Set Covering Problem

PIA: Part III

Operations Research, 032
Team 6: Alan Alejandro Vargas González, 2086183

May, 2025

1 Part III: Testing and Optimization

1.1 Test Cases

To evaluate the performance of my solver, I selected instances of varying sizes from the OR-Library:

- **Small Instances**:
  - `scp41.txt` (200 rows, 1000 columns)
  - `scp51.txt` (200 rows, 2000 columns)
- **Medium Instances**:
  - `scpa1.txt` (300 rows, 3000 columns)
  - `scpb1.txt` (300 rows, 3000 columns)

Instance	Subsets	Cost	Time (s)	Optimal?
`scp41.txt`	65	429.0	0.3435	Yes [1]
`scp51.txt`	62	253.0	0.6362	Likely
`scpa1.txt`	67	253.0	1.5724	Likely
`scpb1.txt`	37	69.0	2.7163	Yes [1]

1.2 Adjustments and Optimization

No algorithmic improvements were made. I just created `results.py` to automatically generate a `csv` file with test results.

```
import os
import pandas as pd
from set_covering_problem import read, validation, solving
```

```
files = ["scp41.txt", "scp51.txt", "scpa1.txt", "scpb1.txt"]
directory = "data/pia-3rd/"
results = []

for file in files:
    file_path = os.path.join(directory, file)
    m, n, costs, cov = read(file_path)
    validation(m, n, costs, cov)

    sol, cost, amount_time = solving(m, n, costs, cov)

    results.append({
        "Instance": file,
        "Subsets": len(sol),
        "Total cost": cost,
        "Time": round(amount_time, 4)
    })

df = pd.DataFrame(results)
df.to_csv("results/results_3rd.csv", index = False)
print("Was saved successfully")
```

1.3 Result Analysis

- **Correctness**:
  - All solutions matched known or expected optima
  - All constraints were satisfied in all test cases
- **Efficiency**:
  - Sublinear time growth observed in small and medium instances
  - `scpb1.txt`'s higher runtime (2.7163s) suggests sensitivity to instance sparsity (subsets covering few rows)
- **Robustness**:
  - No failures across selected sets (4, 5, A, B)

1.4 Documentation

Generated files for reproducibility:

- `results.csv`: Instance-level metrics (cost, time, subsets)
- `data/pia-3rd/`: Data folder used

- `results/results_3rd.csv/`: Output location

References

[1] Beasley, J. E. (1987). An algorithm for set covering problems. *European Journal of Operational Research*, 31(1), 85-93.