

Inhaltsverzeichnis

Abbildungsverzeichnis	XV
Tabellenverzeichnis	XVII
Abkürzungsverzeichnis	XXI
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	2
1.3 Gang der Untersuchung	3
2 Knowledge Discovery in Databases	5
2.1 Begriffsdefinition	5
2.2 Einordnung des Forschungsgebietes	9
2.3 Vorgehensmodell	13
2.3.1 Zielsetzung und Aufgabenstellung	15
2.3.1.1 Erstellung eines Klassifikationsmodells	16
2.3.1.2 Erstellung eines Regressionsmodells	17
2.3.1.3 Bildung von Clustern	18
2.3.1.4 Entdeckung von Abhängigkeiten	19
2.3.2 Datenvorbereitung	20
2.3.2.1 Datenauswahl und Datenintegration	21
2.3.2.2 Datenbereinigung	23
2.3.2.3 Attributsmanipulation	27
2.3.2.4 Datentransformation	28
2.3.2.5 Aufteilung des Datenbestandes	29
2.3.3 Data Mining	30
2.3.3.1 Ausgewählte Verfahren	31
2.3.3.1.1 Entscheidungsbaumverfahren	33

2.3.3.1.2	Künstliche Neuronale Netze	37
2.3.3.1.3	Clusterverfahren	41
2.3.3.2	Abgrenzung zu anderen Analyseansätzen	44
2.3.3.2.1	Data Access	46
2.3.3.2.2	On-Line Analytical Processing	47
2.3.4	Ergebnisinterpretation	49
3	Assoziationsanalyse	51
3.1	Anwendungsgebiete	52
3.2	Einfache Assoziationsregeln und Kenngrößen	53
3.3	Algorithmen zur Bestimmung von einfachen Assoziationsregeln	58
3.3.1	Apriori-Algorithmus	58
3.3.2	Weitere Algorithmen	66
3.4	Erweiterungen der Assoziationsregeln	70
3.4.1	Item-Restriktionen	71
3.4.2	Taxonomien	72
3.4.3	Quantitative Merkmale	78
3.4.4	Sequenzielle Assoziationsregeln	80
3.5	Bewertung von Assoziationsregeln	82
3.5.1	Kritik am Support- und am Konfidenzmaß	82
3.5.2	Interessantheitsmaße	85
3.5.2.1	Lift	85
3.5.2.2	Conviction	87
3.5.2.3	Chi-Quadrat-Wert	90
3.5.2.4	Relatives Risiko	91
3.5.3	Filtertechnik	92

4	Konzepte zur Bildung und Anwendung von Stichproben	95
4.1	Definitionen	95
4.2	Verfahren zur Stichprobenbildung	97
4.2.1	Bernoulli-Sampling	97
4.2.2	Zufallsauswahl mit festem Umfang	99
4.2.3	Systematische Zufallsauswahl	101
4.2.4	Geschichtete Zufallsauswahl	102
4.3	Kenngrößen eines Datenbestandes	103
4.3.1	Univariate Kenngrößen	103
4.3.2	Multivariate Kenngrößen	105
4.4	Stichproben im Rahmen des statistischen Schätzens	107
4.4.1	Univariate Schätzer	109
4.4.1.1	Punktschätzer	109
4.4.1.2	Bereichsschätzer	114
4.4.1.3	Stichprobenumfangsbestimmung	118
4.4.1.3.1	Konfidenzintervalle	119
4.4.1.3.2	Chernoff-Schranken	123
4.4.2	Multivariate Schätzer	127
4.4.2.1	Punktschätzer	127
4.4.2.2	Bereichsschätzer	128
4.4.2.3	Stichprobenumfangsbestimmung	132
4.5	Stichproben im Rahmen des statistischen Testens	136

5	Stichprobenbasierte Assoziationsanalyse	139
5.1	Grundprinzip	139
5.2	Erstellung der Konzepte	141
5.2.1	Genauigkeit bei vorliegender Stichprobe	141
5.2.1.1	Univariater Ansatz	142
5.2.1.2	Multivariater Ansatz	144
5.2.2	Stichprobenumfangsbestimmung	145
5.2.2.1	Univariater Ansatz	146
5.2.2.1.1	Absolute Abweichung	146
5.2.2.1.1.1	Ansatz über Konfidenzintervalle	147
5.2.2.1.1.2	Ansatz über Chernoff-Schranken	148
5.2.2.1.1.3	Vergleich	149
5.2.2.1.2	Relative Abweichung	151
5.2.2.1.2.1	Ansatz über Konfidenzintervalle	151
5.2.2.1.2.2	Ansatz über Chernoff-Schranken	153
5.2.2.1.2.3	Vergleich	154
5.2.2.2	Multivariater Ansatz	155
5.2.3	Regelentdeckung	157
5.2.3.1	Ansatz über Konfidenzintervalle	158
5.2.3.2	Ansatz über Testtheorie	161
5.2.4	Handlungsempfehlung	163
5.3	Anwendung der erstellten Konzepte	164

6	Anwendung der stichprobenbasierten Assoziationsanalyse für Problemstellungen aus der Telekommunikationsbranche	171
6.1	Umfeld der Anwendung	171
6.1.1	Struktur des festnetzgebundenen Fernsprechnetzes	172
6.1.2	Einbindung der Interconnectionspartner	174
6.1.2.1	Netzstruktur der Interconnectionspartner	175
6.1.2.2	Abrechnung von Interconnectionsvermittlungen	179
6.1.2.3	Verkehrsströme	183
6.2	Problembeschreibung und Aufgabenstellung	185
6.2.1	Problembeschreibung	185
6.2.1.1	Fehlrouting	185
6.2.1.2	Bündelvertauschung	187
6.2.1.3	Zusammenhang zwischen Fehlrouting und Bündelvertauschung	190
6.2.2	Aufgabenstellung	190
6.3	Durchführung der stichprobenbasierten Assoziationsanalyse	192
6.3.1	Datenvorbereitung	193
6.3.2	Analyse	197
6.3.3	Ergebnisinterpretation	199
6.3.3.1	Regelauswahl	199
6.3.3.2	Regelbewertung	204
6.4	Beurteilung der durchgeführten stichprobenbasierten Assoziationsanalyse	206
6.4.1	Zeitbetrachtung	207
6.4.2	Genauigkeitsbetrachtung	216
7	Zusammenfassung und Ausblick	225
	Literaturverzeichnis	231
	Anhang	261