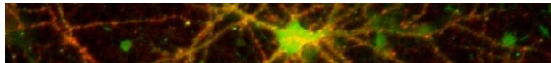




Professur
Künstliche Intelligenz
Prof. Dr. Fred Hamker



TITEL

Frederik Beuth, Fred. H. Hamker



Professur
Künstliche Intelligenz
Prof. Dr. Fred Hamker

Gliederung



Einführung Clusteranalyse

Aufgabe: Erkennung von Gruppen oder Mustern in beliebigen Daten (Mustererkennung)

Beispiel: Durchführung einer Umfrage über den TV-Konsum
Erkennung von Konsumentengruppen

Personennummer	1	2	3	4	5
Sport	3	2	8	9	8
Entertainment	1	2	1	0	8

in h pro Woche



Einführung Clusteranalyse

Aufgabe: Erkennung von Gruppen oder Mustern in beliebigen Daten (Mustererkennung)

Beispiel: Durchführung einer Umfrage über den TV-Konsum
Erkennung von Konsumentengruppen

Personennummer	1	2	3	4	5
Sport	3	2	8	9	8
Entertainment	1	2	1	0	8

in h pro Woche



Warum Clusteranalyse?

- Analyseverfahren zur Mustererkennung
- Breites Anwendungsspektrum für Clusteranalyseverfahren,
z. B. Bilderkennung, Datamining, Information-Retrieval,
...
- Existenz sehr vieler Algorithmen, aber kein perfekter
- Bedarf nach neuen Algorithmen gegeben



Warum Clusteranalyse?

- Analyseverfahren zur Mustererkennung
- Breites Anwendungsspektrum für Clusteranalyseverfahren,
z. B. Bilderkennung, Datamining, Information-Retrieval,
...
- Existenz sehr vieler Algorithmen, aber kein perfekter
- Bedarf nach neuen Algorithmen gegeben



Warum Clusteranalyse?

- Analyseverfahren zur Mustererkennung
- Breites Anwendungsspektrum für Clusteranalyseverfahren,
z. B. Bilderkennung, Datamining, Information-Retrieval,
...
- Existenz sehr vieler Algorithmen, aber kein perfekter
- Bedarf nach neuen Algorithmen gegeben



Warum Default-Artmap?

- Default Artmap: **Adaptive Resonanztheorie** mit überwachtes Clustering (map), Standardverfahren
- Schnelligkeit, gute Resultate, Transparenz, Adaptive Clusteranzahl
- Entwickelt von Gail A. Carpenter an der University of Boston, 2003
- Neuheit (2003) verursacht ein Problem:
Wenig Kenntnisse bezüglich praktischen Einsatzes und Nachteilen des Algorithmus



Warum Default-Artmap?

- Default Artmap: **Adaptive Resonanztheorie** mit überwachtes Clustering (map), Standardverfahren
- Schnelligkeit, gute Resultate, Transparenz, Adaptive Clusteranzahl
- Entwickelt von Gail A. Carpenter an der University of Boston, 2003
- Neuheit (2003) verursacht ein Problem:
Wenig Kenntnisse bezüglich praktischen Einsatzes und Nachteilen des Algorithmus



Warum Default-Artmap?

- Default Artmap: **Adaptive Resonanztheorie** mit überwachtes Clustering (map), Standardverfahren
- Schnelligkeit, gute Resultate, Transparenz, Adaptive Clusteranzahl
- Entwickelt von Gail A. Carpenter an der University of Boston, 2003
- Neuheit (2003) verursacht ein Problem:
Wenig Kenntnisse bezüglich praktischen Einsatzes und Nachteilen des Algorithmus



Ziele der Diplomarbeit

- Aufbereitung des Algorithmus für künftige Forschung
- Evaluatblendation des Default Artmap-Verfahrens
- Nachteile herausarbeiten



Ziele der Diplomarbeit

- Aufbereitung des Algorithmus für künftige Forschung
- Evaluatblendeton des Default Artmap-Verfahrens
- Nachteile herausarbeiten



Ziele der Diplomarbeit

- Aufbereitung des Algorithmus für künftige Forschung
- Evaluationsdetektion des Default Artmap-Verfahrens
- Nachteile herausarbeiten



