UNIVERSITY OF DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

COLLABORATIVE RESEARCH CENTER 531

Design and Management of Complex Technical Processes and Systems by means of Computational Intelligence Methods

Merkmalsbasiertes Lernen von Platzierungsregeln im Rahmen der Aufstellungsplanung von Chemieanlagen

Oliver Ritthoff und Bernd Hicking

No. CI-179/04

Technical Report ISSN 1433-3325 September 2004

Secretary of the SFB 531 · University of Dortmund · Dept. of Computer Science/XI
44221 Dortmund · Germany

This work is a product of the Collaborative Research Center 531, "Computational Intelligence," at the University of Dortmund and was printed with financial support of the Deutsche Forschungsgemeinschaft.

Merkmalsbasiertes Lernen von Platzierungsregeln im Rahmen der Aufstellungsplanung von Chemieanlagen

Oliver Ritthoff[1] und Bernd Hicking[2]

[1] Universität Dortmund, Fachbereich Informatik,
Lehrstuhl für künstliche Intelligenz, 44221 Dortmund
E-mail: ritthoff@ls8.cs.uni-dortmund.de
[2] Universität Dortmund, Fachbereich Bio- und Chemieingenieurwesen,
Lehrstuhl für Anlagentechnik, 44221 Dortmund
E-mail: bernd.hicking@bci.uni-dortmund.de

Zusammenfassung Die Aufstellungsplanung hat im Rahmen der Projektierung einer Chemieanlage die Aufgabe, die Maschinen und Apparate der Anlage unter Berücksichtigung verschiedenster Randbedingungen in einem Anlagengerüst zu positionieren. Zur Lösung dieser Aufgabe wurde am Lehrstuhl für Anlagentechnik ein Werkzeug entwickelt, das unter Verwendung einer Wissensbasis einen sowohl funktionellen als auch möglichst kostengünstigen Platzierungsvorschlag generiert. Allerdings können Informationen im Sinne von neuen Platzierungsregeln mit dem System naturgemäß nicht erzeugt werden. An dieser Stelle setzt der im Folgenden beschriebene Lernansatz an, der solche Regeln aus bewerteten Platzierungsbeispielen extrahiert. Bei diesem Ansatz wurde das Lernen von Platzierungsinformationen auf der Basis aussagenlogischer Repräsentationen mit Verfahren aus dem Bereich der evolutionären Algorithmen und des induktiven Lernens realisiert. Verwendet wurde dafür die am Lehrstuhl für künstliche Intelligenz entwickelte Lern- und Experimentierumgebung YALE. Der vorgestellte Ansatz wurde erfolgreich an zwei Anlagen mit unterschiedlichen Repräsentationen erprobt.

1 Einleitung

Das Ziel der Aufstellungsplanung von Chemieanlagen ist die Positionierung von Maschinen und Apparaten einer Chemieanlage in einem Stahlgerüst (Stahlbau). Bei der Platzierung müssen sowohl Anforderungen von Equipment-, Sicherheits- und Prozessseite, als auch Betreiber- bzw. Standortanforderungen berücksichtigt werden. Die gefundene Lösung muss aber nicht nur funktionell, sondern soll auch möglichst kostengünstig sein.
Zur Lösung dieser Problemstellung wurde am Lehrstuhl für Anlagentechnik ein System zur computerunterstützten Aufstellungsplanung entwickelt. Darin werden für ein konkretes Planungsszenario (vorgegebenes Baufeld inklusive des Anlagengerüstes und der Anlagenwege sowie eine Anzahl zu platzierender Equipmentmodelle) die zu erfüllenden Anforderungen unter Verwendung einer in einer

Datenbank hinterlegten Regelbasis abgeleitet. Die inferierten Anforderungen, d.h. die zu berücksichtigenden Abhängigkeiten zwischen einzelnen Equipments bzw. zwischen einzelnen Equipments und dem Baufeld oder Stahlbau, werden in einem zweiten Schritt in eine numerische Form übertragen. Auf der Basis dieser numerischen Anforderungsrepräsentation wird nun ein heuristisches Suchverfahren (ein *Simulated Annealing* (*SA*) *Algorithmus* [Aar92]) zur Minimierung der Regelverstöße bzw. zur Maximierung der erfüllten Anforderungen verwendet.
Die bisherigen Arbeiten haben gezeigt, dass eine Platzierung von Equipments im Anlagengerüst mit dem beschriebenen Verfahren möglich ist. Allerdings ist diese Methode für große Anlagen sehr aufwändig und erfordert viel Erfahrung bei der Festlegung der *SA*-Strategieparameter (insb. der Abkühlvorschriften).
Im Folgenden soll nun ein Lernansatz zur Lösung der beschriebenen Probleme vorgestellt werden, der Wissen über die Güte einzelner Aufstellungen aus bewerteten Platzierungsbeispielen extrahieren kann. Bei diesem Ansatz soll das Lernen guter Platzierungen auf der Basis aussagenlogischer Repräsentationen (d.h. Merkmalsmengen) mit Verfahren aus dem Bereich der evolutionären Algorithmen und des induktiven Lernens realisiert werden (siehe [DDE+02,RK03]). Umgesetzt wurde dieser Ansatz mit der am Lehrstuhl für künstliche Intelligenz entwickelten Lern- und Experimentierumgebung YALE (siehe hierzu auch [FKMR02,MKFR03]). Die umfangreichen empirischen Untersuchungen, die im Abschnitt 3 vorgestellt werden, bestätigen, dass die durch die Merkmalsmenge gegebene Repräsentation einer Lernaufgabe ein entscheidender Faktor für den Lernerfolg im Sinne der Generalisierungsfähigkeit der erzeugten Modelle ist.
Der vorliegende Report ist wie folgt aufgebaut. Der erste Abschnitt gibt eine kurze Einführung in die Grundlagen der Aufstellungsplanung und die in diesem Zusammenhang auftretenden Problemstellungen, gefolgt von einer Beschreibung der bisherigen Verfahren und Werkzeuge zur Aufstellungsplanung von Chemieanlagen, die dem Experten lediglich ein adäquates Umfeld für die **manuelle** Aufstellungsplanung schaffen. Das am Lehrstuhl für Anlagentechnik entwickelte Planungswerkzeug *CAPD* zur computerunterstützten Aufstellungsplanung ist Inhalt des folgenden Abschnittes - besonderes Gewicht wurde bei den Ausführungen auf den Aspekt der Equipmentplatzierung gelegt. Der maschinelle Lernansatz, der unter Verwendung einer *Support Vector Machine* (*SVM*) [Bur98,SS98] aus bewerteten Aufstellungen Platzierungswissen kompiliert (um damit die Aufstellungsgüte einer bisher nicht bewerteten Anlage zu ermitteln) wird im Abschnitt 3 vorgestellt und ausführlich empirisch untersucht[1].

2 Grundlagen der Aufstellungsplanung

Die Aufstellungsplanung hat die Aufgabe, die Maschinen und Apparate einer Chemieanlage unter Berücksichtigung verschiedenster konstruktiver und verfahrenstechnischer Randbedingungen in einem Anlagengerüst zu positionieren.

[1] Zur Durchführung der in Abschnitt 3 beschriebenen Experimente wurde die *SVM*-Implementation `mySVM` von Stefan Rüping [Rü00] verwendet. Diese ist sowohl zur Lösung von Klassifikations- als auch von Regressionsproblemen geeignet.

Zusätzliche Bedingungen ergeben sich aus der Sicht der Montage, des Anlagenbetriebes sowie aufgrund sicherheitstechnischer Anforderungen. Die Entwicklung von Aufstellungsentwürfen wird daher zu einer sehr komplexen Aufgabe, zu deren Lösung ein großes Maß an Fachwissen aus unterschiedlichsten Fachrichtungen erforderlich ist. Es ist zu betonen, dass die Aufstellungsplanung einen Entwurf der Anlage erzeugt, der in der Planungsphase des *Extended Basic Engineering* erstellt und benötigt wird. Dieser Planungsabschnitt ist dadurch gekennzeichnet, dass alle Equipments, dies sind Maschinen und Apparate einer Anlage, zwar bekannt sind und in ihren Hauptabmessungen im Wesentlichen festliegen, genaue Spezifikationen und exakte Konstruktionsmaße in der Regel aber noch nicht vorliegen. Dennoch muss zu diesem Zeitpunkt entschieden werden, wie der Stahlbau für die Anlage aussieht und welche Positionen die einzelnen Equipments dort einnehmen. Es ist zu beachten, dass mit dieser Festlegung sehr wesentliche kostenrelevante Entscheidungen bereits getroffen werden, die während der späteren Planungsphase des *Detail Engineering* aus Kosten- und Termingründen nicht mehr geändert werden können. Damit wird deutlich, wie wichtig es ist, den Aufstellungsentwurf, z.B. mit Hilfe von Variantenkonstruktionen, zu optimieren und die Konsequenzen auf nachfolgende Arbeiten, wie z.B. die Rohrleitungskonstruktion, möglichst genau zu ermitteln.

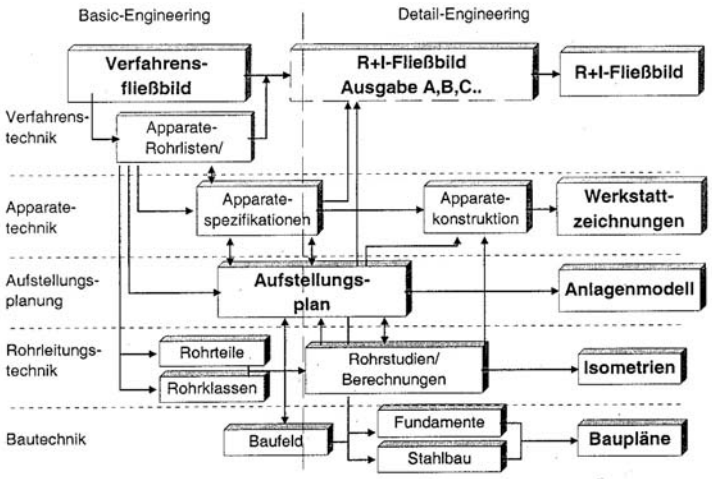


Abbildung 1. Einordnung der Aufstellungsplanung in den Planungsprozess für Chemieanlagen

Abbildung 1 zeigt den Zeitraum und die fachlichen Verknüpfungen der Aufstellungsplanung im Kontext des gesamten Engineerings. Die heutige Praxis beruht im Wesentlichen noch auf einer manuellen Erstellung des Aufstellungsentwurfes, die durch ausgewählte Rohrstudien ergänzt wird. Es ist aus Kostengründen z.Zt. nicht Stand der Technik, entsprechende Variantenkonstruktionen durchzuführen.

