Control experiment: correctness verification and sanity checks

The algorithms that we propose in the paper to calculate the different features have several tricky implementation details. We already formally proved that these algorithms are correct, but this does not rule out potential implementation bugs. Apart from carefully implementing these algorithms, we also systematically verified that we did not introduce any bugs by comparing the output of our algorithms with the output of several brute force algorithms that we also implemented. These brute force algorithms are very slow, but very easy to implement and are useful tools to experimentally verify the correctness of faster, but harder algorithms. More specifically, we also implemented a brute force algorithm that explicitly generates the set of all IMSs X by considering all 2^n solutions and filtering out the IMSs. The features were then calculated by directly using the definition and compared with the features that were calculated by the dynamic programming algorithm from the paper. The brute force algorithm also allowed us to do a sanity check to verify that the bound that we obtained in Theorem 2 is correct, by checking the inequality for every IMS in X. To do a sanity check for Theorem 3, we implemented another brute force algorithm that generates every partition of the set of items $\{1, 2, \ldots, n\}$ and checks the inequality for every IMS. A slight modification of this algorithm was also used to calculate $h(g)$ and g^* (the optimal centroids that minimize $h(g)$ are given by the average of the weights of each group induced by the partition). The output of all brute force algorithms was compared with the output of the algorithms proposed in the paper for 10^4 small instances that were generated by choosing the number of items n uniformly at random between 5 and 12, choosing the knapsack capacity c uniformly at random between 2 and 10^8 and choosing the profits and weights of each item uniformly at random between 1 and c. Through this control experiment we were able to experimentally verify that the results obtained by our algorithms matched the expected results, as desired.