Supplementary File of "Evolutionary Optimization for Proactive and Dynamic Computing Resource Allocation in Open Radio Access Network"

Gan Ruan, Leandro L. Minku, *Senior Member, IEEE,* Zhao Xu, and Xin Yao, *Fellow, IEEE*

I. Dataset Description

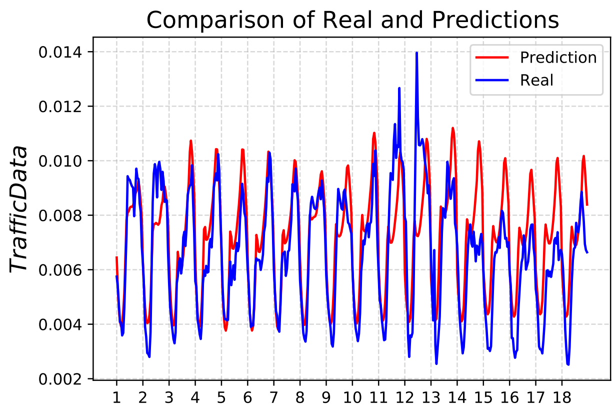
There are two sets of datasets to be used, one of which is the real-world datasets found online while another is the artificial datasets. The real-world datasets are used to verify whether the proposed approach is able to solve the real-world problems better than the greedy algorithm. However, the available realworld datasets have limited types of location and traffic dataset. In order to verify the effectiveness of the proposed algorithm on datasets with more properties, several artificial datasets with the combination of different location and different traffic dataset are generated.

A. Real-world Datasets

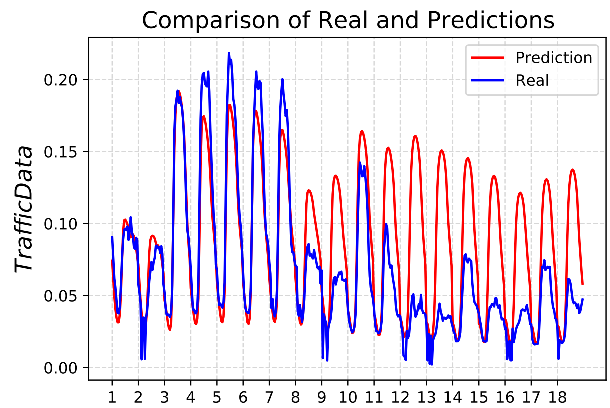
There are four real datasets found in the literature. However, considering that Archive dataset does not have the location information, three location datasets are generated to complement it. Therefore, there are six real-world datasets in total. The datasets are explained below:

- The dataset of Milan includes two months of network traffic data from 11/01/2013 to 12/31/2013 from the Telecom Italia Big Data Challenge dataset [1]. The city of Milan is partitioned into 100×100 grids with grid size of about 235×235 square meters. In each grid, the traffic volume is recorded on an hourly basis. We compile a base station dataset from CellMapper.net, which consists of the locations and coverage areas of active base stations observed in the two months. Based on the location and coverage of each base station, we find the corresponding covered grids and calculate their traffic volume. Finally, we normalize the traffic volumes of each base station to the [0;1] range for the convenience of analytics.
- The Songliao Basin dataset [2] contains movements of near 3-million anonymized cellular phone users among 167 divisions (henceforth locations), covering 4 geographically adjacent areas (Changchun City, Dehui City, Yushu City, and Nongan County) for a one-week period starting on August 7, 2017. Tis total geographic area, located in the southeast Songliao Basin in the center of the Northeast China Plain, Northeast China, covers more than 20 square kilometers. It has two files, one of which is 'Mobility.txt' describing the hourly-mobility network for the entire week. In this file, each row represents the total number of hourly movements by people from locations i to j in the corresponding day. Another file 'GPS.txt' includes the latitude and longitude information for each location in the mobility network. The pre-processing on the 'Mobility.txt' is to calculate the data of each point through adding all weights of coming to this point and originating from this point. After that, if there is one or more hours when some points do not have the data, delete those points from the 'GPS.txt' and delete those rows with the data for those points. Finally, normalize the data of each point to the $[0, 1]$ range.
- C2TM [3]: This dataset consists of individuals' activities during a continuous week (August 20 to 26, 2012), with accurate timestamps and location information indicated by the longitude and latitude of connected points. From geographic view, the monitored link covers a cellular area of around 50km60km. In order to specify principle spatio-temporal properties of cellular traffic a subarea, around 28km $*$ 35km including more than 85% of total population in both city center and suburbs, is selected for our analysis. Specifically, the selected part consists of 13K BTSs that serve more than 452K users, totally generating 379 millions of HTTP records in the measurement week.
- Archive [4]: this dataset has 57 points and the data is collected in approximately 1 year x 24 hours x 57 points. However, there are some hours when not all points have the data. Therefore, points without traffic data are deleted. There are 54 points left with the date from 12/02/2018 to 20/07/2018. In addition, this dataset does not have the location information. Therefore, three location datasets are generated. The details of the generation is explained as follows:
  - 54 points are randomly selected from the location dataset of Milan;
  - 54 points are randomly selected from the location dataset of Songliao;