Die Aufstellungsplanung muss daher von Planungsingenieuren mit entsprechender Erfahrung ausgeführt und in Diskussionen mit den betroffenen Aufgabengebieten, wie z.B. dem Anlagenbetrieb, erstellt werden.

2.1 Klassische Werkzeuge zur Aufstellungsplanung

In der Praxis existieren derzeit keinerlei Planungswerkzeuge zur computerunterstützten Generierung von Aufstellungsentwürfen mit automatischer Optimierung bzw. Erstellung von Optimierungsvorschlägen. Als Planungstools werden lediglich Lösungen angeboten, die dem Experten ein adäquates *Computer Aided Engineering* (*CAE*)-Umfeld für die manuelle Aufstellungsplanung schaffen. Sie stellen ihm aber keine wissensbasierten Methoden, die den Prozess der Aufstellung selber unterstützen, zur Verfügung. Das *CAE*-Umfeld, welches kommerzielle Planungstools anbieten, konzentriert sich auf die manuelle Positionierung der Equipments mit Hilfe graphischer Software und die Abschätzung der Rohrleitungskonstruktion unter Einsatz von Routingalgorithmen.
Im Bereich der Optimierung von Aufstellungsentwürfen sind bereits zahlreiche Ansätze diskutiert worden. Ein gemeinsames Manko dieser Arbeiten ist sicherlich die Reduktion des Problems auf reine Kostenaspekte. Die Arbeit von Amorese [ACM77] minimiert lediglich die Kosten der Rohrleitungsverbindungen. Malingriaux [MHS70] zieht zusätzliche Kostenfaktoren hinzu, z.B. Bodenkosten, Montagekosten, Inbetriebnahmekosten und sogar Betriebsführungskosten. Eine große Schwierigkeit stellt dabei das exakte Bestimmen dieser Kosten dar. Zusätzlich ist die Aufstellungsplanung nicht ausschließlich dadurch bestimmt, dass lediglich die bekannten Kosten minimiert werden müssen. Vielmehr sind, wie bereits in Abschnitt 2 erwähnt, zahlreiche Aspekte zu berücksichtigen.
Die meisten Ansätze nehmen eine weitere Einschränkung vor, wie z.B. auch der von Georgiadis [GSRM99]. Gemeint ist die Reduzierung der möglichen Aufstellungsplätze auf Rastereinheiten. In der Praxis dürfte es sehr schwierig sein, Ausrüstungen verschiedener Größe allesamt mit dem gleichen Raster zu beschreiben. Auf diese Weise sind keine optimal ausgenutzten Stellflächen zu realisieren. Es ist daher unbedingt anzustreben, die Objekte praktisch frei platzieren zu können, was z.B. bei einer Platzierung in mm-Abständen gewährleistet ist.
Der Ansatz von Penteado und Ciric [PC96] versucht neben den reinen Kostenaspekten auch sicherheitstechnische Kriterien zu berücksichtigen. Diese Idee ist ein kleiner Schritt in die Richtung, der Komplexität der Anforderungen an eine Aufstellung gerecht zu werden. Ein Nachteil den dieses Projekt mit vielen anderen teilt, ist die Formulierung des Problems in *MINLP*-Form[2]. Dadurch wird

[2] *Mixed Integer Nonlinear Programming (MINLP)* [Flo95] bezeichnet die mathematische Programmierung mit kontinuierlichen und diskreten Variablen, wobei sowohl die Zielfunktion als auch die Nebenbedingungen (Constraints) Nichtlinearitäten aufweisen. *MINLP* kann zur Formulierung von Problemen verwendet werden, bei denen eine simultane Optimierung sowohl der Systemstruktur (diskreter Anteil) als auch der Systemparameter (kontinuierlicher Anteil) erforderlich ist. Ein entscheidender Nachteil von *MINLP*-Problemen ist allerdings, dass sie in ihrer allgemeinen Formulierung NP-vollständig sind.

entweder der Rechenaufwand für das verwendete Lösungsprogramm nicht mehr bewältigbar oder die realen Objekte aus dem Anlagenbau werden geometrisch nur unzureichend beschrieben.

2.2 Computer Aided Plant Design (CAPD)

Am Lehrstuhl für Anlagentechnik werden bereits seit einigen Jahren Methoden und Software zur Unterstützung der Aufstellungsplanung und der Rohrleitungskonstruktion entwickelt [ST98]. Die ersten Arbeiten befassten sich schwerpunktmäßig mit der Strukturierung dieses vielschichtigen Planungsprozesses sowie der Entwicklung und Erprobung rechnergestützter Methoden zur Identifikation von Prozessstrukturen und dem Routing von Rohrleitungen [Mut95,Has95]. Während der weiteren Entwicklungen, die stets in Rückkopplung mit Fachleuten aus der Industrie durchgeführt wurden, stellte sich heraus, dass die ursprüngliche Idee, reine Expertensysteme zu entwickeln, kaum zum Ziel führt. Stattdessen zeigte sich, dass die Quantifizierung von Anforderungen und die Weiterentwicklung von Routingalgorithmen erfolgreicher ist [Kös98,Nip00,STBMN97]. Nachdem sich hiermit die Möglichkeit eröffnete, praxisrelevante Probleme zu bearbeiten, stellte sich die kombinatorisch anwachsende Komplexität als neues Hindernis heraus und führte zu weiteren Strukturierungen der Planungsaufgaben durch Schaffung von Standardmodulen [STHKL98,STHL98,Hol00]. Auf dieser Basis gelang es erstmalig, den Aufstellungsplan einer Anlage mit Hilfe eines Routingalgorithmus praxisrelevant zu entwerfen [Leu02,STLL00,STBLL01]. Die Struktur des heutigen *CAPD* (*Computer Aided Plant Design*)-Systems ist in Abbildung 2 dargestellt.

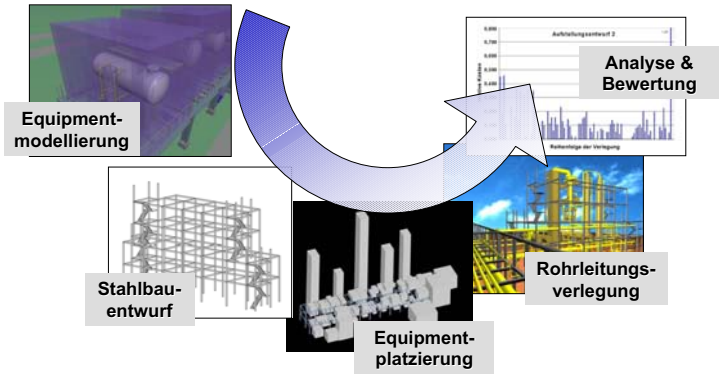


Abbildung 2. Phasen des Aufstellungsentwurfes beim *CAPD*-System

Das Verfahrensfließbild mit den dazugehörigen Maschinen- und Apparatelisten sowie die Abmessungen und äußeren Randbedingungen des Baufeldes sind Vorgabewerte für das Planungssystem. Die Equipment-Modellierung legt auf der

Grundlage der äußeren Abmessungen der Equipments inklusive aller Anbauten Räume für Bedienung und Wartung sowie die Nahverrohrung der Equipments inklusive der Anschlusspunkte für die Prozessleitungen fest. Dahinter steht das Konzept, die Komplexität der Planungsaufgabe zu reduzieren, indem sämtliche volumenmäßig relevanten Teile und Anforderungen, die sich auf ein Equipment beziehen und sich mitbewegen, sobald sich die Position des Equipments verändert, in einem Gesamtvolumen zusammengefaßt werden. Ein Beispiel für ein Equipment in der Modellierungsansicht ist der in Abbildung 3 dargestellte liegende Behälter mit dem ihn umgebenden Volumenkörper, der den Gesamt-Platzbedarf des Behälters mit Anbauten und der ihm zugeordneten Räume kennzeichnet.

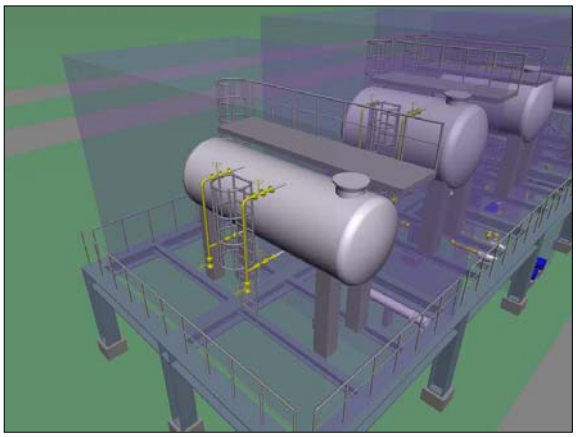


Abbildung 3. Equipment-Modell eines liegenden Behälters

Im nächsten Schritt wird der Stahlbau als rechtwinkliges Fachwerk entworfen, wobei die Grundfläche auf der Grundlage des Platzbedarfs der Equipments, die im Erdgeschoss platziert werden müssen, abgeschätzt wird. Nach der Platzierung der Equipments in dem unzerklüfteten, auf der geschätzten Grundfläche basierenden Stahlbau, werden die bei der Platzierung nicht belegten Stahlbauraster entfernt.
Für die Platzierung der Equipments bzw. ihrer Volumenkörper ist eine neuartige Methode entwickelt worden, die auf dem *force directed placement* beruht. Die Grundidee besteht darin, dass Equipments untereinander sowie zu der Struktur des Stahlbaus und Anschlusspunkten am Rand des Baufeldes in Beziehung stehen. Diese Beziehungen ergeben sich z.B. aufgrund des Prozesses, der u.a. durch die verbindenden Rohrleitungen sowie durch Funktionseinheiten, wie z.B. Kolonnensysteme, beschrieben wird. Weitere Anforderungen ergeben sich aufgrund der Sicherheitstechnik, des Anlagenbetriebes oder der Anlagenmontage.

