VNUHCM

University of Information Technology

Report for Knapsack solutions

Nguyen Duy Hoang
22520467
CS106.O21

Contents

1 Experiment Set-up **1**
1.1 Data extraction 1
1.2 Using Google OR-Tools 1

2 Statistics and Evaluation of Knapsack Problem Results **1**
2.1 Statistics 1
2.2 Evaluation 3

1 Experiment Set-up

1.1 Data extraction

In the process of experiment setup, the `data.zip` file was acquired from the provided repository located at `https://github.com/likr/kplib`, and subsequently underwent extraction. Due to the substantial magnitude of the dataset, consisting of 20,800 `.kp` files, executing all samples was deemed unfeasible. Consequently, a random sampling methodology was devised to ensure the selection of a representative subset. This sampling method was performed on the files:

- *change_folder_name.py*: I compiled all folder names into a single array named `folder_name`. Subsequently, for each folder within this array, I systematically renamed its subfolders with numerical labels ranging from 1 to 8, mirroring their sequential order within the original directory structure.
- *random_data.py*: After that, I iterated through each subfolder from 1 to 8. And with each subfolder, I randomly selected 5 files from the subsubfolders $R01000$ and $R10000$, respectively. Then, I wrote the paths to the selected files into the *testcases_path.txt* file.

1.2 Using Google OR-Tools

Google OR-Tools is an open-source optimization toolkit that offers efficient optimization algorithms. When applied to the knapsack problem, OR-Tools provides effective solving methods to find ways to pack items into a limited-capacity bag, aiming to maximize value.

When using this tool for the Knapsack problem, I will utilize the function `set_time_limit(180)` to restrict the solution time to 180 seconds, equivalent to 3 minutes. According to my research at `https://developers.google.com/optimization/reference/python/algorithms/pywrapknapsack_solver#set_time_limit`, when a finite time limit is set, the solution obtained might not be optimal if the limit is reached.

Therefore, if the problem finds a solution before the time limit, we refer to it as an optimal or best solution. However, if the problem-solving time is 180 seconds, corresponding to the set time limit, then the solution obtained is not optimal.

2 Statistics and Evaluation of Knapsack Problem Results

2.1 Statistics

The **output** folder contains all the solutions of the randomly selected test cases as described in the above **Experiment Set-up**. For this report, we only sampled a subset of test cases from 13 folders: 00Uncorrelated, 01WeaklyCorrelated, 02StronglyCorrelated, 03InverseStronglyCorrelated, 04AlmostStronglyCorrelated, 05SubsetSum, 06UncorrelatedWithSimilarWeights, 07SpannerUncorrelated, 08SpannerWeaklyCorrelated, 09SpannerStronglyCorrelated, 10MultipleStronglyCorrelated, 11ProfitCeiling, 12Circle. In each folder, we selected one test case for each N_items value ranging from n00500 to n20000.

- **Test case:** This represents the path to the test case file.
- **Time:** The time taken to find the solution and measured in seconds. This time is limited to a maximum of 180 seconds.
- **Value:** The total value of the optimally packed items.
- **Weight:** The total weight of the optimally packed items.
- **Optimal:** True if it is an optimal solution. If the time taken to find the solution is less than 180 seconds, it will be marked as True. However, if the time taken equals 180 seconds, which is equivalent to the set_time_limit(), it will be marked as False.

