KNN algorithm in machine learning

February 9, 2017

# Overview

## KNN Algorithm Description

K-nearest neighbors (KNN) algorithm is one of the simplest machine learning algorithm used in classification and regression. In classification, KNN algorithm classify objects by a majority vote of k nearest reference object. The input to the algorithm is k closest training examples in the feature space. Output is different for classification and regression. Class membership is the output for KNN classification, and property value for the object is the output for KNN regression. Distance calculation and distance rank are the two most calculation intensive process in KNN. KNN algorithm calculate the distance between reference objects and query objects, which result in N distances calculations. Distance rank then sorts each row of the matrix and choose the k smallest distance for each row.

The KNN algorithm can also assign weight to the contribution of the neighbors, so closer member contribute more to the average than the more distant ones. Common weighting scheme would weight neighbors by **1/d**, where d is distance between neighbor and reference object.

Despite the simplicity of KNN algorithm, it provide very competitive results. It can be applied in many different problems, such as robot control, data mining, and prediction. KNN can also be used as a base classification method before tweaking or changing to a more complex classification technique.

kNN remarks

* K must not be a multiple of the number of classes. (ex. K != 4, when class = 2, to avoid confusion of class)

## High-Level Requirements

* Implement KNN algorithm in opencl kernel, and achieved an improvement compared to sequential
* Compare with existing KNN algorithm over results on datasets

## Deliverables

* KNN Algorithm Kernel
* Details comparison of existing KNN algorithm kernel
* Testing result against large dataset (possibly a real life example)

## Implementation Plan / Deliverable

* Provide summary of related research by 02/16/17
* Implement KNN algorithm in kernel 03/02/17
* Optimization (VTune) 03/09/17
* Report 03/16/17

## Reference

1. <https://www.youtube.com/watch?v=09mb78oiPkA> (MIT OpenCourseWare)
2. <http://vlm1.uta.edu/~athitsos/nearest_neighbors/>
3. An Efficient FPGA Implementation for odd-even sort based KNN algorithm using OpenCL, Peng, Huang
4. Mult-GPU Implementation of Machine Learning Algorithm using CUDA and OpenCL (Masek, Burget)