

Zusammenfassung - Mustererkennung und Kontextanalyse

Andreas Ruschinski

20. Oktober 2015

Korrektheit und Vollständigkeit der Informationen sind nicht gewährleistet. Macht euch eigene Notizen oder ergänzt/korrigiert meine Ausführungen!

Inhaltsverzeichnis

1 Überblick Klassifikation	1
2 Grundlagen Signalverarbeitung	2

1 Überblick Klassifikation

1.1 Einführendes Beispiel

- Ziel: Bestimmung von Fischen auf der Basis von Kamerainformationen
- Verwendung der Kamera zum Merkmale (Features) des aktuellen Fisches zu bestimmen: Länge, Helligkeit, Breite
- Annahme: Die Modelle (Beschreibung) der Fischen unterscheiden sich
- Klassifikation: finde zu gegebene Merkmalen das am besten passende Modell (Welche Beschreibung der Fische passt am besten zu den aktuellen Fisch)

1.2 Mustererkennungssysteme

- 1) **Sensing:** Erfassung der Umwelt mittels Sensoren, z.B: Kamera, Bewegungssensoren, Mikrophon, RFID-Lesegerät, Problem: Eigenschaften und Begrenzungen (Bandbreite, Auflösung, Empfindlichkeit, Verzerrung, Rauschen, Latenz, ...)des Sensors beeinflussen Problemschwierigkeit
- 2) **Segmentation:** Identifikation der einzelnen Musterinstanzen (Identifikation der für unser Problem relevanten Daten), z.B: Fisch aufn Fließband
- 3) **Merkmalsberechnung:** Bestimmung von Features, gesucht sind dabei Features welche eine Diskriminierungsfähigkeit haben (Zwischen zwei gleichen Klassen ähnlicher Wert, Zwischen zwei verschiedenen Klassen großer Werteunterschied) und Invariant gegenüber Signaltransformationen (Rotation, Translation, Skalierung, perspektivische Verzerrung) sind; Problem: Wie kann ich aus einer großen Auswahl an Merkmalen die besten geeigneten finden?
- 4) **Klassifikation:** Annahme: Modelle (Grundlegende Eigenschaften) der Klassen (verschiedene Fische) unterscheiden sich; Ziel: Zuweisung von Probleminstanzen aufgrund deren Merkmale zu der am besten entsprechenden Klasse
- 5) **Nachbereitung:** Entscheidung auf Basis Klassifikation, Problem: Wie können wir Kontextinformationen nutzen? Können wir verschiedene Klassifikatoren zusammen nutzen?

1.3 Entwurf von Mustererkennungssystemem

- 1) **Daten sammeln:** Wie kann man wissen, wann eine Menge von Daten ausreichend groß und repräsentativ ist, um das Klassifikationssystem zu trainieren und zu testen?

- 2) **Merkmale bestimmen:** Gesucht: Einfach zu extrahierende Merkmale mit hoher Diskriminativität, Invariant gegenüber irrelevanten Transformationen, unempfindlich gegenüber Rauschen, Problem: Wie kann ich A-priori-Wissen nutzen?
- 3) **Modell auswählen:** Wie erkennt man, wann ein Modell sich in der Klassifikation signifikant von einem anderen Modell – oder vom wahren Modell – unterscheidet? Wie erkennt man, dass man eine Klasse von Modellen zugunsten eines anderen Ansatzes ablehnen sollte? Versuch-und-Irrtum oder gibt es systematische Methoden?
- 4) **Klassifikator trainieren:** Verwende gesammelte Daten, um Parameter des Klassifikators zu bestimmen
- 5) **Klassifikator evaluieren:** Wie bewertet man die Leistung? Wie verhindert man Overfitting/Underfitting?

Für weitere Informationen sind folgende Referenzen zu konsultieren: [1, S. 3-16]

2 Grundlagen Signalverarbeitung

2.1 Digitalisierung

In diesem werden die Grundlagen der Digitalen Signalverarbeitung beschrieben. In der Digitalen Signalverarbeitung werden Methoden und Techniken behandelt welche aus den analogen Sensorwerten eine digitale Information erstellen.

Herkömmliche Sensoren liefern aktuelle Werte anhand von Spannungen (Bsp: Drucksensor: $1g = 0,1V$). Dies analoge Signale werden im ersten Schritt zeitlich diskretisiert d.h. die kontinuierlichen Signale werden durch Abtastung angenähert. Im nächsten Schritt werden die abgetasteten Werte diskretisiert (Diskretisierung der Amplituden) d.h. jeden abgetasteten Wert wird ein Wert digitaler Wert zugewiesen.

Unter Abtastung versteht man die Erhebung eines Wertes zu einem Zeitpunkt. Die Häufigkeit der Abtastung wird in Herz (Hz) angegeben d.h. Abtastung mit 10 Hz entspricht 10 maliges abtasten des analogen Signales innerhalb von einer Sekunde. Die Diskretisierung der Amplitude erfolgt mittels einem Analog-Digital-Wandler der entsprechend von Grenzwerten ($1/2$ Spannung, $1/4$ Spannung, $1/8$ Spannung) entsprechende Bits setzt und diese Information ausgibt.

2.1.1 Dithering

Ein Problem bei der Diskretisierung der Amplitude ergibt sich dadurch dass bei einem sehr geringen Sensorwerte entsprechende Bits nichts gesetzt werden. Dies hat zur Folge dass Informationen verloren gehen. Um dies zu verhindern wird Zufallsrauschen auf den aktuellen Sensorwert addiert. Dadurch wird der Grenzwert manchmal überschritten. Dadurch nähert sich der Erwartungswert den Realwert an.

2.1.2 Abtastung

Abtasttheorem. Ein Signal nur dann kann korrekt abgetastet werden, wenn es keine Frequenzanteile enthält, die oberhalb der halben Abtastrate liegen.

$$f_{max} \leq \frac{1}{2} f_{sample}$$

Aus dem Abtasttheorem folgt wenn wir ein Signal mit f_{max} korrekt abzutasten müssen wir dieses Signal mit einer Frequenz von $2 * f_{max}$ abtasten.

2.2 Lineares System

- f ist linear gdw. $f(c*(a+b)) = c*(f(a)+f(b))$ d.h. wenn f linear ist können wir die Konstante herausziehen und die einzelnen Bestandteile untersuchen
- Homogenität: $f(c*a) = c*f(a)$ d.h. wir können Konstanten aus unseren Berechnungen ausschließen und unser Ergebnis mit der Konstante multiplizieren
- Additivität: $f(a+b) = f(a) + f(b)$ d.h. wir können die zusammengesetzte Argumente von f getrennt betrachten und unsere Teilergebnisse addieren

Literatur

[1] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.