

# Neuronale Netzwerke und Clusteringverfahren für die Analyse von Geodaten


Lehrstuhl für Geoinformatik

Robin Bially

[HTTPS://GITHUB.COM/ROBINBIA/PROJEKTARBEIT-GEOINFORMATIK.GIT](https://github.com/ROBINBIA/PROJEKTARBEIT-GEOINFORMATIK.GIT)

Projektarbeit unter der Betreuung von PD Dr. Dr.-Ing. Wilfried Linder von 11.2017 - 09.2018 als Vorbereitung der sich anschließenden Masterarbeit.

*Fertigstellung, August 2018*



# Inhaltsverzeichnis

<b>1</b>	<b>Motivation .....</b>	<b>5</b>
<b>2</b>	<b>Geodaten und Geoinformation .....</b>	<b>7</b>
2.1	Definition und Gestalt von Geodaten	7
2.2	Geografische Koordinaten	7
2.3	Qualitätsmerkmale	7
2.4	Georeferenzierung	9
2.5	Geoinformationssysteme	9
2.6	Beispiele	9
2.6.1	Google Maps .....	9
2.6.2	Normbasierte Austauschchnittstelle (NAS) .....	9
2.6.3	Geoobjekte .....	9
2.6.4	Mobile Mapping .....	9
2.7	Algorithmen in der Geoinformatik	9
2.8	Verschiedene Arten und ihre Anwendungszwecke	9
2.9	Beschaffung von Geodaten	9
<b>3</b>	<b>Geoanalyse .....</b>	<b>11</b>
3.1	Beispiele	11
3.2	Offene Probleme	11

---

<b>4</b>	<b>Neuronale Netzwerke .....</b>	<b>13</b>
4.1	Anwendungszwecke	13
4.2	Tensorflow	13
4.3	Anwendung auf Satellitendaten	13
<b>5</b>	<b>Clusteringverfahren .....</b>	<b>15</b>
5.1	Probabilistisches und Possibilistisches Clustering	15
5.1.1	FCM und PFCM .....	15
5.1.2	Vorraussetzungen für die Anwendung auf Geodaten .....	15
5.1.3	Eigener Algorithmus (noch ohne Name) .....	15
5.2	CVI	15
5.2.1	NPC .....	15
5.2.2	FHV .....	15
5.2.3	Otsu-Binarisierung .....	15
5.2.4	VAT-Algorithmus .....	15
<b>6</b>	<b>Forschungsbedarf .....</b>	<b>17</b>
6.1	Ausblick - Mein Thema für die Masterarbeit	17





## 1. Motivation





## 2. Geodaten und Geoinformation

### 2.1 Definition und Gestalt von Geodaten

Geodaten sind digitale Informationen, welche Sachdaten mit Geometriedaten<sup>1</sup> (und Chronometriedaten) vereinen, z.B. {Luftdruck 1 bar, Ort Düsseldorf, Datum 26.11.2017}. Die räumliche Information kann in unterschiedlichen Formen vorliegen, z.B. symbolisch als Ortsname oder Postleitzahl, aber auch als mathematisch atomare Referenz auf Positionen der Erde mittels Koordinaten. Diese können in unterschiedlichster Dimensionalität vorliegen:

- Kugelkoordinaten mit Bezug auf jeden Punkt im Volumen der Erde als Geoid oder Rotationsellipsoid (3D)

- Gauß-Krüger oder geografische Koordinaten mit Bezug auf die Oberfläche der Erde ohne Berücksichtigung von Höhenunterschieden (2D)

- 2D-Koordinaten mit einer zusätzlichen Sachinformation für die Höhe über dem Geoiden (2.5D).

### 2.2 Geografische Koordinaten

Ein geeignetes und weit verbreitetes Koordinatensystem zur verzerrungsarmen Darstellung sind die Geografischen Koordinaten. Abbruch der Arbeit an dieser Stelle, da Relevanzzweifel...

