

Verslag kNN - Deel 1

Leon van der Veen (3009637) Tom de Ruijter (?)
Andrew Wolters (?) Hans Harmannij (?)

22 februari 2011

Inleiding Bla bla bla.

Condensing Bla bla bla.

Editing Bla bla bla.

Implementatie Voor de implementatie van dit project is gekozen voor de programmeer taal C++. Deze beslissing is voortgekomen uit een noodzaak van een taal die hoge performance heeft. Ook een groot voordeel is dat C++ een goed geschikte taal is voor object georiënteerd programmeren. Dit maakt een modulaire structuur in de code een stuk eenvoudiger.

Structuur en dataset We hebben voor de structuur van het (C++) project gekozen om de Header (.hpp) betanden en de Source (.cpp) bestanden te scheiden. Deze zijn ook in het filesysteem verdeeld over verschillende mappen. Dit heeft misschien in specifieke project niet extreem veel zin, maar over het algemeen bij een software project heeft een groot voordeel om deze twee zaken te scheiden. Zo kan het project makkelijk omgezet worden in een zogenaamde 'library' waardoor code weer makkelijk hergebruikt zou kunnen worden in andere projecten. Dit alles is weer in de naam van 'modulair' programmeren.

Verder is de structuur in gedeeld in verschillende mappen. Zo krijgt de basis van de kNN (het kNN algoritme zelf, de Dataset, etc.) een eigen map waar de rest van het project op verder bouwt. Ook is er een speciale dataset (FileDataset) die zorgt voor het laden en opslaan van Datasets in bestanden. Deze is weer in een aparte map geplaatst: datasets. Ook de verschillende 'optimizers' (de condensing en editing algoritmen) zijn weer in een eigen map 'optimizers' geplaatst. Dit maakt specifieke code vinden een stuk makkelijker en schijnt de functionliteit op een overzichtelijke manier.

In de dataset hebben wij voor bepaalde termen gekozen. Zo heeft elke 'instance', van features en class label, de naam 'Stimulus' gekregen. De dataset is dan ook niets anders als een lijst van Stimuli met wat extra code om deze lijst (efficient) op te bouwen in het geheugen. Er is geprobeerd de dataset zo effiecent mogelijk te houden wat zeer belangrijk is met grote aantallen Stimuli. We hopen met deze opbouw en keuzen het uiteindelijke resultaat zo snel mogelijk te houden (en makkelijk omzetbaar voor een multithreaded variant).

Conclusie Bla bla bla.

Evaluatie

Leon In het begin kwam ik lastig op gang met wekelijks (voor)onderzoek doen over algoritmen. Mijn bijdragen aan deel 1 is dan ook vooral gericht op de Development kant. Ik heb dan ook de structuur van het project opgezet, en de basis gelegd waarop de rest verder kan implemmenteren. Qua 'research' zou ik dus meer moeten doen (in deel 2). Qua overleg en teamwork ging het prima. Er is genoeg overleg geweest met elkaar en problemen werden onderling goed overlegd en opgelost (zo nodig).

Hans Bla bla bla.

Andrew Bla bla bla.

Tom Bla bla bla.