Multivariate Erweiterung von univariaten Skalierungsverfahren für Zugehörigkeitswerte

Martin Gebel

28. April 2009

Im Zeitalter des Computers ist die Bedeutung der Datenanalyse immer weiter gestiegen. Eine wichtige Analysemethode ist hier die Klassifikation mit überwachtem Lernen, für die neue Verfahren in zwei unterschiedlichen wissenschaftlichen Disziplinen, Statistik und Informatik, entwickelt werden. Aufgrund der zahlreichen verschiedenen Klassifikationsverfahren ist eine Vereinheitlichung und Vergleichbarkeit der Klassifikationsergebnisse, der so genannten Zugehörigkeitswerte, ein wünschenswertes Ziel.
Ziel dieser Arbeit ist es, einen vereinheitlichenden Rahmen für die Entwicklung von vergleichbaren Zugehörigkeitswerten zu bieten. Klassifikationsverfahren, die auf der statistischen Modellierung basieren, berechnen Zugehörigkeits–Wahrscheinlichkeiten, die eine statistische Unsicherheit in der Zuweisung widerspiegeln. Dagegen erzeugen Maschinelle Lernverfahren in der Regel Zugehörigkeitswerte, die nicht die Kriterien einer Wahrscheinlichkeit erfüllen. Zugehörigkeits–Wahrscheinlichkeiten bieten im Gegensatz zu diesen unstandardisierten Zugehörigkeitswerten die Möglichkeit, die Klassifikationsresultate sowohl weiter zu verarbeiten als auch zu vergleichen. Durch eine Skalierung der unstandardisierten Zugehörigkeitswerte in Zugehörigkeits–Wahrscheinlichkeiten wird dies auch für Ergebnisse von maschinellen Lernverfahren ermöglicht.
Für den Fall einer dichotomen Klassifikation existieren diverse Methoden, die für die eine Klasse die Zugehörigkeits–Wahrscheinlichkeit ermitteln und für die andere Klasse die Gegenwahrscheinlichkeit verwenden. Ein umfassender Vergleich dieser Methoden in dieser Arbeit zeigt, dass die Skalierungsverfahren mittels Logistischer Regression von Platt bzw. mittels der Beta–Verteilungsfunktion von Garczarek die besten Resultate erzielen. Für den multivariaten Klassifikationsfall mit Klassenanzahl $K > 2$ ist die Methode mittels *Pairwise Coupling* von Hastie und Tibshirani bzw. eine Verallgemeinerung für ECOC-Matrizen von Zadrozny das Standardvorgehen. Grundla-

ge hierfür ist eine Reduktion des multivariaten Klassifikations–Problems auf mehrere Zweiklassen–Probleme und eine anschließende Skalierung der Resultate zu Zugehörigkeits–Wahrscheinlichkeiten für die diversen Zweiklassen–Probleme. Anschließend werden diese binären Zugehörigkeits–Wahrscheinlichkeiten anhand des Pairwise Coupling Algorithmus zu einer $N \times K$–Matrix von Zugehörigkeits–Wahrscheinlichkeiten für die multivariate Klassifikation transformiert.
Diese Arbeit stellt dem Verfahren des Pairwise Coupling eine neue Skalierungsmethode unter Verwendung der Dirichlet–Verteilung gegenüber. Die *Dirchlet–Skalierung* hat den Vorteil, dass hiermit aus den unskalierten Zugehörigkeitswerten für die Zweiklassen–Probleme in nur einem Schritt die Matrix von Zugehörigkeits–Wahrscheinlichkeiten für den multivariaten Klassifikationsfall erstellt werden kann. Außerdem ist dieses Verfahren nicht ausschließlich auf die Zugehörigkeitswerte der Zweiklassen–Probleme anwendbar. Es ist das erste Skalierungsverfahren, für das auch eine direkte Anwendung auf eine (unskalierte) Matrix von Zugehörigkeits–Wahrscheinlichkeiten für den multivariaten Klassifikationsfall möglich ist. In der Parameter–Optimierung im Rahmen dieser Dirchlet–Skalierung wird ein Pareto-Optimum für die folgenden Kriterien gefunden: Korrektheitsrate des Klassifikators, Widerspiegelung der Korrektheit innerhalb einer Zuweisungsklasse und hohe Effektivität in der Klassenzuweisung.
Abschließend werden in Experimenten die Ergebnisse einer Anwendung der multivariaten Skalierungsverfahren auf Zugehörigkeitswerte maschineller Lerner verglichen. Diese Analysen umfassen Künstliche Neuronale Netze sowie die Support Vector Machine. Hier zeigt sich, dass die Dirichlet–Skalierung, besonders für balancierte Datensätze, vergleichbare Ergebnisse erzielt.