Zur Verdeutlichung des *force directed placements* siehe Abbildung 4. Dargestellt sind zwei Pumpenpaare P_1 und P_2, eine Kolonne K_1 sowie ein Anlagenweg. Die Anforderungen (z.B. "P_1 sollte in der Nähe von K_1 stehen" und "K_1 muss am Anlagenrand stehen") werden durch die entsprechenden Verbindungslinien zwischen den Equipments bzw. zwischen den Equipments und dem Anlagengelände dargestellt, wobei die Stärke der Verbindungen ein Produkt aus deren Länge und Gewichtung ist.

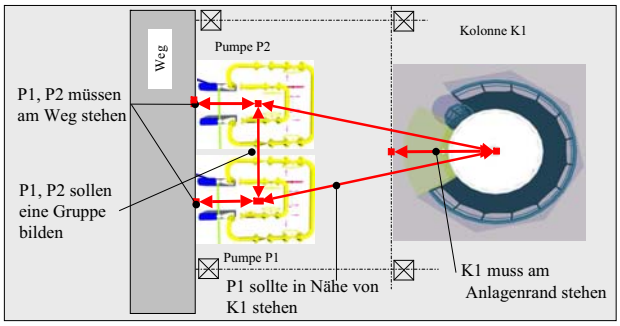


Abbildung 4. Beispiel für das *force directed placement* mit zwei Pumpenpaaren und einer Kolonne

Vor dem eigentlichen Positionieren der Ausrüstungen müssen diese Anforderungen zunächst mit Hilfe einer Wissensbasis generiert werden. Dazu wird der Bedingungsteil (die Prämisse) jeder Regel aus der Wissensbasis mit den technischen Spezifikationen der Ausrüstungen verglichen. Im Fall einer Übereinstimmung ist der Folgerungsteil (die Konklusion) der Regel eine für diese Ausrüstung zu erfüllende Anforderung. Dieses Verfahren produziert bis zu einige hundert Anforderungen für eine zu planende Anlage.
Abbildung 5 erläutert das Prinzip der Inferenz von Platzierungsanforderungen an einem Beispiel. Gegeben ist eine Pumpe `P-1310` sowie die Prämissen und Konklusionen der für diesen Apparat zutreffenden Equipment- und Lageanforderungen. Durch Anwendung der Equipmentregel `E313` ("Falls das Equipment eine Pumpe ist, muss es bedient und gewartet werden und es wird ein Gabelhubwagen zur Wartung benötigt") wird zunächst abgeleitet, dass für die Pumpe Bedienung und Wartung, sowie ein Gabelhubwagen zur Wartung notwendig ist. Die Anwendung der Lageregeln `L005` ("Falls ein Gabelhubwagen zur Wartung benötigt wird, *soll* das Equipment *neben dem Weg* platziert werden") sowie `L034` ("Falls eine Bedienung erfolgt, *sollte* das Equipment *neben dem Weg* platziert werden") ergibt die Anforderungen, dass die Pumpe neben dem Weg positioniert werden *soll* bzw. *sollte*. Die Lageregel `L100` ("Falls das Equipment eine Pumpe ist, *muss* dieses *im Erdgeschoß* platziert werden") ergibt zusätzlich, dass die Pumpe `P-1310` im Erdgeschoß platziert werden *muss*.

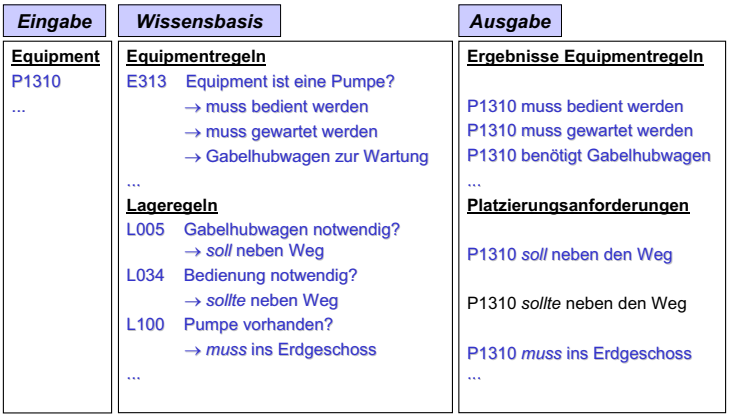


Abbildung 5. Inferenz der Platzierungsanforderungen auf Grundlage der Platzierungsregeln der Wissensbasis und der zu platzierenden Apparate

Den Anforderungen werden dabei durch die Regeln unterschiedliche Gewichtungen zugeordnet. Eine Anforderung "*muss*", "*soll*" oder "*sollte*" erfüllt sein. Falls, wie im obrigen Beispiel, gleiche Anforderungen mit unterschiedlichen Gewichtungen abgeleitet werden, erfolgt keine Addition der Gewichtungen, sondern es wird stets die stärkere Gewichtung (in diesem Fall *soll*) übernommen. Um diese Anforderungen für das nachgeschaltete numerische Optimierungsverfahren, den *SA*-Algorithmus, verwenden zu können werden sie in eine Matrix umgewandelt. Darin werden die Gewichtungen als Zahlen abgelegt, die die Stärke der Anziehung zwischen den beteiligten Objekten repräsentiert. Stellt man sich vor, die beiden Objekte seien mit einer mechanischen Feder verbunden, dann entspricht dieser Zahlenwert der Federkonstanten. Je größer der Abstand der Objekte, um so länger ist die Feder und die anziehende Kraft nimmt zu. Wären alle anziehenden Kräfte befriedigt, indem die entsprechenden Objekte auf derselben Position lägen, so wäre die Summe aller Kräfte gleich Null und der energieärmste und somit optimale Zustand erreicht. Da dies schon aufgrund der Geometrie der Ausrüstungen nicht möglich ist, aber auch konkurrierende Anforderungen ein nebeneinander liegen aller Ausrüstungen verhindern, wird eine Minimierung der Summe aller Kräfte angestrebt. Dabei ist zu bedenken, dass lediglich hinsichtlich der zuvor formulierten Anforderungen und entsprechenden Kräfte optimiert werden kann.

Eine Lösungsmethode für derartige Probleme, die sich durch zahlreiche lokale Minima auszeichnen, ist das *Simulated Annealing* [Aar92]. Hierbei hat sich gezeigt, dass die "Abkühlrate" beim *Simulated Annealing* durch empirische Parameter gesteuert werden muss, um Lösungen in vertretbarer Zeit zu erhalten. Gleichzeitig ist festzustellen, dass die Lösungen nur begrenzt reproduzierbar sind. Da bislang alle Equipments einer Anlage einzeln und gleichberechtigt berücksichtigt werden, steigt die Rechenzeit des Systems exponentiell mit

der Anzahl der Equipments. Sie begrenzt damit die Anlagengröße (auf ca. 100 Komponenten) und schränkt die Möglichkeit optimierende Variantenrechnungen durchzuführen ein. Die wichtigsten Schritte des *CAPD*-Platzierungsschemas sind in der Abbildung 6 noch einmal zusammenfassend dargestellt.

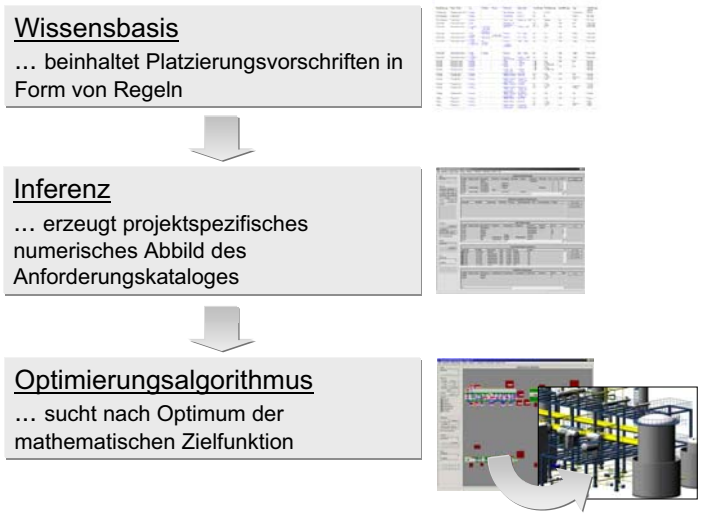


Abbildung 6. Vereinfachtes Ablaufschema des *CAPD*-Platzierungsansatzes

Nach der Platzierung der Ausrüstungen im Stahlbau werden im nächsten Modul die ungefähren Längen der verbindenden Rohrleitungen ermittelt. Dazu wird ein sogenannter Manhattan-Router eingesetzt, der die Ausrüstungen auf kürzestem Wege miteinander verbindet, dabei aber jegliche Hindernisse ignoriert. Genauere Rohrleitungsentwürfe erhält man durch den Einsatz eines Detailrouters, der alle Hindernisse umgeht, damit aber auch sehr rechenaufwendig ist.
Das letzte Modul (siehe Abbildung 2) ist das Bewertungsmodul. Dieses ermöglicht die Analyse des Konstruktionsentwurfes und kann dem Konstrukteur Hinweise geben, welche Equipments verschoben werden sollten, um die Rohrleitungskonstruktion zu verbessern oder welche Rohrleitungen durch Equipments bzw. andere Rohrleitungen behindert werden und wo Änderungen aus Kostengründen zweckmäßig sind.