Beispiel Clusteranalyse

Aufgabe Erkennung von Konsumentengruppen

Datensatz

Patterns Nr.	1	2	3	4	5
Sport in h	3	8	8	2	9
Entert. in h	1	1	8	2	0
Merkmal a_{sport}	0.3	0.8	0.8	0.2	0.9
Merkmal a_{entert}	0.1	0.1	0.8	0.2	0.0
Zielklasse K	0	1	2	-	-
Datenmengen	Trainingsmenge			Testmenge	



Beispiel Clusteranalyse

Aufgabe Erkennung von Konsumentengruppen

Datensatz

Patterns Nr.	1	2	3	4	5
Sport in h	3	8	8	2	9
Entert. in h	1	1	8	2	0
Merkmal a_{sport}	0.3	0.8	0.8	0.2	0.9
Merkmal a_{entert}	0.1	0.1	0.8	0.2	0.0
Zielklasse K	0	1	2	-	-
Datenmengen	Trainingsmenge			Testmenge	



Definitionen für die Clusteranalyse

Aufgabe Erkennung von Gruppen oder Mustern in beliebigen Daten

Rohdaten Datenquelle aus der realen Welt

Merkmale M Vorverarbeiteten Rohdaten, erzeugen einen M -dim. Merkmalsraum

Datensatz Zusammenfassung von Merkmalen zu Patterns, Patterns zu Datensätzen

Cluster C Gruppe von Patterns mit ähnlichen Merkmalen

Klasse L Clustername, mehrere Cluster pro Klasse möglich



Definitionen für die Clusteranalyse

Aufgabe Erkennung von Gruppen oder Mustern in beliebigen Daten

Rohdaten Datenquelle aus der realen Welt

Merkmale M Vorverarbeiteten Rohdaten, erzeugen einen M -dim. Merkmalsraum

Datensatz Zusammenfassung von Merkmalen zu Patterns, Patterns zu Datensätzen

Cluster C Gruppe von Patterns mit ähnlichen Merkmalen

Klasse L Clustername, mehrere Cluster pro Klasse möglich



Professur
Künstliche Intelligenz
Prof. Dr. Fred Hamker

Gliederung



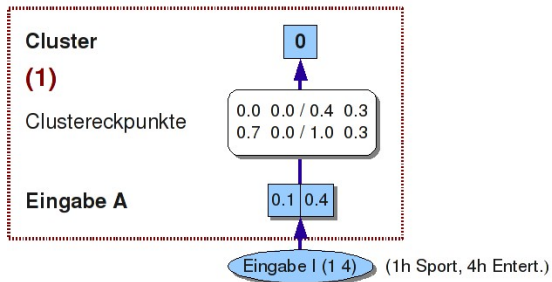
Trainingsalgorithmus

- Initialisierung: Erstelle
aus dem ersten Pattern
den ersten Cluster
- Füge sequentiell Patterns
dem jeweils ähnlichsten
Cluster hinzu
Pattern einem Cluster
hinzufügen



