Phase Order Search with Feature Selection
João Martins & David Henriques

Major Changes

The main change we foresee in our project is with our benchmark suit. On one hand, some cases in the SPEC200 suite take quite a while to run and we may not have time to run them all for all diferent optimization orderings. On the other hand, we found several compression algorithms that seem interesting to analyse and compile with LLVM, so we may just end up using those.

Acomplishments

We have managed to integrate a fair number of our case studies with LLVM (which is harder than it sounds). It is clear to us that this is a bottleneck and that once this is done, the remaining of the project should go smoothly. We've settled in PCA as feature selection technique since it explicitly tells us which features are most important. We've discussed about what to do to restrict the Phase Order Space and came up with tho proposals

- we partition the optimizations into classes under some criterium (e.g. effect, like motion, rewriting or elimination of code) and then we try all orderings of those classes, which, unlike all ordering of optimizations, are a manageable number.
- we randomly pick a couple of optimizations and, using only those optimizations, fix an ordering on that pair by trying out some benchmarks. Each time we do this we cut the phase order space in half. We do this until the space is of manageable size.

Meeting the Milestone

We have not yet achieved our milestone since integration with LLVM is taking way longer than expected. However, having laid down the workflow, we now believe the second half of the project will be faster than initially predicted.

Surprises

Besides having a hard time running our benchmarks with the passes we've chosen, we were very surprised to find that clang's higher optimization levels (O2 and O3) are often slower than no optimization at all (O0). If a similar phenomenon occurs with the optimization passes we run, then the end results may end up being not very convincing. It is however too soon to speculate,a s we only have tried this in a couple of benchmarks.

Revised Schedule

Once we have all our benchmarks running with whatever passes we want them to (which should happen by the end of the week), we will spend one week gathering training data and teaching the classifier. The last week will be devoted to testing the classifier and writing up.

Resources Needed

We have secured both SPEC2006 and several other benchmarks. Unfortunately we may need to throw away some of the slower benchmarks, since we are using our own laptops, which are not very fast.