Testcase	N_items	Execution Time	Capacity	Total Value	Total Weight	Optimal
/00Uncorrelated/1/R10000/s030	50	0.000	127555	201461	127530	True
/00Uncorrelated/2/R01000/s051	100	0.000	26220	45629	26204	True
/00Uncorrelated/3/R10000/s030	200	0.001	474730	809165	474569	True
/00Uncorrelated/4/R01000/s049	500	0.000	122191	196691	122178	True
/00Uncorrelated/5/R01000/s022	1000	0.000	247104	400290	247103	True
/00Uncorrelated/6/R10000/s010	2000	0.001	4924266	8166357	4924263	True

Continued on next page

Testcase	N_items	Execution Time	Capacity	Total Value	Total Weight	Optimal
/00Uncorrelated/7/R01000/s024	5000	0.000	1192393	2067317	1192392	True
/00Uncorrelated/8/R10000/s017	10000	0.000	24670351	40300050	24670348	True
/01WeaklyCorrelated/1/R01000/s094	50	0.000	11888	13175	11885	True
/01WeaklyCorrelated/2/R01000/s025	100	0.000	23220	25875	23220	True
/01WeaklyCorrelated/3/R10000/s024	200	0.000	502927	561884	502915	True
/01WeaklyCorrelated/4/R10000/s079	500	0.000	1201558	1336791	1201556	True
/01WeaklyCorrelated/5/R01000/s066	1000	0.000	248350	273740	248350	True
/01WeaklyCorrelated/6/R01000/s064	2000	0.000	494251	548275	494251	True
/01WeaklyCorrelated/7/R01000/s047	5000	0.032	1243366	1375240	1243366	True
/01WeaklyCorrelated/8/R01000/s047	10000	0.016	2484197	2742854	2484197	True
/02StronglyCorrelated/1/R10000/s083	50	0.386	122235	156235	122235	True
/02StronglyCorrelated/2/R10000/s091	100	1.758	270311	338311	270311	True
/02StronglyCorrelated/3/R01000/s050	200	20.606	47460	61760	47460	True
/02StronglyCorrelated/4/R01000/s075	500	180.046	122676	157500	122400	False
/02StronglyCorrelated/5/R01000/s073	1000	180.029	241402	311690	241090	False
/02StronglyCorrelated/6/R10000/s018	2000	180.001	4963941	6359261	4959261	False
/02StronglyCorrelated/7/R10000/s036	5000	179.968	12330501	15846553	12325553	True
/02StronglyCorrelated/8/R10000/s015	10000	179.709	24775197	31808958	24769958	True
/03InverseStronglyCorrelated/1/R01000/s092	50	1.083	16307	14524	16224	True
/03InverseStronglyCorrelated/2/R10000/s026	100	0.000	289321	259320	289320	True
/03InverseStronglyCorrelated/3/R01000/s016	200	180.043	56460	50260	56460	False
/03InverseStronglyCorrelated/4/R01000/s020	500	180.038	143201	128001	143201	False
/03InverseStronglyCorrelated/5/R01000/s034	1000	180.012	306206	273966	306166	False
/03InverseStronglyCorrelated/6/R10000/s017	2000	180.024	5959872	5326865	5955865	False
/03InverseStronglyCorrelated/7/R10000/s030	5000	179.978	14777352	13204795	14772795	True
/03InverseStronglyCorrelated/8/R01000/s088	10000	179.901	2964911	2648835	2964435	True
/04AlmostStronglyCorrelated/1/R01000/s036	50	0.000	13946	17336	13946	True
/04AlmostStronglyCorrelated/2/R01000/s055	100	0.440	24611	31424	24611	True
/04AlmostStronglyCorrelated/3/R10000/s011	200	180.076	497719	638830	497716	False
/04AlmostStronglyCorrelated/4/R10000/s079	500	2.579	1201558	1554738	1201557	True
/04AlmostStronglyCorrelated/5/R10000/s011	1000	180.011	2512521	3207714	2508949	False
/04AlmostStronglyCorrelated/6/R01000/s092	2000	0.487	494716	635011	494716	True
/04AlmostStronglyCorrelated/7/R10000/s050	5000	179.945	12361648	15875198	12356757	True