3 Numerisch basiertes Lernen von Platzierungsregeln

Der in Abschnitt 2.2 beschriebene Ansatz des Planungswerkzeuges *CAPD* ist ein reiner Konstruktionsansatz mit dem Ziel der Erzeugung eines möglichst guten Aufstellungsvorschlages auf der Grundlage einer gegebenen Regelbasis und einer

Equipment- und Rohrleitungsliste. Dieser Prozess muss jedoch für jede zu konstruierende Anlage erneut angestoßen werden. Informationen im Sinne von neuen Platzierungsregeln können mit dem System naturgemäß nicht abgeleitet werden. An dieser Stelle setzt nun der numerisch basierte Lernansatz mit einer induktiven Herleitung von Informationen über die Relevanz einzelner repräsentierter Aufstellungseigenschaften (wie z.B. Nachbarschaftsrelationen zwischen einzelnen Equipments oder zwischen Equipments und dem Anlagengerüst) an. Vereinfachend formuliert soll dieser Ansatz zur Beantwortung der Frage beitragen, was gute von schlechten Aufstellungen unterscheidet.
In diesem Zusammenhang wurden zwei Anlagen betrachtet, eine stark abstrahierte Anlage aus der Dissertation von P. Leuders [Leu02] sowie eine reale Anlage zur Gasbehandlung. Von diesen Anlagen wurden zur Untersuchung des Lernansatzes zahlreiche Anlagenvariationen erzeugt, d.h. unterschiedliche Platzierungen der Equipments im Anlagengerüst. Bevor die Anlagenvariationen als Eingabe für einen induktiven Lernansatz verwendet werden konnten, wurden diese um verschiedene Bewertungsmaße ergänzt.
Dabei war das erste Ziel des Lernansatzes, auf der Grundlage der Trainingsdaten ein Modell mit einer möglichst hohen Vorhersagegenauigkeit auf den gegebenen Testdaten zu erzeugen. Als erstes Kriterium wurde dazu der Erfüllungsgrad der Platzierungsanforderungen einer konkreten Aufstellung gewählt. Dafür wurden Szenarien mit verschiedenen prozentualen Anteilen zu erfüllender *Muss*- und *Soll*-Anforderungen untersucht. Das bedeutet, dass eine konkrete Aufstellung, je nach gegebenem Szenario, einen entsprechenden Anteil an *Muss*- oder *Soll*-Anforderungen erfüllen muss, um als gute Aufstellung (entspricht einem positiven Beispiel) gewertet werden zu können. Andernfalls wurde diese als schlechte Aufstellung im Sinne des untersuchten Szenarios eingestuft (entspricht einem negativen Beispiel). Als weiteres Kriterium für die Güte des gelernten Modells wurde die Länge der aus einer Aufstellung resultierenden Rohrleitungsführung betrachtet.
Die zweite Zielsetzung betraf die Optimierung der ursprünglichen Anlagenrepräsentation in Bezug auf die gegebenen Lernverfahren, Evaluationskriterien und Datensätze. Hierbei sollten Informationen darüber gewonnen werden, welche Merkmale für die Beschreibung der untersuchten Anlagen relevant bzw. welche irrelevant sind. Der Hintergrund ist, dass beim induktiven maschinellen Lernen die Wahl geeigneter Merkmale einen wichtigen Einfluß auf den Lernerfolg hat und die Selektion von Merkmalen zur Erleichterung der Lernaufgabe und damit zur Erhöhung der Performanz führen kann.

3.1 Anlage I

Die erste untersuchte Anlage ist eine Abstraktion einer Chemieanlage, die jedoch bereits einige wichtige Eigenschaften realer Anlagen erfüllt (siehe Tabelle 1). Im Gegensatz zu realen Anlagen sind in diesem Modell keine Equipmentklassen wie z.B. *Pumpen*, *Kolonnen*, etc. vorhanden. Es wird lediglich zwischen Equipments verschiedener Größe unterschieden. Dabei belegen von den 30 Equipments 24 jeweils nur ein Feld, vier umfassen zwei nebeneinander liegende Felder und jeweils

Eigenschaften der Anlage	
Anzahl der Equipments	30
Für die Equipments zur Verfügung stehende Felder	44
Anzahl der Anforderungen	44 *Muss*-Anforderungen
	28 *Soll*-Anforderungen
Anzahl der Etagen	3
Equipmentfixierungen in (x,y,z)	ja
Equipmentfixierungen in z	ja
Rotationsfixierungen	ja
Anziehung in z	ja
Platzierung außerhalb des Stahlbaus	ja
Anziehung von Equipments an den Anlagenrand	ja
Verwendung von unterschiedlich großen Equipments	ja

Tabelle 1. Überblick über die Eigenschaften der Anlage I

ein Equipment belegt zwei bzw. drei übereinander liegende Felder. Des Weiteren ist die Platzierung der Equipments nur auf Feldern eines vorgegebenen Rasters, und nicht an beliebiger Stelle, erlaubt.
Die Anlage (siehe Abbildung 7) besteht aus drei Stockwerken (Etage 0 bis Etage 2), wobei jedes Stockwerk 36 Felder umfaßt. Von diesen 36 Feldern stehen allerdings in der ersten Etage nur 12 und in den beiden anderen Etagen nur 16 für die Equipments zur Verfügung. Auf den Feldern 14 bis 17 im Erdgeschoß (Etage 0) befindet sich ein Anlagenweg - die entsprechenden Felder sind für die Platzierung gesperrt. Die Randfelder der Etagen sind für die korrekte Platzierung der Equipments, die zwei nebeneinander liegende Felder belegen, notwendig. Dabei sind die Felder 31 bis 36 und 61 bis 66 Randfelder für jeweils zwei Etagen. Das bedeutet, dass für die 30 vorhandenen Equipments insgesamt 44 Felder zur Verfügung stehen. Die restlichen Felder werden nicht mit Equipments besetzt.
Auf der linken Seite der Abbildung ist die vollständige Positionsmatrix abgebildet. Die Felder im Anlagengerüst (Stahlbau) sind hellgrau unterlegt. Die Felder, die belegt werden können, sind durch fette schwarze Ziffern, die gesperrten Felder durch dünne blaue Ziffern gekennzeichnet.
Auf der rechten Seite der Abbildung ist die optimale Equipmentbelegung nach Etagen aufgeschlüsselt dargestellt. Dabei belegen die Equipments in diesem Fall alle Felder innerhalb des Anlagengerüstes. Das dreistöckige Equipment steht auf Feld 27 außerhalb des Anlagengerüstes und belegt aufgrund seiner Höhe auch über diesem Feld liegende Felder. Entsprechendes gilt für das zweistöckige Equipment auf Feld 28. Neben diesen beiden Equipments, die sich in z-Richtung über mehr als eine Etage ausdehnen, gibt es vier Equipments, die sich innerhalb einer Etage in der Breite (x-Richtung) bzw. der Länge (y-Richtung) über

mehr als ein Feld ausdehnen. Die Schraffierungen auf den jeweiligen Feldern geben an, ob ein Equipment an einer bestimmten Position platziert werden muss (*Fixiert in (x,y,z)*), die Rotation eines Equipments festgelegt ist, eine zulässige Etage vorgegeben wird (*Fixiert in (z)*) bzw. eine Abhängigkeit von einem zweiten Equipment in z-Richtung besteht (*Anziehung in (z)*). Zu beachten ist dabei, dass für ein Equipment zwei Arten von Fixierungen vorgegeben sind und dass bei den beiden mehrstöckigen Equipments die Fixierung auf allen Etagen gekennzeichnet ist.

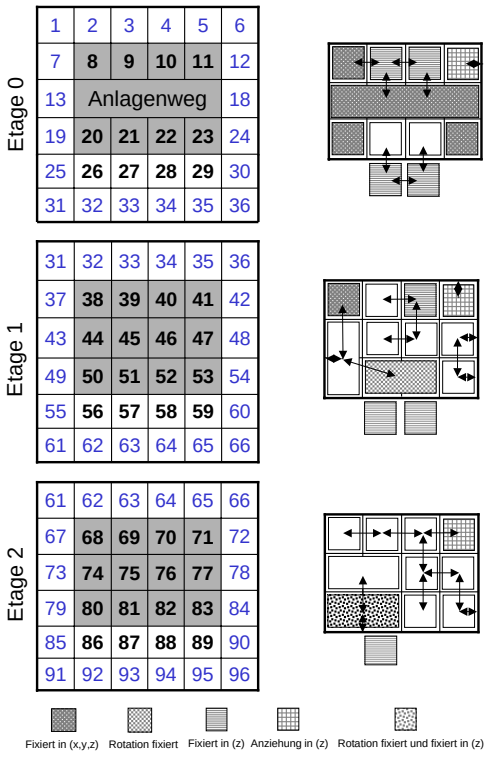


Abbildung 7. Visualierung der Positionsmatrix der Anlage I

Alle Equipments die keiner dieser Kategorien zugeordnet sind (in Abbildung 7 weiß gekennzeichnet), können jeweils auf freien Feldern innerhalb des Baufeldes positioniert werden. Die beschriebenen Fixierungen entsprechen 14 der

44 *Muss*-Anforderungen, die bei jeder Anlagenplatzierung erfüllt sein müssen. Hinzu kommt für jedes der 30 Equipments die Vorgabe, dass es innerhalb (Equipment 1 - 28) bzw. außerhalb (Equipment 29 und 30) des Stahlbaus zu stehen hat. Die 28 *Soll*-Anforderungen sind in der Abbildung 7 durch Pfeile zwischen den Equipments bzw. zwischen Equipments und dem Anlagenrand angedeutet. So *soll* z.B. im Erdgeschoß das Equipment auf der Position 27 neben dem Equipment auf Position 21 platziert werden und das Equipment auf Position 53 *soll* am östlichen Anlagenrand stehen.
Die in [Leu02] verwendete und in Abbildung 7 gezeigte Anlage ist optimal im Sinne einer minimalen Anzahl verletzter Platzierungsanforderungen für eine gegebene Menge von Equipments und einem vorgegebenem Baufeld. Mit Hilfe eines Zufallsgenerators erzeugte Varianten dieser Anlage entsprechen einzelnen Beispielen und dienen als Grundlage für den nachfolgend beschriebenen Lernansatz. Das Labeln der Beispiele geschieht dabei durch die Evaluierung der Varianten anhand unterschiedlicher Kriterien (Erfüllungsgrad der *Muss*- und *Soll*-Anforderungen bzw. resultierende Rohrleitungslänge der Anlagenplatzierung).
In den folgenden Abschnitten wird zunächst eine Beispielrepräsentation der Anlage vorgestellt, die eine relationale Beschreibung der Lagebeziehungen einzelner Equipments untereinander, sowie einzelner Equipments zum Anlagengerüst enthält. Auf der Grundlage dieser Repräsentation werden unterschiedliche Lernverfahren zur Ermittlung der relevanten Beziehungen untersucht. Zur Einordnung der mit der relationalen Anlagenbeschreibung erzielten Ergebnisse wird ein Vergleich mit einer Repräsentation durchgeführt, die in kodierter Form eine direkte Abbildung von Equipments auf einzelne Anlagenpositionen enthält.

