

Inhaltsverzeichnis

1 Einführung	1
1.1 Problem	1
1.2 Ziel	2
1.3 Aufbau.....	2
1.4 Hintergrund.....	2
1.5 Forschungsmethodik.....	4
2 Konzept und Stand der Datenmustererkennung	5
2.1 Rahmenkonzept.....	5
2.2 Problemfelder.....	6
2.2.1 Autonomie.....	6
2.2.2 Allgemeine Verwendbarkeit	8
2.2.3 Datenproblematik	8
2.2.4 Verständlichkeit	10
2.2.5 Sicherheit.....	10
2.2.6 Interessantheit.....	10
2.2.7 Verfahrenseffizienz	11
2.3 Modell eines Data-Mining-Systems	12
2.3.1 Steuerung.....	12
2.3.2 Datenbankschnittstelle	13
2.3.3 Wissensbasis (Domänenwissen und Benutzereingaben).....	13
2.3.4 Fokussierung	14
2.3.5 Analysealgorithmen	15
2.3.6 Bewertung	15
2.3.7 Präsentation	16
2.4 Stand der Forschung	16
2.4.1 Überblick.....	16
2.4.2 Abgrenzung zu verwandten Gebieten	19
2.4.3 Systembeispiele.....	20
2.4.3.1 Abweichungsanalyse.....	27
2.4.3.2 "Wenn-Dann"-Regelsuche	29
2.4.3.3 Werkzeugpools.....	30
2.4.3.4 Grafisches Data Mining	30
2.4.3.5 Entscheidungsbaumverfahren	31
2.4.3.6 Clusternde Systeme.....	32
2.4.3.7 Neuronale Netze.....	32
2.4.3.8 Genetische Algorithmen	33
2.4.4 Aktuelle Entwicklungen.....	33

3 Datenmustererkennung im Ergebniscontrolling	35
3.1 Spezifika des Ergebniscontrollings	35
3.1.1 Mehrdimensionalität des Ergebnisses	35
3.1.2 Atomare Datengrundlage	36
3.1.3 Umfang der Kombinatorik	37
3.2 Grenzen bisheriger Ansätze	37
3.3 Domänenspezifische Anforderungen	40
3.3.1 Dimensionen und Merkmale	40
3.3.2 Grenzen der Interpretation	40
3.3.3 Verdichtung	41
3.3.4 Segmentierung	42
3.3.5 Fokussierung	42
3.3.6 Nachbehandlung der Daten	43
3.3.7 Datendeformation	43
3.3.8 Bewertung der Auffälligkeit	44
3.3.9 Betriebswirtschaftliche Muster	45
3.3.9.1 Konzentrationen	46
3.3.9.2 Wiederholungen	46
3.3.9.3 Abweichungen	46
3.3.9.4 Abhängigkeiten	46
3.3.9.5 Ähnlichkeiten (Portfoliomuster)	47
3.3.9.6 Rangfolgen	48
3.3.9.7 Entwicklungen	48
3.3.9.8 Sequenzen	49
3.3.9.9 Zusammenfassung	49
3.4 Verfahren für Mustererkennung und Musterbeschreibung	51
3.4.1 Mustererkennung	51
3.4.1.1 Clusteranalyse	52
3.4.1.1.1 Überblick	52
3.4.1.1.2 Proximitätsmaße	52
3.4.1.1.3 Clusteranalysealgorithmen	53
3.4.1.1.4 Bewertung	54
3.4.1.2 Bayes-Verfahren	55
3.4.1.2.1 Beschreibung	55
3.4.1.2.2 Bewertung	56
3.4.1.3 Fuzzy-Datenanalyse	56
3.4.1.3.1 Überblick	56
3.4.1.3.2 Fuzzy-Clusteranalyse	56
3.4.1.3.3 Fuzzy-Mustererkennung	57
3.4.1.3.4 Bewertung	59

3.4.2	Musterbeschreibung	60
3.4.2.1	Statistische Verfahren	60
3.4.2.1.1	Regressionsanalyse	60
3.4.2.1.2	Varianzanalyse	60
3.4.2.1.3	Diskriminanzanalyse	61
3.4.2.1.4	Bewertung	61
3.4.2.2	Maschinelles Lernen	62
3.4.2.2.1	Überblick	62
3.4.2.2.2	Induktionsalgorithmus ID3	62
3.4.2.2.3	Rules-per-class-Ansatz	65
3.4.2.2.4	Erweiterung von ID3 auf den n-Klassen-Fall	66
3.4.2.2.5	Bewertung	68
3.4.3	Hybridverfahren	68
3.4.3.1	Rough Sets	68
3.4.3.1.1	Überblick	68
3.4.3.1.2	Fallbeispiel	69
3.4.3.1.2.1	Grundlagen	69
3.4.3.1.2.2	Analyse von Attributabhängigkeiten	70
3.4.3.1.2.3	Eliminierung überflüssiger Attribute	72
3.4.3.1.2.4	Eliminierung überflüssiger Attributausprägungen	73
3.4.3.1.2.5	Ergebnisse	73
3.4.3.2	Konzeptionelles Clustern	74
3.4.3.2.1	Überblick	74
3.4.3.2.2	Partitioning Module von CLUSTER/2	74
3.4.3.2.3	Bewertung	81
3.4.3.3	RULEARN	81
3.4.3.3.1	Überblick	81
3.4.3.3.2	Fallstudie I: Werkzeugindustrie (WERKZEUG AG)	83
3.4.3.3.2.1	Betrachtetes Unternehmen	83
3.4.3.3.2.2	Datenauswahl	83
3.4.3.3.2.3	Analyseläufe	84
3.4.3.3.2.4	Ergebnisse	85
3.5	Methodenauswahl	88