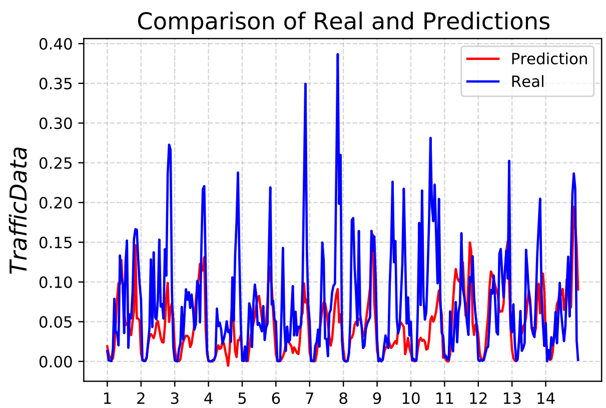
Gan Ruan, Leandro L.Minku and Xin Yao are with CERCIA, School of Computer Science, University of Birmingham, Edgbaston Birmingham B15 2TT, UK (e-mail: GXR847@student.bham.ac.uk, L.L.Minku@bham.ac.uk, xinyao@ln.edu.hk).
Zhao Xu are with NEC Laboratories Europe, 69115 Heidelberg, Germany. (email: zhao.xu@neclab.eu)
Xin Yao is also with the School of Data Science, Lingnan University, Hong Kong, China.



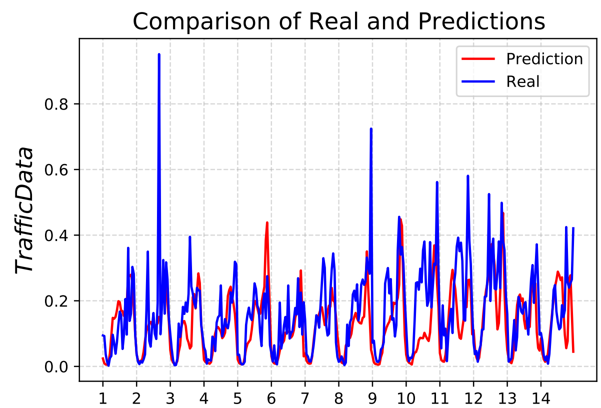
(a) The base station located in a residential district of Milan dataset from 12/14/2013 to 12/31/2013.



(b) The base station located in a business district of Milan dataset from 12/14/2013 to 12/31/2013.



(c) A random picked base station of Archive dataset from 1/6/2018 to 14/6/2018.



(d) A random picked base station of Archive dataset from 1/6/2018 to 14/6/2018.

Fig. 1. RRH traffic prediction results by LSTM for two base stations of Milan and Archive datasets. One of the reasons for large prediction error in the peaks for the Archive dataset might be that there are many high fluctuations in the period between any two days for the Archive data (Fig. 2(c) and (d)), which is difficult for the LSTM to learn and predict. Another reason might be that the change magnitude of the Archive data (from 0 to almost 0.4 in Fig. 2(c) and from 0 to almost 1 in Fig. 2(d)) is very large, compared to the change magnitude of the Milan data (from 0.002 to 0.014 in Fig. 2(a) and from 0 to around 0.22 in Fig. 2(b)).

- The third location dataset is randomly generated within a range with 54 points.

Those three datasets are named as Archive-Milan, Archive-Songliao and Archive-Random, respectively.

B. Artificial Datasets

This paragraph describes the generated artificial datasets and how they are combined with different types of generated location dataset and generated traffic dataset. There are 8 generated artificial datasets with seven days, which includes 1a, 2a, 3a 100/158, 3a 120/158, 1c-Milan, 1c-Songliao, 2b-Np=10 (Nt=174) and 2b-Np=5(Nt=185), where Np is the maximal number of points in each group for the location dataset, Nt is the total number of points in the solution. Note that in dataset 2a, the maximal number of points (Np) in each group for the location dataset is set as 5. Among those datasets, 1, 2 and 3 means the first, second and third location dataset, respectively; a, b and c means the first, second and third traffic dataset, respectively.

