Algorithmische Methoden der Netzwerkanalyse

Marco Gaertler

03. Februar, 2009

Sinnvolle Algorithmen?

Frage

Nach welchen Kriterien könnten Algorithmen vorgehen, um sinnvolle Clusterungen zu identifizieren, wenn das Berechnen einer optimalen Struktur bezüglich der meisten Qualitätsindizes schwer ist?

Sinnvolle Algorithmen?

Frage

Nach welchen Kriterien könnten Algorithmen vorgehen, um sinnvolle Clusterungen zu identifizieren, wenn das Berechnen einer optimalen Struktur bezüglich der meisten Qualitätsindizes schwer ist?

Wie könnte ein *greedy* Clusteralgorithmus vorgehen?

Greedy Merge: Linkage

Definition

Sei $G = (V, E, \omega)$ ein Graph und $\mathcal{C}_1$ eine beliebige Clusterung von G. Weiter seien entweder eine globale Kostenfunktion $c_{\text{global}} \colon \mathcal{A}(G) \to \mathbb{R}_0^+$ oder eine lokale Kostenfunktion $c_{\text{local}} \colon 2^V \times 2^V \to \mathbb{R}_0^+$ gegeben.
Das *allgemeine Linkage* Verfahren berechnet eine Folge von Clusterungen $\mathcal{C}_2, \ldots, \mathcal{C}_\ell$, wobei $\mathcal{C}_{i+1}$ aus $\mathcal{C}_i = \{C_1, \ldots, C_k\}$ wie folgt entsteht:

Greedy Merge: Linkage (Fortsetzung)

Definition (Fortsetzung)

global: $\mathcal{C}_{i+1} := \operatorname*{argmin}_{\mathcal{C}' \in M} c_{\text{global}}(\mathcal{C}')$ mit

$$M := \left\{ \{C_\mu \cup C_\nu\} \cup \left(\mathcal{C}_i \setminus \{C_\mu, C_\nu\} \right) \middle| \mu \neq \nu \right\}$$

local:

$$\mathcal{C}_{i+1} := \operatorname*{argmin}_{\{C_\mu \cup C_\nu\} \cup (\mathcal{C}_i \setminus \{C_\mu, C_\nu\})} c_{\text{local}}(C_\mu, C_\nu)$$

Kostenfunktionen

Fragen:

- Was könnten globale/lokale Kosten konkret messen?
- Was unterscheidet die globalen Kosten von den lokalen?
- Welche Vor-/Nachteile haben globale/lokale Kosten?

Existenz von Sinnvolle Algorithmen?

Frage

Kann es eigentlich *sinnvolle* Algorithmen zur Clusterungen geben?

Wie kann man *sinnvoll* formalisieren?

Axiomatisierung?

Definition

Ein Clusterungsalgorithms ist eine Funktion $f\colon \mathbb{R}_0^{n\times n} \rightarrow \mathrm{Part}(n)$.

Anschaulich: Ein Algorithmus berechnet aus den paarweisen Abständen/Ähnlichkeiten von n Elementen eine Partition.

Axiomatisierung?

Definition

Ein Clusterungsalgorithms ist eine Funktion $f\colon \mathbb{R}_0^{n \times n} \to \mathrm{Part}(n)$.

Anschaulich: Ein Algorithmus berechnet aus den paarweisen Abständen/Ähnlichkeiten von n Elementen eine Partition.

Frage

Welche Eigenschaften sollte f nun erfüllen?

Axiomatisierung?

Axiom: Scale-Invariance

$\forall\, D \in \mathbb{R}_0^{n \times n}, \forall\, \alpha \in \mathbb{R}^+ \colon f(D) = f(\alpha D)$

Axiom: Richness

f ist surjektiv.

Axiomatisierung?

Axiom: Scale-Invariance

$\forall\, D \in \mathbb{R}_0^{n\times n}, \forall\, \alpha \in \mathbb{R}^+ \colon f(D) = f(\alpha D)$

Axiom: Richness

f ist surjektiv.

Für $D \in \mathbb{R}_0^{n\times n}$ definiere den f-Support durch die Menge

$$\left\{ D' \in \mathbb{R}_0^{n\times n} \;\middle|\; \forall\, 1 \leq i,j \leq n \colon \begin{cases} D'_{ij} \leq D_{ij}, \text{ falls } i \sim_{f(D)} j \\ D'_{ij} \geq D_{ij}, \text{ falls } i \not\sim_{f(D)} j \end{cases} \right\}$$

Axiom: Consistency

$\forall\, D, D' \in \mathbb{R}_0^{n\times n} \colon \Big(D' \in f\text{-}\mathrm{Support}(D) \Longrightarrow f(D) = f(D') \Big)$