National Aeronautics and Space Administration

COMPACT
Cubesat Cost Estimation Using K-Nearest Neighbors

Joseph Mrozinski, Michael Saing, Melissa Hooke, Alex Lumnah, James Johnson

Summary

- CubeSat builders need a method to validate their cost estimates when soliciting NASA
- COMPACT tool will estimate the total project costs based on actual historic flown mission data using K-Nearest Neighbors
- A cost estimation tool for users to create quick CubeSat/ SmallSat (≤100 kg) cost estimates with only a few high level known parameters
- Used by those who needs early project cost estimation and/or validation such as: project manager, project systems engineer, cost engineers, cost estimators/analysts, and evaluators/reviewers

Data Summary

- The COMPACT team collects and normalizes technical, cost and programmatic data for CubeSats
- Data comes from interviews with project personnel (PIs, Managers, Systems Engineers)
- Total completed record for the KNN analysis: 25 launched missions

Total # of Mission Records by Form Factors	25
6U	5
3U	13
Others (1, 1.5, 2)	7

Leading Development Organization Type	Tech Demo	Science	Communication
Commercial	2	0	0
University	7	3	0
Civil*	3	3	1
JPL	4	0	2

*Civil = NASA

CubeSat	Mass (kg)	U's	Leading Developer Organization
PharmaSat (1)	5	3	Civil
O/OREOS	5.2	3	Civil
NanoSail-D (2)	4	3	Civil
RAX 1 (USA 218)	3	3	University
PSSC-2	3.7	2	Civil
M-Cubed/COVE (1)	1	1	University
CINEMA (1)	3.15	3	University
CSSWE	3	3	University
Firefly (1)	3.51	3	Civil
PolySat (CP8) "IPEX"	1	1	University
M-Cubed/COVE 2	1	1	University
SkyCube	1.3	1	Commercial
SporeSat-1	5.2	3	Civil
RACE	5	3	University
GRIFEX	4	3	University
LMRST	4.6	3	JPL
EDSN	2	1.5	Civil
CSUNSat-1	2.7	1.5	University
ASTERIA	11	6	JPL
ISARA	5	3	JPL
MarCO	12.7	6	JPL
RainCube	12	6	JPL
Tempest-D	14	6	JPL
NEAScout	12.3	6	Commercial
MiTEE	4	3	University

NASA COMPACT is interested in your project data, and if you would like to contribute, please contact michael.saing@jpl.nasa.gov

Conclusion

COMPACT's Cost Estimations using KNN estimates are a vast improvement over relying on simple mean and median estimates, as evidenced by the MRE curves to the right. Further, the KNN predictions come within +/- 50% of the actual cost on half of test missions.

K-Nearest Neighbors

KNN is a simple form of analogy cost estimation. Here's how it works:

User Inputs
- Mass (kg)
- # of U's
- # of Spacecraft
- Developer type (university, commercial, civil, JPL)

→ Find K "closest" observations in historical data → Calculate weighted average of those K costs →

Algorithm Outputs
- Names of K most similar CubeSats
- Cost estimate for set of inputs

"Closest" here is determined by Euclidean distance between points. Now, the only thing left to do is to choose the number of neighbors, K.

Choosing K

K is chosen by leave-one-out cross (LOOCV) validation:

1. *Set k*
2. *Remove one observation, the "test mission", from the data*
3. *Build KNN model on the remaining observations or "training missions"*
4. *Predict cost for the test mission using weighted average of k neighbors*
5. *Compute MRE statistic for the test mission*
6. *Repeat Steps 2-5 for all n missions*
7. *Compute MRE summary statistics given n errors from Steps 6*
8. *Repeat 1-7 for each new k and choose the k that minimizes the most parameters in Step 7*

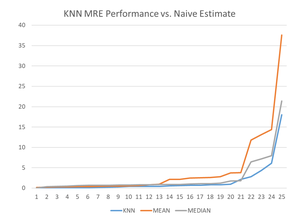
K	25th	50th	75th	Mean
1	0.29	0.63	1.05	1.87
2	0.16	0.45	0.89	1.66
3	0.22	0.42	0.89	1.75
4	0.27	0.60	0.96	1.85
5	0.21	0.57	0.96	1.96

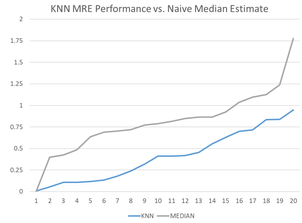
MRE Summary for Different K Values

LOOCV is preferred over k-fold cross validation in order to maximize the number of observations that are used in the training algorithm since there are only 25 data points. This gives us a better idea of how the model will perform in practice (minimizes variance), but gives an increased risk of overfitting (increases bias).

Evaluation

Leave-one-out cross validation is used to evaluate the model. The MRE for each mission is calculated based on inputting each mission and outputting a KNN prediction after removing that CubeSat from the training data. The results are compared to MREs from using the mean and the median of the training data which is considered to be a naïve estimate good for model assessment.





Next Steps

- Continue data collection and normalization
- Integrate K-Nearest Neighbors methodology into software tool
- Create Parametric Cost Models using Principle Component Analysis
- Release the COMPACT tool via NASA's www.oncedata.com

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California,
URS #282538

www.nasa.gov

ACKNOWLEDGEMENTS: We appreciate all the projects that contributed their project data to NASA's COMPACT tool.