

Schätzung notwendiger Versuchszahlen für Überwachtes Lernen von Objektklassifikation auf Bilddaten

Projektbeschreibung

Unter Objektklassifikation versteht man die Zuordnung von Objekten zu vorab bekannten Klassen. Die Zuordnung basiert häufig auf Fotografien der Objekte. Bei der wiederholten Anwendung von (teils verschiedenen) Lernalgorithmen zur Klassifikation ergibt sich aus einem Lerndatensatz eine Matrix, die abbildet welche Objekte wie oft welcher Klasse zugeordnet wurden. Ausgehend von der Annahme, dass die Häufigkeiten innerhalb einer Zeile einer Multinomialverteilung folgen, kann eine Parameterschätzung (Bayes) dieser Verteilung erfolgen. Hierbei kann vorab berechnet werden, wieviele Lerndaten notwendig sind, um einen vorgegebenen Vertrauensbereich bezüglich dieser Parameterschätzung einzuhalten. Als konjugierte Verteilung dient hierbei die Dirichlet-Verteilung.

Das verwendete Dirichlet-Multinomial-Modell kann auf den Fall von Intervallwahrscheinlichkeiten verallgemeinert werden, siehe: imprecise Dirichlet model (IDM)¹. Mit den IDM lässt sich Unwissen besser modellieren, und die Konvergenz der Ober- und Unterschranken der Intervallwahrscheinlichkeiten kann anhand realistischer Beispiele aus dem robotischen Anwendungsfeld untersucht werden.



Objektklasse	Geschätzte Klasse			
	Tasse	Flasche	Glas	Teller
Tasse	10	1	3	0
Flasche	2	8	5	1
Glas	4	1	5	1
Teller	1	1	3	9

Mögliche Aufgabenstellungen

Mögliche Fragestellungen können je nach Art der Arbeit (Werkstudententätigkeit, Abschlussarbeit, etc.) folgende Punkte umfassen.

- Vergleich verschiedener Varianten zur Bestimmung der notwendigen Versuchszahlen im Lernprozess.
- Vergleich des Dirichlet-Multinomial-Modells mit dem IDM.
- Untersuchung des Einflusses frei wählbarer Parameter im IDM.

Fachliche Anforderungen

Bewerber sollten Kenntnisse oder die Bereitschaft sich die Kenntnisse anzueignen aus mindestens einem der folgenden Gebiete mitbringen:

- Erfahrung in der Softwareentwicklung und Programmierung in Python, R oder C++
- Erfahrung in der statistischen Datenverarbeitung
- Kenntnisse in Bayes'scher Statistik und/oder Intervallwahrscheinlichkeiten (coherent lower previsions).

Ein abgeschlossenes Grundstudium oder Bachelor in Statistik, Informatik, Elektrotechnik, Mathematik oder einem verwandten Studiengang ist hilfreich, aber nicht notwendig.

Ansprechpartner

Christian Rink, Tel. 08153 28-3080, E-Mail: christian.rink@dlr.de

¹<http://dx.doi.org/10.1016/j.ijar.2004.10.002>