Trainingsalgorithmus

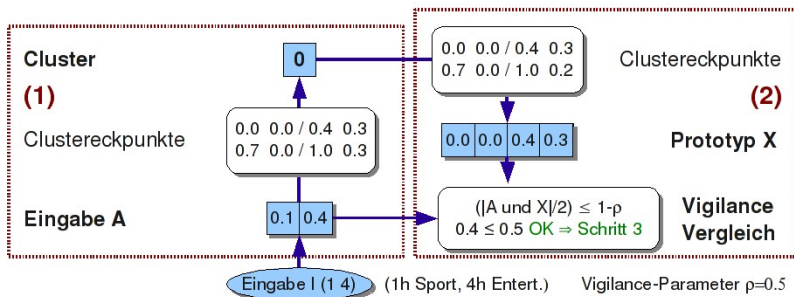
(1) Clusterauswahl





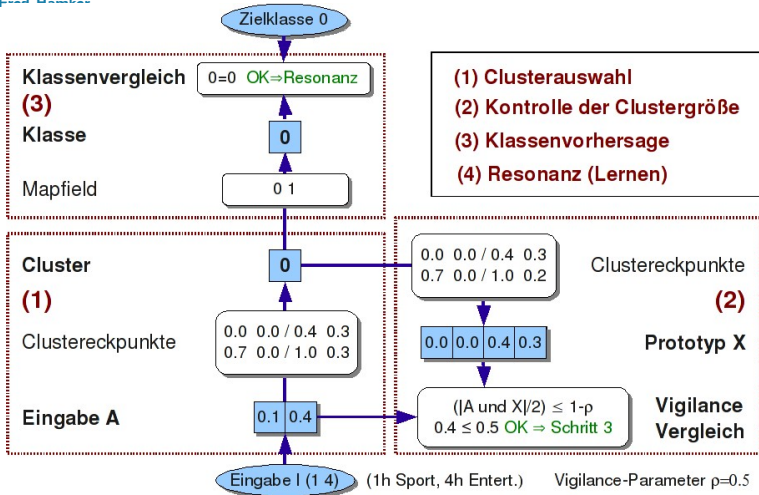
Trainingsalgorithmus

- (1) Clusterauswahl
- (2) Kontrolle der Clustergröße





Trainingsalgorithmus





Professur
Künstliche Intelligenz
Prof. Dr. Fred Hamker

Gliederung



Datensätze

- Generischer **2d3c** - Verwendung für Clustervisualisierung
- **Iris** Schwertlilien
- **Glass** Chemische Elemente von Glas

Trainiert mit

- Default Artmap
- Künstliches Neuronales Netz (KNN) / Backpropagation Learning

Resultate

- Klassifizierungsrate $K = \frac{\text{Korrekte Patterns}}{\text{Gesamtanzahl}}$
- Struktur des Netzes $\#C$ (Anzahl Cluster bzw. Hidden-Neuronen)



Datensätze

- Generischer **2d3c** - Verwendung für Clustervisualisierung
- **Iris** Schwertlilien
- **Glass** Chemische Elemente von Glas

Trainiert mit

- Default Artmap
- Künstliches Neuronales Netz (KNN) / Backpropagation Learning

Resultate

- Klassifizierungsrate $K = \frac{\text{Korrekte Patterns}}{\text{Gesamtanzahl}}$
- Struktur des Netzes $\#C$ (Anzahl Cluster bzw. Hidden-Neuronen)



Datensätze

- Generischer **2d3c** - Verwendung für Clustervisualisierung
- **Iris** Schwertlilien
- **Glass** Chemische Elemente von Glas

Trainiert mit

- Default Artmap
- Künstliches Neuronales Netz (KNN) / Backpropagation Learning

Resultate

- Klassifizierungsrate $K = \frac{\text{Korrekte Patterns}}{\text{Gesamtanzahl}}$
- Struktur des Netzes $\#C$ (Anzahl Cluster bzw. Hidden-Neuronen)



Datensätze

- Generischer **2d3c** - Verwendung für Clustervisualisierung
- **Iris** Schwertlilien
- **Glass** Chemische Elemente von Glas

Trainiert mit

- Default Artmap
- Künstliches Neuronales Netz (KNN) / Backpropagation Learning

Resultate

- Klassifizierungsrate $K = \frac{\text{Korrekte Patterns}}{\text{Gesamtanzahl}}$
- Struktur des Netzes $\#C$ (Anzahl Cluster bzw. Hidden-Neuronen)



Professur
Künstliche Intelligenz
Prof. Dr. Fred Hamker

Gliederung



- Geringer Nutzen der Parameteroptimierung
- Klassifizierungsrate K ist sehr stark von der Reihenfolge der Patterns im Training abhängig (bis zu 20%)

Lösung: Eine oder mehrere randomisierte Trainingsmengen, aber dadurch Black-Box Verfahren

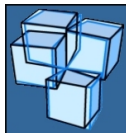


- + Bessere Laufzeit als ein KNN
- + Adaptive Clusteranzahl
- + Kategorisierungsraten sind gut und vergleichbar mit denen eines KNN



Zusammenfassung

- Evaluation: Verfahren ist praktisch einsetzbar
- Forschung: Beheben des Problems der Eingabedatenabhängigkeit



Danksagung: Diese Arbeit wurde im Rahmen des europäischen Projektes “Self Constructing Hyper Wavelet Algorithms For Extrapolating Linguistics (SCHWAFEL)” gefördert.