Der maschinelle Lernansatz

Der nachfolgend beschriebene numerische Lernansatz basiert auf der Lernaufgabe der Merkmalsauswahl [LM98]. Das Ziel hierbei ist, aus einer vorgebenen Menge von Merkmalen einer gegebenen Repräsentation (entspricht hier den Eigenschaften einer Anlagenplatzierung) die jeweils relevanten Merkmale (Eigenschaften) zu bestimmen. Methodisch basieren alle untersuchten Lernansätze auf dem sogenannten *Wrapper*-Ansatz [KJ97]. Bei diesem Ansatz (siehe Abbildung 8) wird die Suche nach einer optimalen Merkmalsmenge durch die Performanz eines vorgegebenen Induktionsverfahrens (hier der *SVM*) gesteuert. Mit diesem Ansatz erreicht man aufgrund der Verwendung des induktiven Bias des gegebenen Lernverfahrens für die Steuerung des Suchprozesses i.d.R. eine hohe Generalisierungsfähigkeit. Allerdings ist diese jedoch verfahrensbedingt mit einer hohen Laufzeit, d.h. mit einer großen Anzahl von Performanzberechnungen verbunden. Letztendlich ausschlaggebend für den Erfolg der Merkmalsauswahl ist, wie effektiv das jeweilige Auswahlverfahren den vorliegenden Raum aller Merkmalskombinationen (entspricht der Potenzmenge über der Menge der Merkmale) durchsucht.
Vor diesem Hintergrund wurde ein reiner kreuzvalidierter Klassifikations- bzw. Regressionsansatz (auf der Grundlage der ursprünglichen, unveränderten Repräsentation) mit unterschiedlichen Merkmalsauswahlverfahren verglichen, dar-

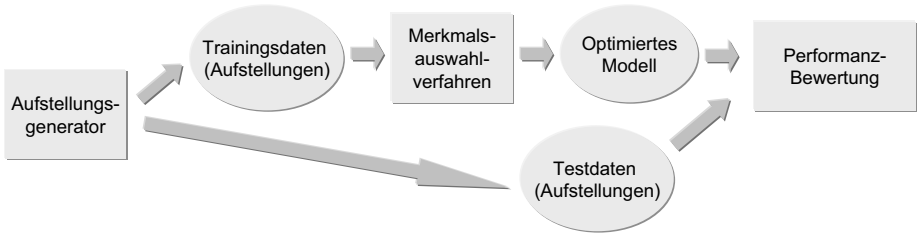


Abbildung 8. Ablaufschema des numerischen Lernansatzes

unter ein *genetischer Algorithmus* (*GA*) und zwei sequentielle Verfahren (*Forward Selection* und *Backward Elimination*). Bei allen untersuchten Auswahlverfahren wird die Merkmalsmenge als binärer Vektor aufgefaßt, in dem einzelne Merkmale durch Selektion zu der betrachteten Merkmalsmenge hinzugefügt bzw. durch Deselektion aus dieser entfernt werden können.

Im Fall des *genetischen Algorithmus* zur Merkmalsauswahl entspricht die Population von Individuen einer Menge von Merkmalsmengen. Der hier verwendete modifizierte *GA* [RK03] optimiert nun diese Menge von Merkmalsmengen mit Hilfe der Operatoren *Crossover* (entspricht dem Austausch von Teilmengen zweier Merkmalsmengen) und *Mutation* (entspricht der Selektion bzw. Deselektion einzelner Merkmale aus einer Merkmalsmenge). Die Fitnessfunktion eines Individuums entspricht dabei der Performanz des induktiven Lernverfahrens auf der betrachteten Merkmalsmenge. Das Ziel dieses Ansatzes ist dabei, die positiven Sucheigenschaften konventioneller genetischer Algorithmen für die Suche im Merkmalsraum zu nutzen[3].

Sequentielle Verfahren [AB96] durchsuchen den Merkmalsraum ebenenweise (d.h. zunächst einelementige, dann zweielementige Merkmalsmengen usw.) durch die Hinzunahme bzw. Entfernung des Merkmals mit der jeweils höchsten Performanzsteigerung innerhalb einer Ebene.

In den Fällen, in denen eine geringe Anzahl von Merkmalsblöcken gegeben war wurde zusätzlich eine vollständige Suche (sogenanntes *Brute-Force-Verfahren*) im Merkmalsraum durchgeführt. Dieses Verfahren ist aufgrund seiner exponentiellen Laufzeit zwar nur bei kleinen Merkmalsmengen anwendbar, bietet aber durch die vollständige Betrachtung aller möglichen Merkmalskombinationen eine exakte Beurteilung der Performanz einzelner Lernansätze. Insbesondere kann ermittelt werden, ob ein Verfahren ein globales Optimum (also eine optimale Merkmalsmenge) findet bzw. wie weit die gefundene Merkmalsmenge vom Optimum entfernt ist. Ebenso können mit Hilfe der vollständigen Suche Eigenschaften des Bewertungsverbandes (also der durch die Auftragung der Performanz über alle Merkmalsmengen gebildeten Struktur) wie Monotonie bzw. Nichtmonotonie abgeleitet werden.

[3] Eine ausfürliche Beschreibung *GA*-basierter Auswahlverfahren wird in [Fre02], Kapitel 9.1 gegeben.

Monotonie: Ein Verband wird als monoton bezeichnet, falls die Hinzunahme eines Merkmals zu einer bestehenden Merkmalsmenge zu einer Verbesserung der Performanz führt.

Eine Auswirkung der Nichtmonotonie innerhalb eines Bewertungsverbandes ist bei sogenannten Hill-Climbing-Verfahren wie *Forward Selection* und *Backward Elimination* eine (vorzeitige) Konvergenz auf lokalen Optima. *Genetische Algorithmen* sind hingegen durch den globalen Suchansatz in der Lage solche lokalen Optima zu verlassen.

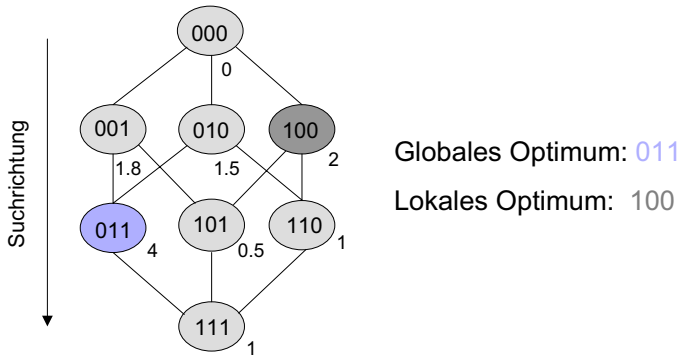


Abbildung 9. Beispiel für einen nichtmonotonen Bewertungsverband

Ein Beispiel für einen trivialen nichtmonotonen Merkmalsverband mit drei Merkmalen ist in Abbildung 9 gegeben. Das erste Merkmal (Knoten "100") ist ein lokales Optimum innerhalb des Verbandes, die Merkmalsmenge bestehend aus dem zweiten und dritten Merkmal (Knoten "011") ist das globale Optimum. Anschaulich läßt sich das nichtmonotone Verhalten beim Übergang vom Knoten "100" zu den Knoten "101" und "110" erkennen (die Hinzunahme eines Merkmals führt zu einer Verschlechterung der Performanz). Die Folge einer solchen Nichtmonotonie ist bei lokalen Verfahren (wie *Forward Selection* und *Backward Elimination*), dass diese das einmal erreichte lokale Optimum nicht mehr verlassen können.

Erzeugung der Beispieldaten

Um die Ermittlung relevanter Platzierungsinformationen mit Hilfe maschineller Lernverfahren durchführen zu können, mußten zunächst geeignete Beispieldaten erzeugt werden. Diese Beispieldaten, d.h. die bewerteten Aufstellungen wurden im Fall der relationalen Repräsentation mit einem am Lehrstuhl für Anlagentechnik entwickelten Generator erzeugt. Dabei wurde die jeweilige Anforderung

des gegebenen Szenarios als Entscheidungskriterium für die Güte einer Aufstellung gewertet. So wurde eine Aufstellung im Szenario *Muss10%Soll* als positives Beispiel gewertet (und entspricht somit einer guten Aufstellung), falls es alle *Muss*- sowie mindestens 10% der *Soll*-Anforderungen erfüllt. Falls es diese beiden Kriterien nicht erfüllt, wurde es als negatives Beispiel deklariert[4]. Der Generator ermöglicht die Ausgabe einer optimalen Anlage, einer Anlage mit zufällig platzierten Equipments und einer Anlage, bei der die Equipments so platziert werden, dass die *Muss*-Anforderungen auf jeden Fall erfüllt sind, ansonsten aber zufällig platziert wird.

Optimierung der Lernparameter

Nachdem die Beispiele mit Hilfe des Generators erzeugt wurden, erfolgt vor dem eigentlichen Lernschritt noch eine Optimierung der Lernparameter der *Support Vector Machine* (Parameter C und *epsilon*) sowie der *Größe des verwendeten Samples*. Dieser Schritt dient der Komplexitätsreduktion der Lernaufgabe, da für alle nachfolgend untersuchten Lernansätze die bei der Parameteroptimierung ermittelten Werte übernommen werden können.
In Tabelle 2 ist für jedes der untersuchten Szenarien (*Muss0%Soll*, *Muss10%Soll*, *Muss25%Soll* und *Muss5%Soll*) der jeweils optimale Wert für die SVM-Parameter C und *epsilon* sowie die *Größe des Samples* eingetragen. In Fällen, in denen zwei Parameterwerte eine identische Performanz aufweisen, sind beide Werte aufgeführt. Gilt für alle Werte die gleiche Performanz, wird dies in der Tabelle durch das '*'-Zeichen angedeutet.