- There are three location datasets with different types of location information:
  1) All points are totally randomly generated within a range;
  2) High cohesion and low coupling w.r.t distance of points. It means that those points that are close to each other are in the same group. More specific, the distance of any two points in the same group is smaller than τ and the distance of any two points in different groups is larger than τ, the maximum number of generated points in each cluster is Np.
  3) Many points are gathered together while others are scattered away from those points. The distance of any two gathered points is much smaller than that of gathered point and that scattered away. Ng and Nt are the number of gathered and total points, respectively.

- There are three traffic datasets with different traffic pattern:
  1) Totally randomly generated from (0,1) for each point at 24 hours of each day;
  2) Generate the traffic data in a way that the optimal value of the objective function is known for the second case of the location dataset. More specifically, for each cluster in which all points have the distance close to each other smaller than τ, just split it into several sub-clusters and then generate the traffic data such that the total traffic data in each sub-cluster is equal to 1 at each hour of a day.
  3) Follow the pattern of existing real dataset Milan and Songliao. Those patterns are extracted from the traffic dataset of Milan and Songliao datasets. For the traffic pattern of Milan dataset, the traffic data for all points firstly decreases at the first five or six hours of each day and then increases until noon. Then, it remains stable at five or six hours and lastly it decreases. As for the traffic pattern of Songliao dataset, it firstly increases until eight or nine of each day and then remains stable for ten or eleven hour and lastly decreases. The traffic dataset based on Milan and Songliao datasets are generated following the traffic pattern of them, just as described before this sentence.

II. Experimental Results

A. Comparison Results of SplitEA and the Greedy Algorithm

1) Prediction Results of LSTM on Milan and Archive Datasets: This section presents the prediction errors of the LSTM model on two real-world datasets Milan and Archive. The RRH traffic prediction results by LSTM for two randomly picked base station from Milan and Archive datasets are shown in Figure 1. For Milan dataset, a residential and a business areas are picked with the results shown in Figure 1 (a) and (b), respectively. For Archive dataset, two randomly selected picked base stations with the results shown in Figure 1 (c) and (d), respectively. It is clear from Figure 1 (b) that at the week of Christmas, the prediction errors on the business ares is much larger than that of other days. The reason is analyzed as follows. Figure 1(b) presents the base station located in a business district of Milan dataset from 12/14/2013 to 12/31/2013. The date from 12/25/2013 to 12/31/2013 is the week of Christmas, so few people went to the business district where their companies locates, resulting in few traffic data. However, the predicted traffic data is learnt based on the week before the Christmas week when people worked as usual at business areas.

2) Comparison results of SplitEA and greedy algorithm on artificial datasets: In order to test the ability of the SplitEA in searching for good solutions over the greedy algorithm on datasets with more properties beyond those of the existing datasets, like different types of location information and traffic patterns, six artificial datasets with different points distribution and different traffic pattern have been generated to do the verification. The comparison results of the greedy algorithm and SplitEA on the artificial traffic dataset are shown in Table 1. It is clear from this table that SplitEA significantly performs better than the greedy algorithm regarding the fitness values on all artificial datasets except for datasets $2a$ and $2b - Np = 5(Nt = 185)$, which proves that the greedy algorithm is able to find better solutions than the greedy algorithm. As for the specific metrics, SplitEA gets significant better values of all metrics except for $Udelay$ on two datasets $1c - Milan$ and $1c - Songliao$. On other datasets, SplitEA gets significantly better K and $Uunder1$ while worse U and $Udelay$.

The reasons are analyzed as follows. It is intuitive to get the reason why SplitEA gets worse $Udelay$ that SplitEA tends to cluster more points together due to the less required number of clusters for fixed number of points and this would inevitably increase the delay in the network. As for the reason why SplitEA gets worse U on all datasets except for $1c - Milan$ and $1c - Songliao$, it is because the traffic datasets for all other datasets except for $1c - Milan$ and $1c - Songliao$ are artificially generated, which results in much larger mean traffic value than that of real traffic datasets. In this case, when SplitEA clusters more points together, it would get much more $Udelay$, further increasing the value of U. As for the reason why SplitEA gets worse fitness value on datasets $2a$ and $2b - Np = 5(Nt = 185)$, mutation process tends to cluster points in each group together on the artificial datasets with the second location dataset especially when the maximal number of points in each group is small. This is the reason why SplitEA gets better fitness value on dataset $2b - Np = 10(Nt = 174)$ where maximal number of points in each group is 10.