4	Detailkonzeption von CLUSMIN	91
4.1	Systemarchitektur	91
4.2	Systemkomponenten	92
4.2.1	Steuerung	92
4.2.2	Datenbankschnittstelle	92
4.2.3	Fokussierung	92
4.2.3.1	Überblick	92
4.2.3.2	Felderfokussierung	92
4.2.3.3	Kennzahlenumrechnung	93
4.2.3.4	Feldergewichtung	93
4.2.3.5	Steuerung der Auffälligkeit	93
4.2.3.6	Datenauswahl	93
4.2.3.6.1	Selektionsvektor	94
4.2.3.6.2	Datensatzlimitgesteuerte Verdichtung	95
4.2.4	Wissensbasis	97
4.2.5	Analysekomponenten	98
4.2.5.1	Mustererkennung	98
4.2.5.1.1	Ähnlichkeitsberechnung	98
4.2.5.1.1.1	Gewichtung	99
4.2.5.1.1.2	Proximitätsmaß für Kennzahlen und Merkmale	99
4.2.5.1.2	Clusterbildung	100
4.2.5.1.2.1	Average-Linkage-Verfahren	102
4.2.5.1.2.1.1	Gruppenbildung	102
4.2.5.1.2.1.2	Abbruchbedingungen	103
4.2.5.1.2.2	Sequentielles heuristisches Verfahren	104
4.2.5.1.2.2.1	Behandlung eindeutig zuordenbarer Datensätze	104
4.2.5.1.2.2.2	Behandlung nicht eindeutig zuordenbarer Datensätze	105
4.2.5.2	Musterbeschreibung	105
4.2.5.2.1	Aufbereitung der Gruppen	106
4.2.5.2.1.1	Aufbereitung der Kennzahlen	106
4.2.5.2.1.2	Aufbereitung der Merkmale	107
4.2.5.2.2	Auswahl der Variablen	107
4.2.5.2.2.1	Automatische Auswahl auffälliger Kennzahlen	107
4.2.5.2.2.2	Automatische Auswahl auffälliger Merkmale	109
4.2.5.2.2.3	Benutzergesteuerte Auswahl der Variablen	111

4.2.5.2.3	Beschreibung der Gruppen	111
4.2.5.2.3.1	Aufbau der Beschreibung.....	112
4.2.5.2.3.2	Verbesserung der Beschreibung.....	112
4.2.6	Bewertung	113
4.2.6.1	Vergleichbarkeit von Aussagen	113
4.2.6.2	Aussagendeckung.....	113
4.2.6.3	Überprüfung der Ausschlußaussagen.....	114
4.2.7	Präsentation	115
4.2.7.1	Aussagenpräsentation.....	115
4.2.7.2	Aussagenausschluß	117
4.2.7.3	Exportfunktion	117
4.2.7.4	Objektauswahl.....	117
5	Realisierung von CLUSMIN.....	119
5.1	Entwicklungsumgebung.....	119
5.1.1	Hardware	119
5.1.2	Software	119
5.2	Systemintegration	120
5.2.1	Systemablauf	120
5.2.2	Datenhaltung	124
6	Evaluation von CLUSMIN.....	125
6.1	Evaluation mit Testdaten	125
6.1.1	Parametereinstellungen	125
6.1.1.1	Ähnlichkeitsberechnung.....	125
6.1.1.2	Clusterbildung.....	125
6.1.1.2.1	Average-Linkage-Verfahren	126
6.1.1.2.2	Sequentielles heuristisches Verfahren	126
6.1.1.3	Aufbereitung der Gruppen	126
6.1.1.4	Auswahl der Variablen.....	126
6.1.1.5	Beschreibung der Gruppen.....	127
6.1.1.6	Präsentation.....	127
6.1.2	Testläufe	127
6.1.2.1	Laufzeitverhalten	127
6.1.2.2	Beurteilung der generierten Regeln	128
6.2	Fallstudien mit Praxisdaten.....	129
6.2.1	Fallstudie II: Werkzeugindustrie (WERKZEUG AG).....	129
6.2.2	Fallstudie III: Pharmazeutische Industrie (PHARMA AG).....	131
6.2.3	Fallstudie IV: Nahrungsmittelindustrie (TEE GmbH & Co. KG).....	133
6.2.4	Fallstudie V: Pharmazeutische Industrie (PHARMA AG).....	134
6.2.5	Zusammenfassende Bewertung.....	136

6.3 Integration in das Unternehmenscontrolling.....	137
6.3.1 Form der Mensch-Maschine-Kommunikation	137
6.3.2 Anwendungsschwerpunkt und Einsatzrhythmus	138
7 Zusammenfassung und zukünftige Entwicklungsmöglichkeiten	141
7.1 Zusammenfassung.....	141
7.2 Zukünftige Entwicklungsmöglichkeiten	141
7.2.1 Weitere Fusion	141
7.2.2 Marktpulse	143
Anhang.....	VII
Literaturverzeichnis	VII
Fallbeispiele	XXVI
Trainingsdatensätze zu ID3 (Grundkonzeption)	XXVI
Trainingsdatensätze zum Rules-per-class-Ansatz	XXVII
Entscheidungsbäume	XXVIII
Trainingsdatensätze zum Vier-Klassen-Fall von ID3	XXX
Berechnungsausschnitt zum Vier-Klassen-Fall von ID3	XXXI
Trainingsdatensätze zum konzeptionellen Clustern	XXXVI
Parameter des CLUSMIN-Systems	XXXVII