Parameter	$Muss0\%Soll$	$Muss10\%Soll$	$Muss25\%Soll$	$Muss50\%Soll$
C	*	*	*	*
epsilon	*	{0.1,0.01}	{0.1,0.01}	{0.1,0.01}
Größe des Samples	500	500	500	500

Tabelle 2. Parameteroptimierungsmatrix der vier Klassifikationsszenarien (*Muss0%Soll*, *Muss10%Soll*, *Muss25%Soll* und *Muss50%Soll*) für die relationale Repräsentation

[4] Es ist im Zusammenhang mit der Bewertung eines Beispieles nicht von Bedeutung, **welche** der *Soll*-Anforderungen erfüllt sind, sondern lediglich **wie hoch der Prozentsatz** der erfüllten *Soll*-Anforderungen **ist**.

Relationale Repräsentation

Nachdem der prinzipielle Experimentaufbau beschrieben und die Bedeutung der Repräsentation als Grundlage für die Ermittlng relevanter Aufstellungseigenschaften verdeutlicht wurden, soll nun die bei der Anlage I verwendete Merkmalsmenge vorgestellt werden. Diese besteht sowohl aus Merkmalen, die sich nur auf ein Equipment beziehen (wie z.B. *liegt in Etage* oder *hat Rotation*) als auch aus Merkmalen, die sich auf eine Relation zwischen zwei Equipments bzw. zwischen einem Equipment und der Anlage beziehen (wie z.B. *liegt neben* oder *liegt am Anlagenweg*).
Neben der getrennten Bertrachtung einzelner Merkmale soll an dieser Stelle die Verwendung von Merkmalsblöcken vorgestellt werden. Merkmalsblöcke (MB) sind Gruppierungen von Merkmalen, die vergleichbare Informationen enthalten. So kann man bei der Verwendung von merkmalsbasierten Lern- und Optimierungsverfahren von einzelnen Merkmalen (z.B. "$Equipment_i$ liegt neben $Equipment_j$") abstrahieren und Merkmale auf einer Meta-Ebene betrachten (die Relation *liegt neben* mit sämtlichen Equipmentkombinationen). Durch diese Abstraktion gewinnt man zum einen Informationen, die man durch die getrennte Betrachtung einzelner Merkmale nicht erzielen würde (z.B. ob die gesamte Relation *liegt neben* eine für die gegebene Lernaufgabe relevante Eigenschaft darstellt). Zum anderen wird die Komplexität des Suchraumes erheblich reduziert und die Lernaufgabe somit wesentlich vereinfacht. Im Fall der nachfolgend aufgeführten relationalen Repräsentation erfolgt beispielsweise eine Reduktion von ursprünglich 990 einzelnen Merkmalen auf sechs Merkmalsblöcke.

Merkmalsblock	Wertebereich der Merkmale	Anzahl der Merkmale je Merkmalsblock
Equipment A_i liegt neben Equipment A_j	{0,1}	435
Equipment A_i liegt über Equipment A_j	{0,1}	435
Equipment A_i liegt außerhalb des Stahlbaus	{0,1}	30
Equipment A_i liegt am Anlagenweg	{0,1}	30
Equipment A_i liegt in Etage	{0,1,2}	30
Equipment A_i hat Rotation	{0,..,16}	30

Tabelle 3. Relationale Repräsentation der Anlage I

Tabelle 3 zeigt nun die vollständige Menge der in dieser Repräsentation verwendeten Merkmalsblöcke, die Wertebereiche der enthaltenen Merkmale sowie die Anzahl der Merkmalskombinationen innerhalb eines Merkmalsblocks.
Bei den relationalen Merkmalen *liegt neben* und *liegt über* bedeutet ein Merkmalswert "1", dass die entsprechende Relation für die gegebene Equipmentkombination erfüllt bzw. ein Wert von "0", dass diese nicht erfüllt ist. Ist also in einer konkreten Aufstellung beispielsweise $Equipment_1$ neben $Equipment_2$ platziert, so wird dem entsprechenden Merkmal "$Equipment_1$ liegt neben $Equipment_2$" der Wert "1" zugewiesen. Das Merkmal *liegt neben* ist für zwei Equipments

dann erfüllt, wenn sie benachbarte Felder belegen. Als benachbart zu einem Feld zählen die vier direkt an das Feld angrenzenden Felder. Liegt ein Equipment auf dem Feld direkt über einem Feld oder auf einem dem oberen Feld benachbarten Feld, so *liegt* es *über* einem auf dem unteren Feld liegenden Equipment. Die Anzahl von Merkmalen innerhalb eines Merkmalsblocks beträgt $\frac{E^2-E}{2}$, wobei E der Zahl der zu platzierenden Equipments entspricht. Bei den Merkmalsblöcken *liegt außerhalb des Stahlbaus*, *liegt am Anlagenweg*, *liegt in Etage* und *hat Rotation* ist je Equipment nur jeweils ein Merkmal vorhanden. Liegt beispielsweise $Equipment_5$ **nicht** am Anlagenrand, so erhält das entsprechende Merkmal "$Equipment_5$ liegt am Anlagenrand" den Merkmalswert "0". Das Merkmal *liegt außerhalb des Stahlbaus* ist dann erfüllt, wenn ein Equipment auf einem der Felder liegt, die außerhalb des Stahlbaus (Anlagengerüstes) liegen. Ein Equipment *liegt am Anlagenweg*, wenn es eines der acht am Anlagenweg liegenden belegbaren Felder belegt. Bei den vier Equipments, die zwei nebeneinander liegende Felder belegen, ist das entsprechende Merkmal schon dann erfüllt, wenn nur eines der von dem entsprechenden Equipment belegten Felder die Bedingung erfüllt. Bei den beiden mehrstöckigen Equipments hingegen wird jeweils nur das unterste Feld betrachtet.
Bei der Rotation wird zunächst unterschieden, ob ein Equipment keinen, einen, zwei oder drei Stutzen für Rohrleitungsanschlüsse besitzt. Bei einem Equipment ohne Stutzen macht es (in diesem Fall) keinen Unterschied, welche Rotation das Equipment hat, so dass es die Rotation "0" zugewiesen bekommt. Ein Equipment mit einem Stutzen hat die Rotation "1", "2", "3" oder "4", je nachdem, ob der Stutzen des platzierten Equipments nach *oben*, *rechts*, *unten* oder *links* zeigt. Bei zwei Stutzen wird noch unterschieden, ob die Stutzen gegenüber (Rotationen "5" bis "8") oder im rechten Winkel zueinander (Rotationen "9" bis "12") liegen. Bei drei Stutzen gibt es wiederum nur vier Möglichkeiten (Rotationen "13" bis "16").

Lernergebnisse für die relationale Repräsentation

An dieser Stelle wollen wir die Ergebnisse des in den vorherigen Abschnitten beschriebenen maschinellen Lernansatzes diskutieren. Dazu wurden die Kennzahlen verschiedener Lernansätze für die relationale Repräsentation auf allen vier Szenarien miteinander verglichen. Die untersuchten Kennzahlen sind je Lernansatz die beiden Performanzkriterien *Accuracy* und *MDL*, die im Rahmen des Lernansatzes ermittelte (optimierte) Merkmalsmenge, die Anzahl der Merkmale innerhalb der optimierten Merkmalsmenge sowie die Anzahl der benötigten Performanzberechnungen[5].
Für den Vergleich wurden die nachfolgenden Lernansätze unter Verwendung der genannten Kennzahlen betrachtet. Zunächst wurde für eine untere Abschätzung der erreichbaren Performanz (ohne Optimierung der Merkmalsmenge) ein Lernlauf auf der vollständigen Merkmalsmenge durchgeführt. Anschließend wurden

[5] Zu beachten ist, dass dieser Wert die insgesamt durchgeführten Performanzberechnungen **über alle Kreuzvalidierungsläufe** angibt.

die beiden sequentiellen Merkmalsauswahlverfahren *Forward Selection* und *Backward Elimination* mit einem auf *genetischen Algorithmen* basierenden Suchverfahren verglichen. Für sämtliche mit dem *genetischen Algorithmus* durchgeführten Experimente gilt die in Tabelle 4 aufgeführte Belegung der Strategieparameter.

Parameter	*Parameterwert*
Selektionsverfahren	fitnessproportionale Selektion
Crossoververfahren	uniformes Crossover
Crossoverwahrscheinlichkeit	0.6
Mutationswahrscheinlichkeit	0.1

Tabelle 4. Strategieparameter für die Experimente mit dem *genetischen Algorithmus*

Um festzustellen, ob ein Verfahren die (global) optimale Merkmalsmenge erreicht hat, wurde ebenfalls eine vollständige Suche (*Brute-Force-Ansatz*) durchgeführt. Für alle Lernansätze wurde eine *SVM* mit linearem Kernel verwendet und die Lernläufe wurden auf der Ebene der Merkmalsblöcke durchgeführt.

Zunächst betrachten wir die Kennzahlen im Fall des Performanzmaßes *Accuracy* (Abbildung 10, 11, 12 und 13). Ohne im Detail auf die einzelnen Klassifikationszenarien einzugehen, lässt sich insgesamt eine gleichmäßig hohe Performanz über alle Lernansätze (jeweils zwischen 99.8% und 100% *Accuracy*) feststellen. Eine Ursache für dieses Ergebnis ist in der hohen Anzahl der globalen Optima, d.h. der maximal erreichbaren Performanz je Szenario, begründet. In den betrachteten Suchräumen können sich die unterschiedlichen Sucheigenschaften der Lernverfahren weniger stark auswirken. Eine Ausnahme bzgl. der Performanz bildet jedoch das Szenario *Muss25%Soll* (siehe Abbildung 12), das sich durch eine Vielzahl lokaler Optima auszeichnet. Das einzige globale Optimum dieses Szenarios (der Merkmalsblock *liegt außerhalb*) wird dabei sowohl von der *Forward Selection* als auch vom *genetischen Algorithmus* (*GA*) gefunden[6]. Als einziges Verfahren findet der *GA*, insbesondere im Fall der schwierigeren Lernaufgabe *Muss25%Soll*, in allen vier Szenarien die optimale Merkmalsmenge bei einer vergleichsweise moderaten Anzahl von Fitnessbewertungen. Der Merkmalsblock, der am häufigsten als optimal ausgewählt wurde war der relationale Block (*liegt über*) mit insgesamt 435 Merkmalen.