3) The effect of different parameter settings on the performance of SplitEA and greedy algorithm: In order to check the influence of different parameter settings on the performance of SplitEA and the greedy algorithm, two methods are tested on the Milan Dataset which sets different values for two problem-related parameters w and τ. The comparison results of SplitEA and greedy algorithm on dataset Milan under different parameter settings are presented in Table 2. It is clear from this table that under different settings of w, SplitEA significantly performs better than the greedy algorithm on all metrics and the fitness value. In addition, SplitEA gets significantly better results the greedy algorithm on all metrics under all settings of τ except for the setting of $\tau = 1500$ on the metric $Udelay$. The reason might be that if τ is too large, SplitEA tends to cluster many points together, which inevitably causes more delay than the greedy algorithm causes.

B. Analyses of Random Cluster Splitting in SplitEA

1) Comparison results of SplitEA and two variants on artificial datasets.: In order to test the performance of the SplitEA in searching for good solutions over RandEA and CopyEA on different scenarios with different points distribution and different traffic pattern, several artificial dataset have been generated to do the verification. The property of those datasets are described

TABLE 1
COMPARISON RESULTS OF SPLITEA AND THE GREEDY ALGORITHM ON ARTIFICIAL DATASETS.

Datasets	Dataset 1a		Dataset 2a		Dataset 3a 100/158		Dataset 3a 120/158	
Algorithms	GreedyAlg	SplitEA	GreedyAlg	SplitEA	GreedyAlg	SplitEA	GreedyAlg	SplitEA
K	78.2143	64.8381	78.2333	69.7095	135.9524	67.8381	132.2476	69.3524
U	0.3453	0.4284	0.4197	0.5104	0.4556	0.5489	0.4456	0.5272
Udelay	0.1779	0.3233	0.2150	0.3218	0.0184	0.3585	0.0215	0.3341
Uunder1	0.1675	0.1051	0.2046	0.1886	0.4372	0.1904	0.4241	0.1932
f	1.1275	1.0768	1.2020	1.2075	1.8151	1.2273	1.7681	1.2208

Datasets	Dataset 1c-Milan		Dataset 1c-Songliao		Dataset 2b-Np=10 (Nt=174)		Dataset 2b-Np=5 (Nt=185)	
Algorithms	GreedyAlg	SplitEA	GreedyAlg	SplitEA	GreedyAlg	SplitEA	GreedyAlg	SplitEA
K	60.8381	46.5810	56.3381	46.2476	70.6476	45.9190	76.4095	56.2286
U	0.6018	0.5326	0.6585	0.6049	0.2093	0.3738	0.0707	0.4131
Udelay	0.0043	0.0315	0.0044	0.0153	0.0297	0.3396	0.0438	0.3982
Uunder1	0.5975	0.5011	0.6541	0.5896	0.1796	0.0342	0.0270	0.0149
f	1.2102	0.9985	1.2219	1.0673	0.9158	0.8329	0.8348	0.9753

There are 30 independent runs. The values in this table are the mean value of the metrics under 30 runs. Friedman and Nemenyi statistical tests [5] with the significance level 0.05 are used to indicate the statistical significance between compared algorithms. The metric value obtained by a given algorithm on one dataset is regarded as an observation to compose that algorithms group for the test, following Demsars guidelines [5]. Therefore, there are 30 observations in each group for each metric on each dataset. The significantly better values obtained by the algorithm are highlighted in red color.

TABLE 2
THE EFFECT OF PARAMETERS w AND τ IN THE PROBLEM ON THE PERFORMANCE OF SPLITEA AND GREEDY ALGORITHM.

w	0.001		0.01		1	
Algorithms	GreedyAlg	SplitEA	GreedyAlg	SplitEA	GreedyAlg	SplitEA
K	61.1725	53.2882	61.1863	52.0980	59.9784	51.7137
U	0.8326	0.7623	0.8323	0.7659	0.8295	0.7684
Udelay	0.0236	0.0026	0.0238	0.0077	0.0244	0.0100
Uunder1	0.8090	0.7596	0.8086	0.7582	0.8051	0.7583
f	0.8937	0.8156	1.4442	1.2869	60.8079	52.4821

τ	800		1000		1500	
Algorithms	GreedyAlg	SplitEA	GreedyAlg	SplitEA	GreedyAlg	SplitEA
K	68.9020	62.6255	59.9784	51.7137	92.9392	35.0765
U	0.8397	0.8048	0.8295	0.7684	0.8823	0.6681
Udelay	0.0151	0.0064	0.0244	0.0100	0.0122	0.0185
Uunder1	0.8246	0.7983	0.8051	0.7583	0.8702	0.6496
f	69.7416	63.4303	60.8079	52.4821	93.8215	35.7446

There are 30 independent runs. The values in this table are the mean value of the metrics under 30 runs. Friedman and Nemenyi statistical tests [5] with the significance level 0.05 are used to indicate the statistical significance between compared algorithms. The metric value obtained by a given algorithm on one dataset is regarded as an observation to compose that algorithms group for the test, following Demsars guidelines [5]. Therefore, there are 30 observations in each group for each metric on each parameter setting. The significantly better values obtained by the algorithm are highlighted in red color.

in Section I. The comparison results of three EAs on the artificial predicted traffic dataset are shown in Table 3. It is clear from this table that SplitEA significantly performs best than RandEA and CopyEA regarding the fitness values on all artificial datasets except for two datasets $2b$, which proves that our proposed SplitEA is able to find better solutions than two variant in most cases.

