Use of D-Wave's Quatum Computer for Global Optimization

Tomas Ukkonen, Novel Insight, 2020

D-Wave Hybrid Quantum Optimizer can solve problems using one million variables (9/2020). Their quantum computer can optimize Ising model (quadratic optimization) and do simulated annealing. The one million variables $N=10^6$ mean that using normal linear optimization the memory use would scale as $O(N^2)$. This means that the computer must have 4096 gigabytes of memory which is outside of the scope of normal computers (supercomputers, however, could be used to solve for larger models).

The Ising model originally for simple magnetic materials is:

$$H(\boldsymbol{\sigma}) = -\sum J_{ij}\,\sigma_i\,\sigma_j - \mu\sum h_j\,\sigma_j,\; P_\beta(\boldsymbol{\sigma}) = \frac{e^{-\beta H(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} e^{-\beta H(\boldsymbol{\sigma})}}$$

The optimization of $H(\boldsymbol{\sigma})$ can be mapped to simple optimization problem which maximizes, for example, expected value $E[Y]$ given variables $\boldsymbol{\sigma}$. We can choose $J_{ij} = -\frac{1}{4}E[Y\,|\sigma_i,\sigma_j]$, $\mu=0$ and map σ_i $-1/+1$ variables to be binary indicator $0/1$ variables. $\sigma_i' = (\sigma_i+1)/2$, $\sigma_i = 2\,\sigma_i' - 1$. It is then possible to write by simplifying that $E[Y\,|\boldsymbol{\sigma}]$ depends on only two $\boldsymbol{\sigma}$ variables at the time and that they are independent (it is possible generalize model to more variable combinations by introducting meta variables $\sigma_k = (\sigma_i\,\sigma_j)_k$):

$$E[Y] = \sum_{\boldsymbol{\sigma}} E[Y\,|\sigma_i'\,\sigma_j']\,P(\sigma_i')\,P(\sigma_j')$$

This problem is further simplified so that σ_i' are indicator variables telling probability if some decision is made or not ($P(\sigma_i = p) = p$). For high dimensional data the most of the probability mass in a hypercube $[0,1]^D$ are near edges meaning that one can make rough approximation and only use values $p=0$ and $p=1$. This means that $P(\sigma_i=0)=0$ and $P(\sigma_i=1)=1$ and we have

$$E[Y] = \sum E[Y\,|\sigma_i'\,\sigma_j']\,\sigma_i'\,\sigma_j'$$

$$E[Y] = -\cdot\sum J_{ij}\,(\sigma_i+1)(\sigma_j+1)$$

$$E[Y] = -\cdot\sum J_{ij}\,\sigma_i\,\sigma_j - \sum\left(\sum_j J_{kj} + \sum_i J_{ik}\right)_k \sigma_k - \sum_{i,j} J_{ij}$$

$$E[Y] = E\{H(\boldsymbol{\sigma})\}$$

$$E[Y\,|\boldsymbol{\sigma}] = H(\boldsymbol{\sigma})$$

We have Ising model in which $\mu=1$ and $h_k = \sum_j J_{kj} + \sum_i J_{ik}$ and an additional constant term can be ignored when optimizing or sampling from the Ising model.

In practice, the large scale Ising model allows us to optimize for larger number of parameters than using normal computers. It is then possible to try to do global optimization of problems such as how municipalities operate ($\boldsymbol{\sigma}$ is joint all actions taken by different municipalities each making M same decisions) or *approximately* optimize for total GDP when $\boldsymbol{\sigma}$ variables are all decisions made by different countries (each country makes M same deicisions). It is then possible to run the optimization also separately for each municipality or country and compare results which can be used to look for possible co-operations which can be shown to maybe finally improve each participant's interest by optimization operations as a whole.

When using D-Wave quantum computers we can look for optimimum from $N = 2^{10^6} = 10^{301029}$ possible binary combinations which cannot be easily done using normal computers. Also instead of expetected value it is possible to optimize for uncertainly based the worst case problems like $J_{ij} \sim \max\,(0, E[Y] - \alpha\,\mathrm{StDev}[Y])$ which are maybe more realistic. (*FIXME: Is it possible to have freely connected Ising model or does one have to reduce number of connections to fit to reduced 2d lattice?*)

Application to Municipalities Data (TODO)

Because the quantum computer can solve for one million variables and because there are 310 municipalities in Finland it is possible to have 3225 variables per municipality. However, these are binary variables and in practice continuous variables must be discretized to binary variables using one-hot encoding by using 10 binary variables meaning there are 322 variables available per municipality.

Basically this is quite simple but requires lots of preprocessing by doing ETL, preprocessing the variables and scaling variables to have reasonable range.

Application to Country-Wise GDP Data (TODO)

There are 195 countries in the World. This means there can be 512 continuous variables per country. The optimization target is to maximize total GDP and compare results if each country maximizes GDP independently (by maximizing only own good).

Basically this is quite simple but requires lots of preprocessing by doing ETL, preprocessing the variables and scaling variables to have reasonable range.