Metody analýzy dat
Implementační část

Ondřej Řeháček (REH0063)

December 13, 2016

1 Algoritmy

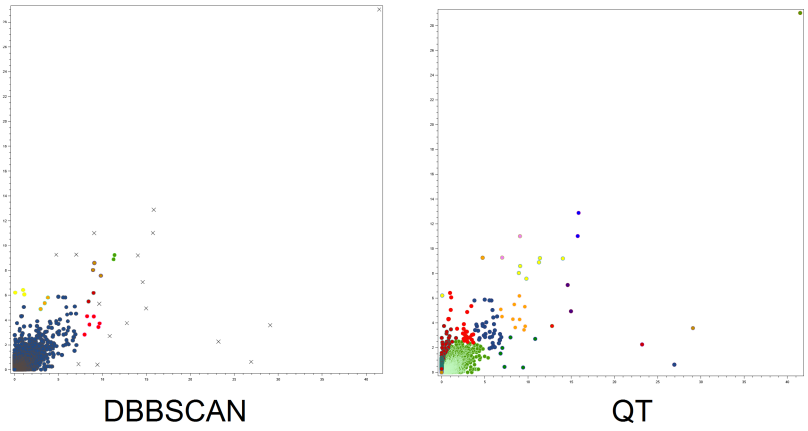
Shlukování je provedeno na všechny videohry v datasetu a jejich atributy EUSales X NASales.

1.1 DBSCAN (Density Based

Algoritmus poměrně hezky shlukoval videohry, které se průměrně prodávají a jsou tak svými prodeji sobě velmi podobné. Problém algoritmu však nastal v případě, když došlo na klasifikací titulů, které jsou svými prodeji "vyjímečné" a od sebe výrazně vzdálené.Tyto body algoritmus opakovaně vyhodnocoval jako šum (noise) a nerozdělil je do shluků, naopak je vyřadil (Křížek v grafu).

1.2 QT (Quality Threshold)

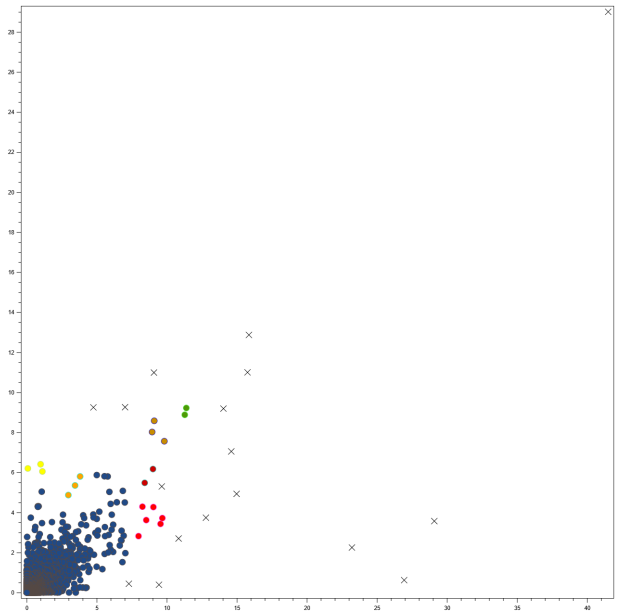
Implementoval jsem pouze na zkoušku - QT tento problém vyřešil u mého datasetu. Generuje větší počet clusterů, ale dovede zařadit i "jedinečné" videohry do samostatných shluků.



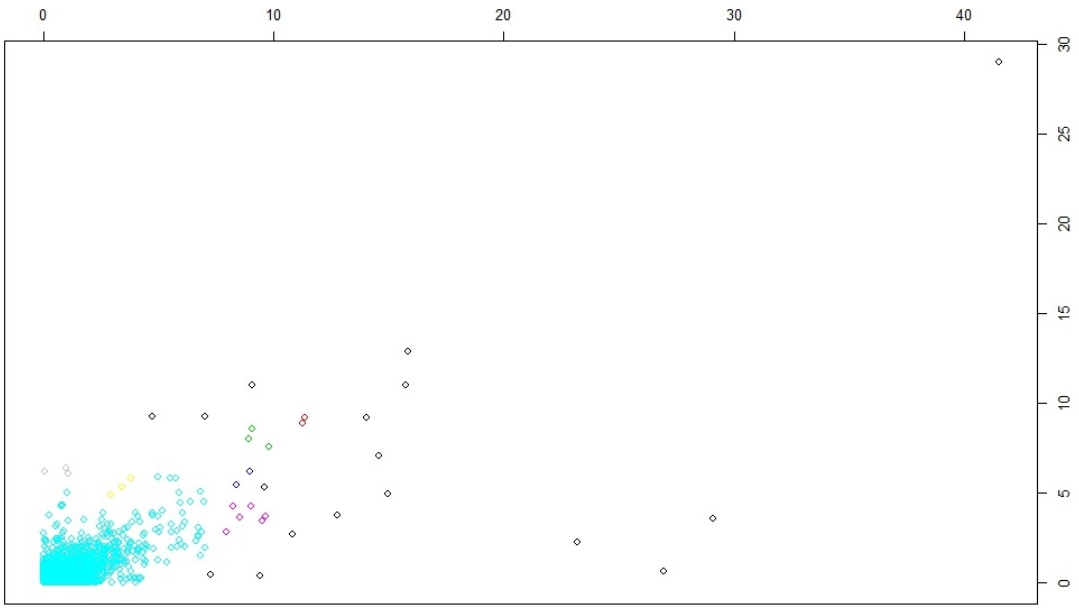
2 Ověření v R

Ověření bylo provedeno pomocí R na algoritmu DBSCAN. Rozdělení do clusteru a vyhodnocení noise bodů je identické.

2.1 Výstup implementace (Noise - Křížek)



2.2 Výstup R (Noise - černý bod)



3 Spuštění a výstupy

Spuštění je podmíněno NuGet balíkem OxyPlot.PDF.

Ve složce **Output** je několik výstupu programu:

- **output.txt** - Informace o datasetu.
- **dbClustering.pdf** - Graf reprezentující DBSCAN shlukování.
- **qtClustering.pdf** - Graf reprezentující QT shlukování.
- **dbClusters.txt** - Rozpis jednotlivých her v clusterech vytvořených DBSCAN algoritmem.
- **qtClusters.txt** - Rozpis jednotlivých her v clusterech vytvořených QT algoritmem.
- **xxxOccurences.csv** - Výskyty hodnot, relativni a kumulativni četnost hodnot, PDF, CDF.

4 Zdroje

- https://www.kaggle.com/gregorut/videogamesales
- https://en.wikipedia.org/wiki/DBSCAN
- https://github.com/mhahsler/dbscan
- https://www.yzuzun.com/2015/07/dbscan-clustering-algorithm-and-c-implementation/
- http://stackoverflow.com/questions/6621630/dbscan-code-in-c-sharp-or-vb-net-for-cluster-analysis
- http://www.c-sharpcorner.com/uploadfile/b942f9/implementing-the-qt-algorithm-using-C-Sharp/
- https://github.com/dennyferra/QTCluster
- https://www.kaggle.com/gregorut/videogamesales
- https://sites.google.com/site/dataclusteringalgorithms/quality-threshold-clustering-algorithm-1