The reason why RandEA gets better fitness values than SplitEA on two datasets $2b$ might be that the diversity introduction through splitting a random cluster into two clusters is not enough for those two datasets to find more clustering structure that gets better U, which can be reflected by the results of U. Following this line of diversity introduction, the reason why RandEA and SplitEA perform equally on Dataset 2a is that the diversity introduction in two EAs is somehow equal while the emphasis is different. In addition, SplitEA performs best regarding all metrics excepts for $Udelay$ on two Datasets $1c$. Similarly, the reason is that when SplitEA tries to cluster more points together, it might result in much delay. Besides, SplitEA gets best values regarding all metrics excepts for K on two Datasets $3a$. The reason might be that the spitting of cluster in SplitEA causes solutions with better fitness values while inevitably increases the number of cluster.

2) Fitness value curve of SplitEA and two variants on artificial datasets.: In order to have a better understanding of the proposed SplitEA and its two variants in the evolution process, the fitness value curve of three EAs on all artificial datasets in a randomly selected run is drawn and presented in Figures 2 and 3. As three EAs all initialize the population randomly at the first day, the curve of all days except for the first day is presented. It clear from those two figures that during the whole optimization process, the fitness vale of SplitEA is always better than that of RandEA and CopyEA on four datasets 3as and 1cs. As for the comparison results on datasets 1a and 2a, SplitEA is only worse than CopyEA on first day of all days.

3) The effect of different parameter settings on the performance of SplitEA and two variants: In order to check the influence of different parameter settings on the performance of SplitEA and two variants, three algorithms are tested on the Milan Dataset which sets different values for two problem-related parameters w and τ and three EA-related parameters $prob$, G and $popsize$. The comparison results of SplitEA and two variants on dataset Milan under different parameter settings are presented in Table 4. It is clear from this table that under different settings of w, SplitEA significantly performs better than the greedy algorithm

TABLE 3
COMPARISON RESULTS OF THREE EAS ON ARTIFICIAL DATASETS.

Datasets	Dataset 1a			Dataset 2a			Dataset 3a 100/158			Dataset 3a 120/158		
Algorithms	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA
K	64.7429	64.5714	64.8381	70.5190	70.1429	69.7095	61.4095	**63.4667**	67.8381	63.6952	**66.3429**	69.3524
U	**0.4343**	0.4531	0.4284	0.5036	0.5127	0.5104	0.6849	**0.6391**	0.5489	0.6307	**0.5809**	0.5272
Udelay	0.3272	**0.3381**	0.3233	0.3120	**0.3195**	0.3218	0.4858	**0.4425**	0.3585	0.4356	**0.3860**	0.3341
Uunder1	0.1072	**0.1149**	0.1051	**0.1916**	**0.1932**	0.1886	0.1991	**0.1966**	0.1904	**0.1951**	**0.1949**	0.1932
f	**1.0818**	1.0988	1.0768	1.2088	**1.2141**	1.2075	1.2990	**1.2738**	1.2273	1.2677	**1.2443**	1.2208
Datasets	Dataset 1c-Milan			Dataset 1c-Songliao			Dataset 2b-Np=10 (d=174)			Dataset 2b-Np=5 (d=185)		
Algorithms	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA
K	48.5571	**47.1667**	46.5810	**48.2952**	**48.4571**	46.2476	**49.0857**	**49.6571**	45.9190	**61.5905**	**61.4000**	56.2286
U	**0.5466**	0.5566	0.5326	**0.6193**	0.6388	0.6049	0.3149	0.4777	**0.3738**	0.3072	**0.4051**	**0.4131**
Udelay	0.0277	0.0403	**0.0315**	0.0135	**0.0226**	0.0153	0.2678	**0.3421**	**0.3396**	0.2839	**0.3350**	0.3982
Uunder1	**0.5189**	**0.5163**	0.5011	**0.6058**	0.6161	0.5896	**0.0471**	0.1356	0.0342	**0.0233**	0.0701	0.0149
f	**1.0322**	**1.0283**	0.9985	**1.1022**	1.1233	1.0673	0.8057	0.9743	**0.8329**	0.9231	1.0191	**0.9753**

