

Geleitwort

Statistisches Matching wurde ursprünglich zur Unterstützung der Marktforschung entwickelt. Um reichhaltigere Informationen über das Verbraucherverhalten gewinnen zu können, verschmolz man in den 1960er Jahren eine Erhebung zum Konsumverhalten mit einer Erhebung über Fernsehgewohnheiten zu einer einzigen Menge von Datensätzen, die dann Informationen zum Konsum- und Fernsehverhalten gleicher Objekte beinhaltet.

Ziel des statistischen Matchings ist es, weitere Informationen über Individuen zu erlangen, indem relevante Attribute ihrer sog. statistischen Zwillinge aus anderen Mengen von Datensätzen hinzugefügt werden. Traditionelle Matchingverfahren ermitteln die statistischen Zwillinge auf Grundlage der Distanzen zwischen den Ausprägungen der Datensätze in den sog. Matchingvariablen, die allen Datensätzen gemein sein müssen.

In der vorliegenden Arbeit wird eine Methode des statistischen Matchings mit Fuzzy Logic entwickelt. Der Autor nennt diese Methode statistisches Fuzzy-Matching. Durch die Verwendung der Theorie unscharfer Mengen zur Vorverarbeitung der Daten kann erstens eine neue Alternative zur Bestimmung der Distanzen zwischen Datensätzen entwickelt werden und zweitens wird die direkte Einbeziehung nominal und ordinal skalierten Variablen in den Matchingprozess ermöglicht. Insbesondere letzteres ist bei traditionellen Methoden nicht ohne aufwändige Vorverarbeitungen der Daten möglich. Die Umwandlung der Matchingvariablen in linguistische Variablen mit zugehörigen linguistischen Termen gestattet es, Distanzen zwischen Datensätzen auf Basis ihrer Zugehörigkeitsgrade zu einer Regelbasis zu bestimmen. Die Erstellung und der Aufbau der Regelbasis werden ebenfalls in dieser Arbeit gezeigt.

Statistisches Fuzzy-Matching dürfte u. a. in solchen Situationen den traditionellen Methoden überlegen sein, wenn kategorielle Variablen eine wichtige Rolle beim Matching spielen. Das in den Werten nicht enthaltene Anwenderwissen kann dann mit Hilfe von Zugehörigkeitsfunktionen eingebracht und für die Ermittlung der statistischen Zwillinge genutzt werden.

Neben der Entwicklung des theoretischen Ansatzes hat der Autor seine Methode auch programmtechnisch umgesetzt. In ausführlich dargestellten Anwendungsbeispielen werden detaillierte Vergleiche des statistischen Fuzzy-Matchings mit traditionellen Methoden gezogen. Gleichzeitig demonstrieren die Beispiele

die Funktionsweise der Methode und verdeutlichen unterschiedliche Ansatzpunkte des statistischen Matchings. Beim Fuzzy-Matching ist zwar bei metrisch skalierten Matchingvariablen ein etwas höherer Aufwand erforderlich, um bspw. die Definitions- und Wertebereiche der linguistischen Terme festzulegen. Dafür kann die Matching-Güte besser ausfallen, für deren Bestimmung der Autor ebenfalls eine neuartige Alternative vorstellt.

In der Praxis kann das Verfahren zur Datenanreicherung von Datenbeständen im Rahmen von Business Intelligence eingesetzt werden, die zunehmend eine wichtige Rolle auch in kleineren Unternehmen spielt, oder um umfangreiche Kundendaten unter Beachtung des Datenschutzes nutzen zu können.

Paul Alpar