good performance"* and the related question *" What is a good diversity measure for designing an ensemble learning algorithm"* An ensemble of size 1 is first created using the heuristic with best individual fitness. New heuristics are then added one at a time until the required ensemble size is reached by greedily selecting the heuristic that will maximise the average diversity of the ensemble according to a diversity metric M, based on a set of *training* instances. Following construction, the ensemble is applied to an unseen *test* set and its performance recorded. 5 different metrics are evaluated (each defined in [2]) with the result shown in figure 2. We find that only one of the diversity metrics (the "disagree metric") results in an ensemble that outperforms a randomly constructed ensemble on the test set.

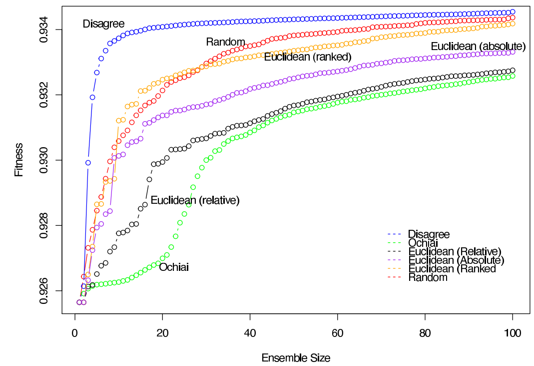

Figure 2: Average fitness of constructed ensembles per diversity metric compared to randomly selected ensemble (training set contains 685 instances, i.e. 50% of full set)

4 CONCLUSION

Using a large set of algorithms in conjunction with a large set of problem instances, we have conducted an in-depth investigation into the factors that underpin ensemble performance and ensemble construction. The full paper [2] provides three main claims with detailed discussion and empirical evidence of each:

- Diversity trumps individual ability
- Precise mechanisms for defining algorithm behaviour are required
- Choice of diversity metric matters for construction

The new insights shed light on how to create good ensembles from existing algorithms. However, by harnessing the power of genetic programming to generate novel algorithms, those insights could additionally be used to evolve even better ensembles, in which diversity is maximised.

REFERENCES

[1] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[2] E. Hart and K. Sim. 2017. On Constructing Ensembles for Combinatorial Optimisation. *Evolutionary Computation* (2017). Posted online January 10 2017.

[3] Fei Peng, Ke Tang, Guoliang Chen, and Xin Yao. 2010. Population-based algorithm portfolios for numerical optimization. *Evolutionary Computation, IEEE Transactions on* 14, 5 (2010), 782–800.

[4] Giorgio Valentini and Francesco Masulli. 2002. Ensembles of learning machines. In *Neural Nets*. Springer, 3–20.

[5] Shiu Yin Yuen and Xin Zhang. 2015. On composing an algorithm portfolio. *Memetic Computing* 7, 3 (2015), 203–214.