There are 30 independent runs. The values in this table are the mean value of the metrics under 30 runs. The best and the second best values obtained by the algorithm are highlighted in red and bold face, respectively. Friedman and Nemenyi statistical tests [5] with the significance level 0.05 are used to indicate the statistical significance between compared algorithms. The metric value obtained by a given algorithm on one dataset is regarded as an observation to compose that algorithms group for the test, following Demsars guidelines [5]. Therefore, there are 30 observations in each group for each metric on each dataset. If results obtained by two or three algorithms are highlighted with the same mark, it meas there is no significant difference between them.

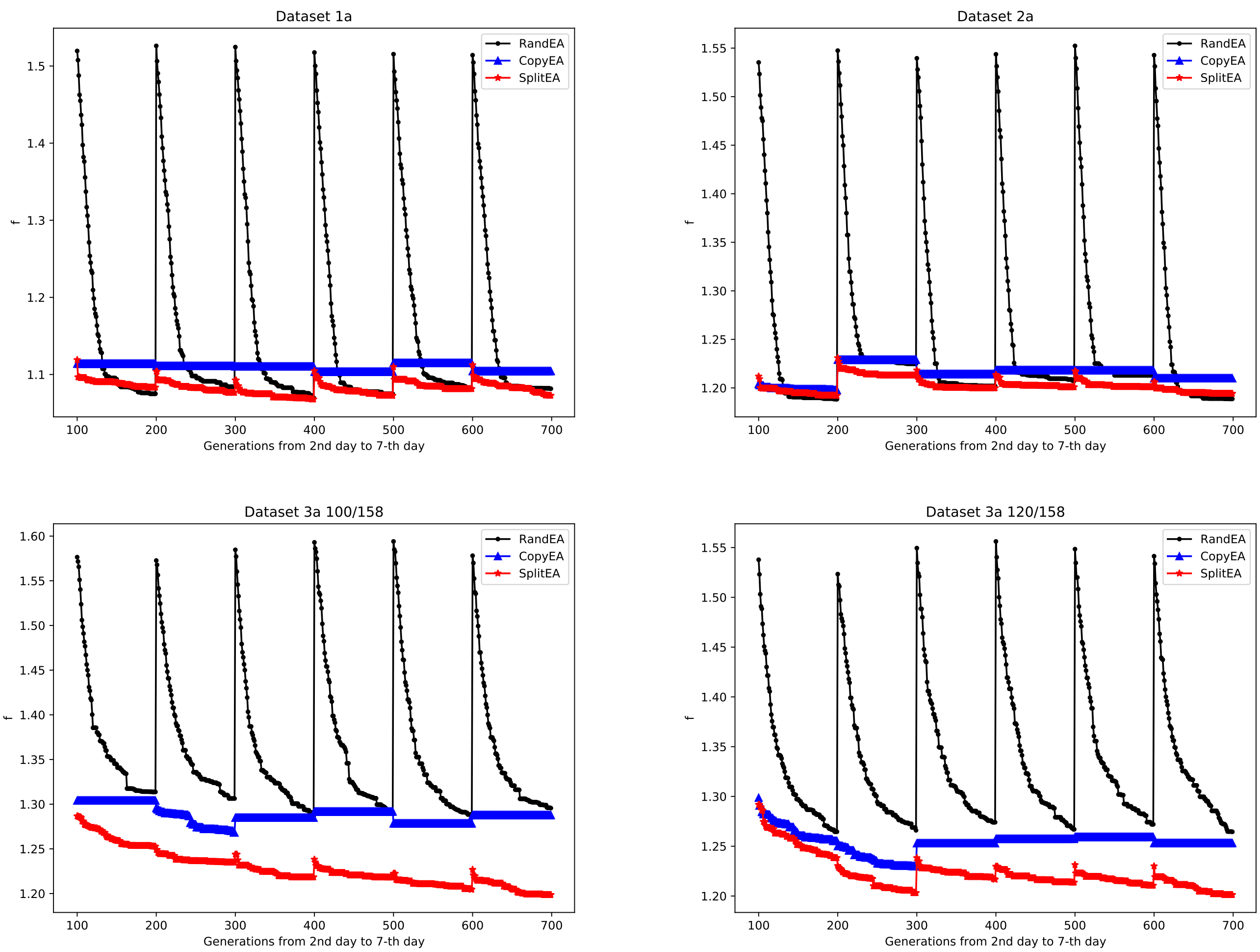


Fig. 2. Curve of the fitness value obtained by SplitEA and two variants on four artificial datasets across the whole evolution process.

on all metrics and the fitness value. It is clear from this table that under different settings of all parameters, SplitEA gets best results, which shows that the proposed SplitEA is not sensitive with parameter setting.