[6] In der Spalte "ermittelte optimale Merkmalsmenge" bedeutet eine grün unterlegte Merkmalsmenge, dass das entsprechende Suchverfahren die durch vollständige Suche ermittelte optimale Merkmalsmenge gefunden hat. Eine rot unterlegte Merkmalsmenge deutet an, dass die gefundene Merkmalsmenge nicht dem globalen Optimum entspricht.

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	100	0	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	100	0,561	435	35	liegt_neben
Merkmalsauswahl mittels Backward Elimination	100	0	990	40	alle Merkmale
Merkmalsauswahl mittels GA	100	0,561	435	150	liegt_über
Vollständige Suche (BruteForce)	100	0,561	435	64	liegt_über
Merkmalsauswahl mittels GA (ohne MB)	90,6	0,845	60	150	-

Abbildung 10. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *Accuracy* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 0% erfüllte *Soll*-Anforderungen

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	99,8	0,002	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	99,8	0,559	435	65	liegt_neben
Merkmalsauswahl mittels Backward Elimination	99,8	0,559	435	70	liegt_neben
Merkmalsauswahl mittels GA	99,8	0,559	435	150	liegt_über
Vollständige Suche (BruteForce)	99,8	0,559	435	64	liegt_über
Merkmalsauswahl mittels GA (ohne MB)	82,8	0,845	56	150	-

Abbildung 11. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *Accuracy* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 10% erfüllte *Soll*-Anforderungen

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	80,2	-0,197	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	83	0,7997	435	65	liegt_außerhalb
Merkmalsauswahl mittels Backward Elimination	80,6	-0,164	960	115	alle Merkmale außer Rotation
Merkmalsauswahl mittels GA	83	0,7997	435	535	liegt_außerhalb
Vollständige Suche (BruteForce)	83	0,7997	435	64	liegt_außerhalb

Abbildung 12. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *Accuracy* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 25% erfüllte *Soll*-Anforderungen

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	99,8	0,002	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	99,8	0,559	435	65	liegt_neben
Merkmalsauswahl mittels Backward Elimination	99,8	0,559	435	100	liegt_neben
Merkmalsauswahl mittels GA	99,8	0,559	435	150	liegt_über
Vollständige Suche (BruteForce)	99,8	0,559	435	64	liegt_über
Merkmalsauswahl mittels GA (ohne MB)	99,8	0,949	49	150	-

Abbildung 13. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *Accuracy* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 50% erfüllte *Soll*-Anforderungen

Ein Vergleich der *GA*-basierten Merkmalsauswahl mit und ohne Verwendung von Merkmalsblöcken ergibt, dass im letzeren Fall bei gleicher Laufzeit eine deutlich geringere Performanz erzielt wird.
Neben dem Performanzmaß *Accuracy* wurde bei der relationalen Repräsentation ein weiteres Maß betrachtet, dass nicht nur die Generalisierungsfähigkeit als Kriterium verwendet, sondern zusätzlich auch die Größe der ermittelten Merkmalsmenge miteinbezieht. Dieses an das Prinzip der minimalen Beschreibungslänge (siehe z.B. [Ris78,Che90]) angelehnte Maß bevorzugt diejenigen Merkmalsmengen, die eine hohe Performanz mit einer geringen Anzahl von Merkmalen erzielen. Die konkrete Formulierung im Fall eines Klassifikationsszenarios lautet wie folgt:

$$MDL(Y) := Accuracy(Y) - k \cdot \frac{\#Merkmale(Y)}{\#Merkmale(X)} \tag{1}$$

Dabei repräsentiert X die vollständige Merkmalsmenge, Y eine betrachtete Teilmenge aus X, $Accuracy(Y)$ die Vorhersagegenauigkeit auf der durch Y gegebenen Beispielmenge und k einen Koeffizient, der die Gewichtung von Komplexität (der Merkmalsbeschreibung) und Performanz festlegt. Der Wertebereich des *MDL*-Maßes umfaßt die Menge der reellen Zahlen im Intervall "-1" bis "1". Die mit dem *MDL*-Performanzmaß durchgeführten Experimente (Abbildung 14, 15, 16 und 17) zeigten ein ähnliches Verhalten der untersuchten Lernansätze wie bei dem *Accuracy*-Maß.
Ein Unterschied ist die deutliche Performanzdifferenz zwischen der ursprünglichen und der optimalen Merkmalsmenge. Ansonsten ist das Verhalten der Selektionsverfahren mit dem der letzten Experimentreihe vergleichbar. Wie bereits bei der Experimentreihe mit dem Bewertungsmaß *Accuracy* ist auch hier der *genetische Algorithmus* das einzige Verfahren, dass in allen Szenarien jeweils die optimale Merkmalsmenge erreicht. Allerdings ist die Zahl der vom *GA* benötigten Fitness- bzw. Performanzberechnungen im Vergleich zur letzten Experimentreihe erhöht (statt durchschnittlich ca. 240 sind es nun 970 Performanzbewertungen), wohingegen sich die Zahl der Berechnungen bei den sequentiellen Suchverfahren kaum verändert hat. Der Merkmalsblock, der in den meisten Szenarien als optimal bewertet wurde, ist der einstellige Block (*liegt am Weg*) mit insgesamt 30 Merkmalen.
Ein Vergleich der *GA*-basierten Merkmalsauswahl mit und ohne Verwendung von Merkmalsblöcken ergibt, wie bei dem zuletzt betrachteten Bewertungsmaß *Accuracy*, dass im letzteren Fall bei gleicher Laufzeit eine deutlich geringere Performanz erzielt wird.
Eine Erkenntnis die aus dieser Experimentreihe gewonnen werden konnte ist, dass die Merkmalsauswahl mit dem *MDL*-Maß zu einer vergleichbaren Generalisierungsfähigkeit bei gleichzeitig deutlich geringerer Merkmalsanzahl (verglichen mit dem *Accuracy*-Maß) gelangt.

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	100	0	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	100	0,9697	30	55	liegt_ausserhalb
Merkmalsauswahl mittels Backward Elimination	100	0	990	40	alle Merkmale
Merkmalsauswahl mittels GA	100	0,9697	30	980	liegt_am_Weg
Vollständige Suche (BruteForce)	100	0,9697	30	64	liegt_am_Weg
Merkmalsauswahl mittels GA (ohne MB)	95	0,899	50	900	-

Abbildung 14. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *MDL* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 0% erfüllte *Soll*-Anforderungen

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	99,8	-0,002	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	99	0,9597	30	55	liegt_außerhalb
Merkmalsauswahl mittels Backward Elimination	99	0,9597	30	105	liegt_außerhalb
Merkmalsauswahl mittels GA	99,8	0,9677	30	995	liegt_am_Weg
Merkmalsauswahl mittels GA (ohne MB)	94	0,883	56	900	-

Abbildung 15. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *MDL* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 10% erfüllte *Soll*-Anforderungen

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	81	-0,19	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	83	0,7997	30	55	liegt_außerhalb
Merkmalsauswahl mittels Backward Elimination	80	0,7394	60	115	liegt_in_Etage u. Rotation
Merkmalsauswahl mittels GA	83	0,7997	30	980	liegt_am_Weg
Vollständige Suche (BruteForce)	83	0,7997	30	64	liegt_am_Weg
Merkmalsauswahl mittels GA (ohne MB)	76	0,703	56	900	-

Abbildung 16. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *MDL* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 25% erfüllte *Soll*-Anforderungen

	Accuracy	MDL	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen	ermittelte optimale Merkmalsmenge
Ohne Selektion	100	0	990	5	alle Merkmale
Merkmalsauswahl mittels Forward Selection	100	0,9697	30	55	liegt_ausserhalb
Merkmalsauswahl mittels Backward Elimination	100	0,9697	30	105	liegt_am_Weg
Merkmalsauswahl mittels GA	100	0,9697	30	920	Rotation
Vollständige Suche (BruteForce)	100	0,9697	30	64	Rotation
Merkmalsauswahl mittels GA (ohne MB)	96	0,9226	37	900	-

Abbildung 17. Vergleich verschiedener Lernansätze unter Verwendung des Bewertungskriteriums *MDL* für das Szenario 100% erfüllte *Muss*-Anforderungen und mindestens 50% erfüllte *Soll*-Anforderungen

Vergleich der Repräsentationen

In diesem Abschnitt wollen wir die beschriebene relationale Repräsentation mit einer Aufstellungsbeschreibung in Absolutkoordinaten vergleichen. In diesem Fall erfolgt die Kodierung einer Aufstellung in Form von absoluten Positionsangaben der Equipments innerhalb des Baufeldes. Für jedes der 30 gegebenen Equipments E_i gilt bezüglich der 96 möglichen Positionen innerhalb des Baufeldes folgende Formel:

$$Val(E_i, j) = \begin{cases} 1, \text{Equip. } E_i \text{ befindet sich an Pos. } j \\ 0, \text{Equip. } E_i \text{ befindet sich } \textbf{nicht} \text{ an Pos. } j \end{cases} \quad (2)$$

Abbildung 18 gibt eine anschauliche Darstellung dieser Repräsentation, wobei aus Gründen der Übersichtlichkeit nur die Positionskodierung für das erste und letzte Equipment abgebildet sind. Equipment 1 soll in diesem Beispiel auf dem Feld 8 platziert werden, während Equipment 30 die Felder 27, 57 und 87 belegt. Der resultierende Merkmalsvektor enthält einen Binärwert für jede Equipment-Feld-Kombination, insgesamt also 2880 $(30 * 96)$ Merkmale. Von diesen weisen jedoch nur 37 einen von Null verschiedenen Merkmalswert auf. Die beschriebene Repräsentation erzeugt somit einem hochdimensionalen, sehr spärlich besetzten Suchraum der insgesamt $\frac{96!}{(96-37)!} = \frac{96!}{59!} = 7.2 \cdot 10^{69}$ mögliche Platzierungen enthält. Für diese Art von Suchräumen hat sich die *Support Vector Machine* als besonders geeignet erwiesen [Joa01].