/04AlmostStronglyCorrelated/8/R10000/s092	10000	179.713	24463602	31512923	24458957	True
/05SubsetSum/1/R10000/s025	50	0.000	113745	113745	113745	True
/05SubsetSum/2/R10000/s005	100	0.000	246212	246212	246212	True
/05SubsetSum/3/R10000/s066	200	0.016	498185	498185	498185	True
/05SubsetSum/4/R01000/s042	500	0.000	124243	124243	124243	True
/05SubsetSum/5/R01000/s013	1000	0.000	243197	243197	243197	True
/05SubsetSum/6/R01000/s093	2000	0.000	503252	503252	503252	True
/05SubsetSum/7/R10000/s007	5000	0.008	12247962	12247962	12247962	True
/05SubsetSum/8/R10000/s029	10000	0.016	24605208	24605208	24605208	True
/06UncorrelatedWithSimilarWeights/1/R10000/s059	50	0.078	2476461	19370	2401115	True
/06UncorrelatedWithSimilarWeights/2/R01000/s075	100	0.110	4952887	37667	4902207	True
/06UncorrelatedWithSimilarWeights/3/R10000/s034	200	0.000	9906026	77405	9904549	True
/06UncorrelatedWithSimilarWeights/4/R10000/s016	500	180.016	24764986	182355	24712578	False
/06UncorrelatedWithSimilarWeights/5/R10000/s014	1000	0.133	49529345	369546	49525194	True
/06UncorrelatedWithSimilarWeights/6/R01000/s012	2000	180.015	99060009	746841	99050084	False
/06UncorrelatedWithSimilarWeights/7/R10000/s089	5000	179.967	247648444	1865038	247624250	True
/06UncorrelatedWithSimilarWeights/8/R10000/s032	10000	179.815	495296685	3717155	495250511	True
/07SpannerUncorrelated/1/R10000/s004	50	175.291	36554	22976	36480	True
/07SpannerUncorrelated/2/R01000/s030	100	180.042	8221	15630	8170	False
/07SpannerUncorrelated/3/R10000/s013	200	180.035	350759	282842	350488	False
/07SpannerUncorrelated/4/R10000/s089	500	180.046	654814	830298	654266	False
/07SpannerUncorrelated/5/R01000/s025	1000	171.168	130450	266204	130435	True
/07SpannerUncorrelated/6/R10000/s007	2000	180.005	1754011	1440553	1753555	False
/07SpannerUncorrelated/7/R01000/s065	5000	179.983	801106	481409	801084	True
/07SpannerUncorrelated/8/R01000/s030	10000	179.752	806538	1700665	806512	True
/08SpannerWeaklyCorrelated/1/R10000/s057	50	179.093	43709	64137	43682	True
/08SpannerWeaklyCorrelated/2/R10000/s068	100	0.110	206737	407337	206729	True
/08SpannerWeaklyCorrelated/3/R10000/s068	200	180.054	392120	771730	391590	False
/08SpannerWeaklyCorrelated/4/R10000/s041	500	180.036	378450	1647110	378334	False
/08SpannerWeaklyCorrelated/5/R10000/s065	1000	180.040	865822	2672544	865592	False
/08SpannerWeaklyCorrelated/6/R10000/s086	2000	180.009	1996343	5485897	1996293	False
/08SpannerWeaklyCorrelated/7/R10000/s028	5000	179.968	1505998	9384304	1505945	True
/08SpannerWeaklyCorrelated/8/R01000/s098	10000	179.837	1451024	1372442	1451020	True
/09SpannerStronglyCorrelated/1/R01000/s053	50	0.016	8751	23937	8737	True
/09SpannerStronglyCorrelated/2/R01000/s082	100	180.036	8237	44635	8235	False
/09SpannerStronglyCorrelated/3/R01000/s033	200	180.041	29374	86673	29373	False
/09SpannerStronglyCorrelated/4/R10000/s018	500	180.043	545122	2379059	545059	False
/09SpannerStronglyCorrelated/5/R01000/s072	1000	180.031	85761	481458	85758	False
/09SpannerStronglyCorrelated/6/R10000/s011	2000	177.481	2483662	8388603	2483603	True
/09SpannerStronglyCorrelated/7/R10000/s020	5000	179.969	9689660	24883885	9688885	True