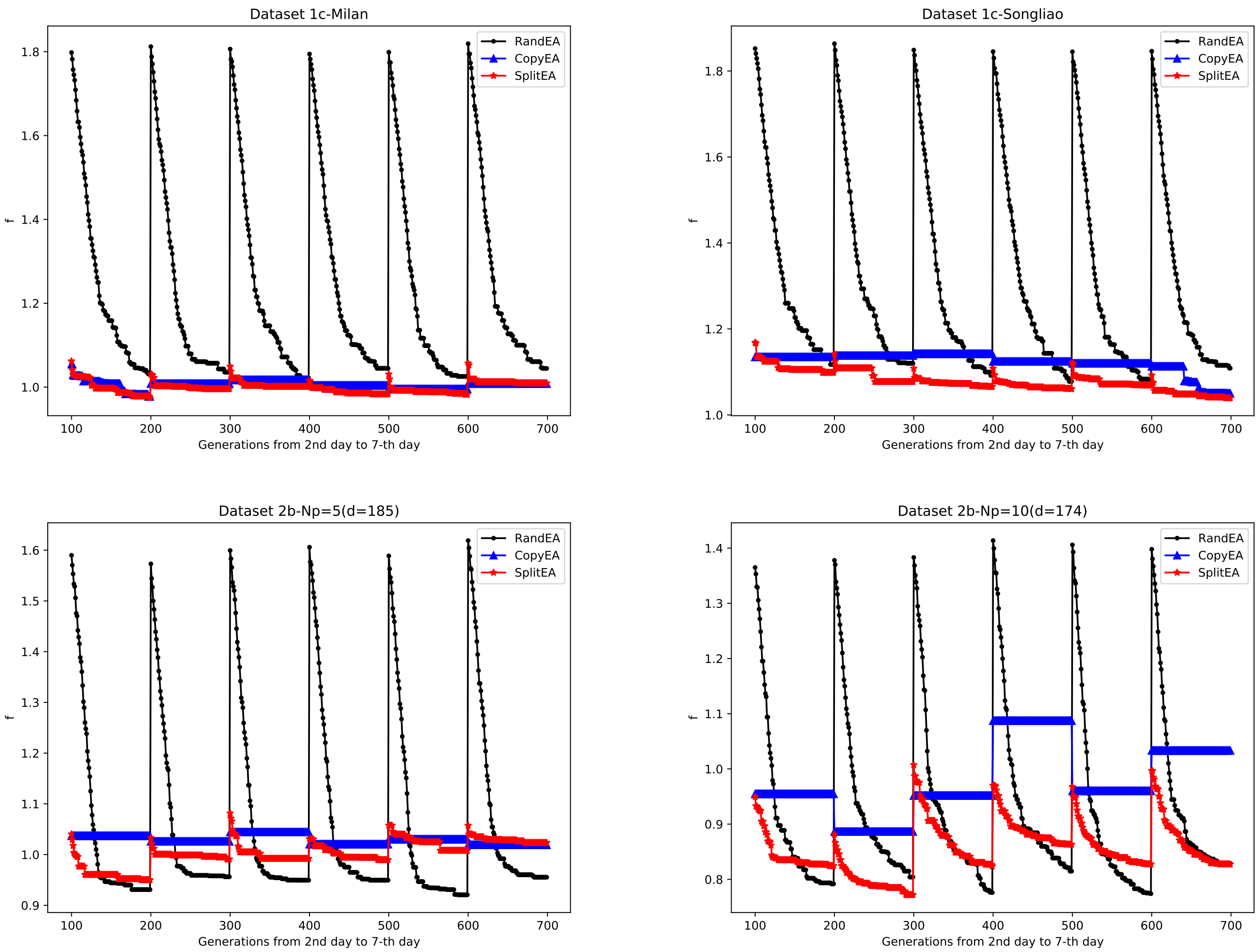


Fig. 3. Curve of the fitness value obtained by SplitEA and two variants on four artificial datasets across the whole evolution process.

References

[1] Gianni Barlacchi, Marco De Nadai, Roberto Larcher, Antonio Casella, Cristiana Chitic, Giovanni Torrisi, Fabrizio Antonelli, Alessandro Vespignani, Alex Pentland, and Bruno Lepri. A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific Data*, 2(1):1–15, 2015.

