

## Subscale Algorythmus

William Mendat,<sup>1</sup> Max Ernst,<sup>2</sup> Steven Schall,<sup>3</sup> Matthias Reichenbach<sup>4</sup>

**Abstract:** Das könnte das abstract sein

**Keywords:** C++; Subscale; Cluster

### 1 Überschrift/Heading

Hallo [?]

### 2 Überschrift/Heading

Hallo [?]

### 3 Berechnung der Dense Units

Dense Units werden in verschiedene Funktionen mit unterschiedlichen Zwecken berechnet. Das Hauptziel dieser Berechnung, ist die Bestimmung einer Teilmenge aus allen möglichen Kombinationen in einem subset. Dabei dürfen keine Wiederholungen der Dense Units auftreten. Die Kombinationen aus einem subset können mittels dem Binomialkoeffizient berechnen,  $\binom{n}{k}$  dabei ist n die Anzahl der Elemente in dem subset und k die minimale Anzahl an punkten in einem subset (e.g. minPoints). Daraus entstehen k-Elemente großes subsets von einem n-Elemente set ohne Wiederholungen der Kombinationen. Wenn zum Beispiel ein Core Set aus folgenden Punkten besteht: [1, 5, 7, 9, 22] und die minimale Anzahl an punkten in einem Subset Drei ist, sind die ersten Drei Dense Units folgende: [1, 5, 7] und [1, 5, 9] und [1, 5, 22]. Die Formel zur Berechnung der Dense Units kann also folgendermaßen betrachtet werden:  $\binom{|CS|}{minPoints}$

---

<sup>1</sup> Hochschule Offenburg, Offenburg, Deutschland w mendat@stud-hs.offenburg.de

<sup>2</sup> Hochschule Offenburg, Offenburg, Deutschland w mendat@stud-hs.offenburg.de

<sup>3</sup> Hochschule Offenburg, Offenburg, Deutschland s schall@stud-hs.offenburg.de

<sup>4</sup> Hochschule Offenburg, Offenburg, Deutschland m reichen@stud-hs.offenburg.de

#### **4 Kollision von Dense Units**

#### **5 Abbildung Dense-Units auf Subspaces**

Für das endgültige Clustering muss die Ausgabe des SUBSCALE-Algorithmus in die Struktur von den Clusterkandidaten abgebildet werden. Jeder Kandidat besteht aus Dimensionen und eindeutigen Punkt IDs. Dabei werden Punkte ausgesucht, welche in mehreren Dense-Units vertreten sind. Diese werden dann zusammen innerhalb eines Subspaces definiert, wobei mehrfach vorkommende Punkte nur einmal eingetragen werden. Die hier zusammengeführten Punkte haben die Eigenschaft, dass sie ziemlich wahrscheinlich auch in dem jeweiligen Unterraum geclustert sind. Diese Eigenschaft macht diese Punkte zu günstigen Kandidaten für das endgültige Clustering, welches im folgenden Unterkapitel ?? näher beschrieben wird.

#### **6 Abschließendes Clustering mit DBSCAN**

Nach dem alle maximalen Subspaces identifiziert worden sind, wird zuletzt das Clustering durchgeführt, welches die maximalen Supspace Cluster finden soll. Zur Bewältigung dieser Aufgabe wird ein volldimensionaler Clustering-Algorithmus verwendet. Der Algorithmus, welcher in der von uns verwendeten Implementierung verwendet wird, nennt sich DBSCAN (Density-Based Spatial Clustering of Applications with Noise). DBSCAN ist ein auf die Dichteverbundenheit basierender Clustering-Algorithmus, der Cluster mit beliebiger Form findet und Rauschpunkte separat zurückliefert.