### 2.3 Qualitätsmerkmale

Ein wichtiger Forschungszweig ist die automatische Beurteilung von Qualitätsmerkmalen von Geodaten hinsichtlich einer bestimmten Fragestellung. Ein geeignetes Maß ist die gewichtete Summe verschiedener Datenmerkmale, welche in der aktuellen ISO-Norm *ISO 19157:2013*<sup>2</sup> spezifiziert sind. Die folgende Auflistung ist eine informelle Beschreibung der oben genannten Norm durch Fragestellungen und Beispiele:

---

<sup>1</sup><https://www.hdm-stuttgart.de/riekert/lehre/gis.pdf>

<sup>2</sup><https://www.iso.org/standard/32575.html>

**Vollständigkeit**

- **Datenüberschuss** - Enthält der Datensatz mehr Objekte und Beziehungen als angegeben?
- **Datenmangel** - Enthält der Datensatz weniger Objekte und Beziehungen als angegeben?

**Logische Konsistenz**

- **Konzeptuelle Konsistenz** - Wurde die Gestalt des Datenmodells bei Aktualisierungen nicht verändert?
- **Wertekonsistenz** - Sind alle Werte sinnvoll?
- **Formatkonsistenz** - Passen die Daten zu angegebenen physikalischen Einheiten?
- **Topologische Konsistenz** - Bleiben topologische Beziehungen bei Änderungen des Datensatzes bestehen (Der botanische Garten befindet sich im Umkreis von 1km von der HHU)?
- **Geometrische Konsistenz** - Ist der digitalisierte Datensatz geometrisch sinnvoll und widerspruchsfrei?

**Positionsgenauigkeit**

- **Äußere Genauigkeit** - Wie gut stimmen die Koordinatenwerte des Datensatzes mit den wahren Koordinaten überein?
- **Innere Genauigkeit** - Wie gut stimmen die relativen Positionen von Objekten zueinander mit den wahren relativen Positionen überein?
- **Rasterdatengenauigkeit** - Wie gut stimmen die Rasterdatenpositionswerte mit den wahren Werten überein?

**Zeitliche Genauigkeit**

- **Genauigkeit von Zeitmessungen** - Wie genau ist die Zeitangabe (minutengenau, tagengenau)?
- **Zeitliche Konsistenz** - Ist die Reihenfolge der Ereignisse korrekt?
- **Zeitliche Gültigkeit** - Ist der Datensatz in Bezug auf das geforderte Zeitformat korrekt?

**Thematische Genauigkeit**

- **Richtigkeit der Klassifikation** - Stimmen Objekte, oder ihre Attribute mit den zugewiesenen Klassen überein, z. B. Zuordnung zu Fluss, statt zu Weg
- **Richtigkeit nichtquantitativer Attribute** - Beispiel: Ist das Grundstück wirklich eine Bananenplantage?
- **Genauigkeit quantitativer Attribute** - Beispiel: Ist die Fläche des Grundstücks korrekt?

Viele der oben genannten Punkte lassen einen subjektiven Spielraum für die Bewertung zu. Sowohl Skalierungen als auch Gewichtungen sind nicht eindeutig definiert, was einen Vergleich verschiedener Datensätze erschwert. Aus diesem Grund ist eine algorithmische Interpretation in Kombination mit verfahren der künstlichen Intelligenz hilfreich. So ließe sich aus der Norm ein universeller und allgemeingültiger Indikator zur Bewertung der Datenqualität ermitteln.



**2.4 Georeferenzierung**

**2.5 Geoinformationssysteme**

**2.6 Beispiele**

**2.6.1 Google Maps**

**2.6.2 Normbasierte Austauschschnittstelle (NAS)**

**2.6.3 Geoobjekte**


**2.6.4 Mobile Mapping**

**2.7 Algorithmen in der Geoinformatik**

**2.8 Verschiedene Arten und ihre Anwendungszwecke**

**2.9 Beschaffung von Geodaten**





## 3. Geoanalyse

3.1 Beispiele

3.2 Offene Probleme





## 4. Neuronale Netzwerke

- 4.1 Anwendungszwecke
- 4.2 Tensorflow
- 4.3 Anwendung auf Satellitendaten







## 5. Clusteringverfahren

### 5.1 Probabilistisches und Possibilistisches Clustering

#### 5.1.1 FCM und PFCM

#### 5.1.2 Voraussetzungen für die Anwendung auf Geodaten

#### 5.1.3 Eigener Algorithmus (noch ohne Name)

### 5.2 CVI

#### 5.2.1 NPC

#### 5.2.2 FHV

#### 5.2.3 Otsu-Binarisierung

#### 5.2.4 VAT-Algorithmus





## 6. Forschungsbedarf

### 6.1 Ausblick - Mein Thema für die Masterarbeit