[2] Zhanwei Du, Yongjian Yang, Zeynep Ertem, Chao Gao, Liping Huang, Qiuyang Huang, and Yuan Bai. Inter-urban mobility via cellular position tracking in the southeast songliao basin, northeast china. *Scientific Data*, 6(1):1–6, 2019.

[3] Xiaming Chen, Yaohui Jin, Siwei Qiang, Weisheng Hu, and Kaida Jiang. Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale. In *2015 IEEE International Conference on Communications (ICC)*, pages 3585–3591. IEEE, 2015.

[4] Rehenuma Tasnim Rodoshi, Taewoon Kim, and Wooyeol Choi. Deep reinforcement learning based dynamic resource allocation in cloud radio access networks. In *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 618–623. IEEE, 2020.

[5] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(Jan):1–30, 2006.

TABLE 4
THE EFFECT OF PROBLEM-RELATED PARAMETERS (w AND τ) AND EA-RELATED PARAMETERS ($prob$, G AND POPSIZE) ON THE PERFORMANCE OF SPLITEA AND TWO VARIANTS.

w	0.001			0.01			1		
Algorithms	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA
K	56.4863	**56.0373**	53.2882	55.5353	**53.2020**	52.0980	55.3647	**52.3333**	51.7137
U	0.7758	**0.7712**	0.7623	0.7788	**0.7687**	0.7659	0.7802	**0.7706**	0.7684
Udelay	**0.0027**	0.0020	**0.0026**	0.0067	**0.0069**	0.0077	0.0079	**0.0099**	**0.0100**
Uunder	0.7730	**0.7691**	0.7596	0.7721	**0.7618**	0.7582	0.7723	**0.7607**	0.7583
f	0.8323	**0.8272**	0.8156	1.3342	**1.3007**	1.2869	56.1449	**53.1040**	52.4821
τ	800			1000			1500		
Algorithms	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA
K	65.4824	**62.9980**	62.6255	55.3647	**52.3333**	51.7137	40.5020	**36.9176**	35.0765
U	0.8124	**0.8056**	0.8048	0.7802	**0.7706**	0.7684	0.7061	**0.6806**	0.6681
Udelay	0.0058	**0.0063**	0.0064	0.0079	**0.0099**	0.0100	0.0137	**0.0169**	0.0185
Uunder	0.8066	**0.7993**	0.7983	0.7723	**0.7607**	0.7583	0.6924	**0.6637**	0.6496
f	66.2947	**63.8037**	63.4303	56.1449	**53.1040**	52.4821	41.2081	**37.5982**	35.7446
prob	0.2			0.5			0.8		
Algorithms	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA
K	57.6667	**53.0333**	52.4118	55.3647	**52.3333**	51.7137	57.6667	**53.0333**	51.4843
U	0.7865	**0.7711**	0.7684	0.7802	**0.7706**	0.7684	0.7865	**0.7711**	0.7673
Udelay	0.0062	0.0085	**0.0083**	0.0079	**0.0099**	**0.0100**	0.0062	**0.0085**	0.0101
Uunder	**0.7802**	0.7626	0.7601	0.7723	**0.7607**	0.7583	0.7802	**0.7626**	0.7572
f	58.4531	**53.8045**	53.1801	56.1449	**53.1040**	52.4821	58.4531	**53.8045**	52.2516
G/popsize	75/20			150/10			300/5		
Algorithms	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA	RandEA	CopyEA	SplitEA
K	57.9608	**52.1157**	51.9588	55.3647	**52.3333**	51.7137	55.5961	**52.4314**	51.9451
U	0.7927	**0.7718**	0.7697	0.7802	**0.7706**	0.7684	0.7788	**0.7708**	0.7680
Udelay	0.0089	0.0110	**0.0101**	0.0079	**0.0099**	**0.0100**	0.0071	0.0098	**0.0093**
Uunder	0.7838	**0.7608**	0.7596	0.7723	**0.7607**	0.7583	0.7717	**0.7610**	0.7587
f	58.7535	**52.8875**	52.7285	56.1449	**53.1040**	52.4821	56.3749	**53.2021**	52.7131

There are 30 independent runs. The values in this table are the mean value of the metrics under 30 runs. The best and the second best values obtained by the algorithm are highlighted in red and bold face, respectively. Friedman and Nemenyi statistical tests [5] with the significance level 0.05 are used to indicate the statistical significance between compared algorithms. The metric value obtained by a given algorithm on one dataset is regarded as an observation to compose that algorithms group for the test, following Demsars guidelines [5]. Therefore, there are 30 observations in each group for each metric on each parameter setting. If results obtained by two or three algorithms are highlighted with the same mark, it meas there is no significant difference between them.