Continued on next page

Testcase	N_items	Execution Time	Capacity	Total Value	Total Weight	Optimal
/09SpannerStronglyCorrelated/8/R10000/s041	10000	179.804	7544845	40225784	7544784	True
/10MultipleStronglyCorrelated/1/R01000/s082	50	0.000	11964	19364	11964	True
/10MultipleStronglyCorrelated/2/R10000/s011	100	0.050	233898	389898	233898	True
/10MultipleStronglyCorrelated/3/R01000/s009	200	17.058	46602	78302	46602	True
/10MultipleStronglyCorrelated/4/R10000/s065	500	180.021	1258980	2016980	1258980	False
/10MultipleStronglyCorrelated/5/R10000/s056	1000	180.032	2481342	4020342	2481342	False
/10MultipleStronglyCorrelated/6/R01000/s079	2000	180.015	491946	801635	491935	False
/10MultipleStronglyCorrelated/7/R01000/s058	5000	179.958	1248000	2019962	1247862	True
/10MultipleStronglyCorrelated/8/R10000/s031	10000	179.741	25008090	40398887	25006887	True
/11ProfitCeiling/1/R10000/s076	50	0.016	132392	132387	132391	True
/11ProfitCeiling/2/R01000/s077	100	180.042	24216	24204	24214	False
/11ProfitCeiling/3/R10000/s041	200	17.206	511610	511584	511610	True
/11ProfitCeiling/4/R10000/s082	500	178.716	1271710	1271664	1271708	True
/11ProfitCeiling/5/R01000/s094	1000	180.029	253114	253020	253114	False
/11ProfitCeiling/6/R01000/s042	2000	180.018	501242	501030	501242	False
/11ProfitCeiling/7/R10000/s007	5000	179.985	12247962	12247500	12247961	True
/11ProfitCeiling/8/R10000/s027	10000	179.844	25002713	25001697	25002712	True
/12Circle/1/R10000/s033	50	0.016	120976	8064659	120976	True
/12Circle/2/R10000/s030	100	180.070	258742	17248594	258742	False
/12Circle/3/R10000/s018	200	186.202	501517	33432774	501517	False
/12Circle/4/R10000/s007	500	180.024	1188095	79202325	1188095	False
/12Circle/5/R10000/s072	1000	180.008	2576715	171772315	2576715	False
/12Circle/6/R10000/s065	2000	180.029	5004504	333616736	5004504	False
/12Circle/7/R01000/s091	5000	179.964	1244805	26229218	1244805	True
/12Circle/8/R01000/s095	10000	179.864	2483752	52334995	2483752	True

Table 1: Test Results

2.2 Evaluation

The following table presents statistics for the selected test cases, as described in the **Experiment set-up** section. The difficulty levels are categorized into three categories: Easy, Moderate, and Most Difficult.

- **Easy Groups (0 False Cases):** Groups 00, 01, 05

  These groups are considered easy as there are no cases where the algorithm failed to find an optimal solution.

- **Moderate Difficulty Groups (Low False Cases):** Groups 04, 06, 10, 11

  These groups exhibit moderate difficulty as the number of cases with non-optimal solutions is relatively low compared to the total number of cases. Specifically, the number of false cases does not exceed 32, equivalent to 40% of the total number of cases.

- **Most Difficult Groups (High False Cases):** Groups 02, 03, 07, 08, 09, 12

  These groups are considered most difficult as the number of cases with failed solutions is relatively high, indicating difficulty in finding optimal solutions.

Group	True	False
00	80	0
01	80	0
02	45	35
03	44	36
04	56	24
05	80	0
06	61	19
07	40	40
08	41	39
09	34	46
10	56	24
11	53	27
12	47	33

Table 2: Optimal Statistics