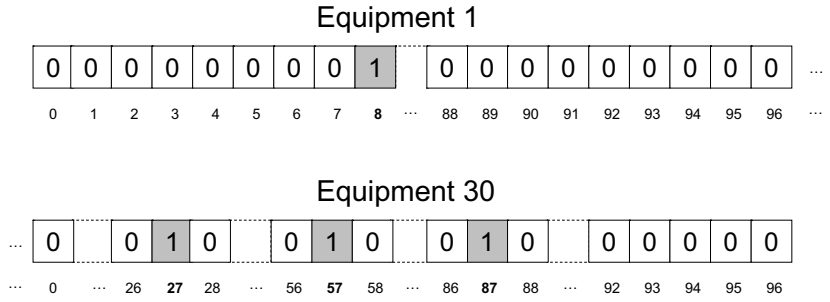


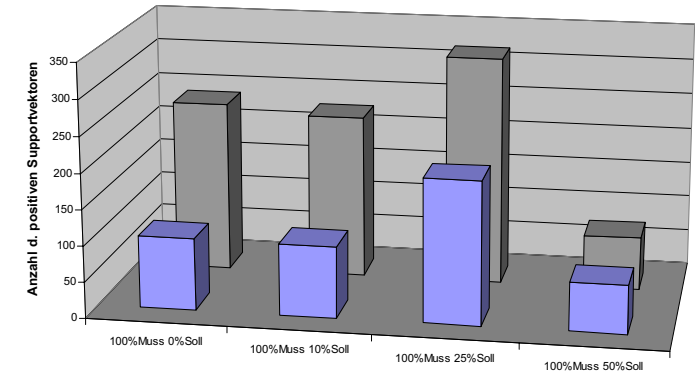
Abbildung 18. Visualisierung der Repräsentation der Anlage I in Absolutkoordinaten

Die Generierung der Beispieldaten erfolgt analog zur relationalen Repräsentation. Ebenso sind die bei dieser Repräsentation verwendeten Parameter übernommen worden. Eine Parameteroptimierung im Vorfeld der Lernläufe hat die in Tabelle 5 aufgeführten optimalen Parametereinstellungen ergeben .
Der eigentliche Vergleich der beiden Repräsentationen geschieht nun anhand der Komplexität der in dem Lernansatz (bestehend aus einem 5-fach kreuzvalidierten SVM-Lernschritt) ermittelten SVM-Modelle sowie deren resultierender Generalisierungsfähigkeit. Als Maß für die Komplexität der Modelle wurde die

Parameter	$Muss0\%Soll$	$Muss10\%Soll$	$Muss25\%Soll$	$Muss50\%Soll$
C	*	*	*	*
epsilon	{0.1,0.01}	*	{0.1,0.01}	*
Größe des Samples	500	500	500	*

Tabelle 5. Parameteroptimierungsmatrix der vier Klassifikationsszenarien (*Muss0%Soll*, *Muss10%Soll*, *Muss25%Soll* und *Muss50%Soll*) für die absolute Repräsentation

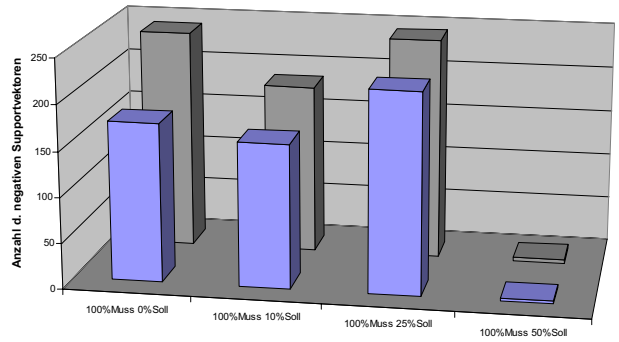
Anzahl der positiven bzw. negativen Stützvektoren (kurz SV) gewählt. Die Generalisierungsfähigkeit wurde im Fall der Klassifikationsszenarien *Muss0%Soll*, *Muss10%Soll*, *Muss25%Soll* und *Muss50%Soll* anhand der *Accuracy* gemessen.



	100%Muss 0%Soll	100%Muss 10%Soll	100%Muss 25%Soll	100%Muss 50%Soll
rel. Repräsentation	100	99	199	67
abs. Repräsentation	239	228	319	74

Abbildung 19. Vergleich der beiden Repräsentationen bzgl. der Anzahl der positiven Stützvektoren

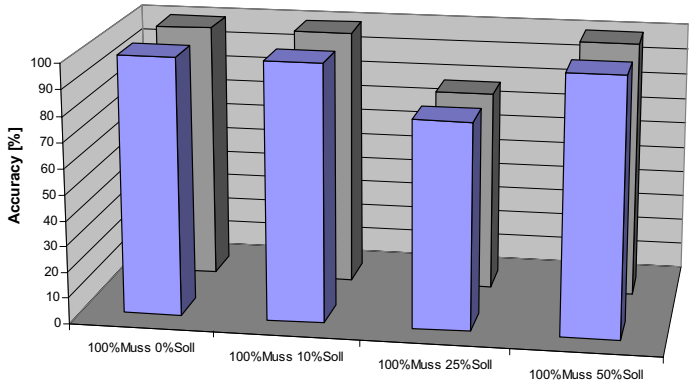
Für die Klassifikationsszenarien gilt, dass bei Verwendung der relationalen Repräsentation sowohl deutlich weniger positive (siehe Abbildung 19) als auch negative (siehe Abbildung 20) Stützvektoren erzeugt werden als bei der Repräsentation in Absolutkoordinaten. Das bedeutet, dass durch die Verwendung der relationalen Repräsentation die erzeugten SVM-Modelle eine deutlich geringere Komplexität aufweisen.
Die relationale Repräsentation, die neben der Beschreibung der Lagebeziehungen einzelner Equipments untereinander, wie z.B. *liegt neben*, auch Beziehungen einzelner Equipments zum Anlagengerüst, wie *liegt in Etage*, enthält, besitzt im



	100%Muss 0%Soll	100%Muss 10%Soll	100%Muss 25%Soll	100%Muss 50%Soll
rel. Repräsentation	174	158	220	3
abs. Repräsentation	242	186	244	4

Abbildung 20. Vergleich der beiden Repräsentationen bzgl. der Anzahl der negativen Stützvektoren

Vergleich zur absoluten Repräsentation eine gleich hohe (Szenario *Muss10%Soll*) bzw. sogar höhere (Szenarien *Muss0%Soll*, *Muss25%Soll* und *Muss5%Soll*) Generalisierungsfähigkeit. Entsprechende Ergebnisse können der Abbildung 21 entnommen werden.



	100%Muss 0%Soll	100%Muss 10%Soll	100%Muss 25%Soll	100%Muss 50%Soll
rel. Repräsentation	100	99,8	80,2	99,8
abs. Repräsentation	99,9	99,8	77,64	99,58

Abbildung 21. Vergleich der beiden Repräsentationen bzgl. der *Accuracy*

Eine Ursache dafür ist, dass die absolute im Vergleich zur relationalen Repräsentation eine ungünstigere Verteilung der Instanzen im Instanzenraum aufweist,

d.h. die Beispiele sind stärker über den Instanzenraum verstreut (sogenannte *concept dispersion*). Kleine Veränderungen einzelner Merkmalswerte führen im Fall der Klassifikation bereits zu einer Änderung des Klassenlabels (siehe Abbildung 22).

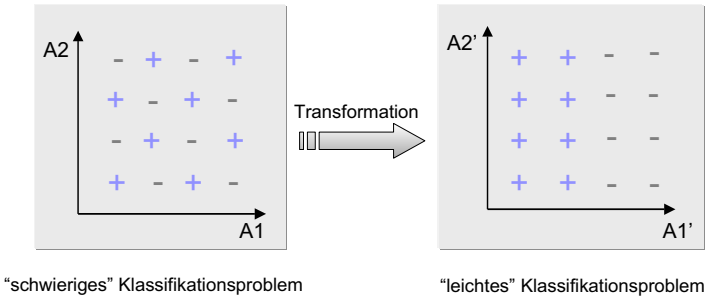


Abbildung 22. Beispiel für eine *Concept Dispersion*

Die Lernaufgabe wird also durch eine vorliegende *Concept Dispersion* [Fre01] erschwert, was in der Konsequenz zu einem komplexeren Modell des verwendeten Lernverfahrens (bei Verwendung einer *SVM* wie gesehen zu einer höheren Anzahl von Stützvektoren) und einer geringeren Generalisierungsfähigkeit führt. Die Auflösung einer solchen *Concept Dispersion* kann durch die Transformation des Instanzenraums erreicht werden.

3.2 Anlage II

Bei der zweiten untersuchten Anlage handelt es sich um eine reale Anlage zur Gasbehandlung, bei der die Platzierung der Equipments mit einer Genauigkeit von einem *mm*, d.h. ohne die Beschränkung auf diskrete Equipmentpositionen, erfolgt. Ein weiterer Unterschied ist die Verwendung von Equipmenttypen wie z.B. *Pumpen*, *Wärmetauschern* und *Kolonnen* bei den nachfolgend beschriebenen Aufstellungsrepräsentationen.
Die 28 zur Anlage gehörenden Apparate werden auf einem etwa 50 m x 50 m großen Baufeld angeordnet, wobei in der Mitte des Baufeldes ein vier Etagen umfassender Stahlbau platziert ist. Abbildung 23 zeigt ein Modell der Anlage inklusive des Baufeldes.
Die innerhalb des Baufeldes eingetragenen Anlagenwege sind in der Abbildung grau dargestellt. Diese Wege haben eine doppelte Bedeutung innerhalb des Planungsprozesses. Zum einen dienen sie als Orientierung für die Apparate, die aufgrund der Bedienbarkeit und Wartung in der Nähe eines Weges platziert werden müssen. Zum anderen können sie als Platzhalter für einen nicht zu be-

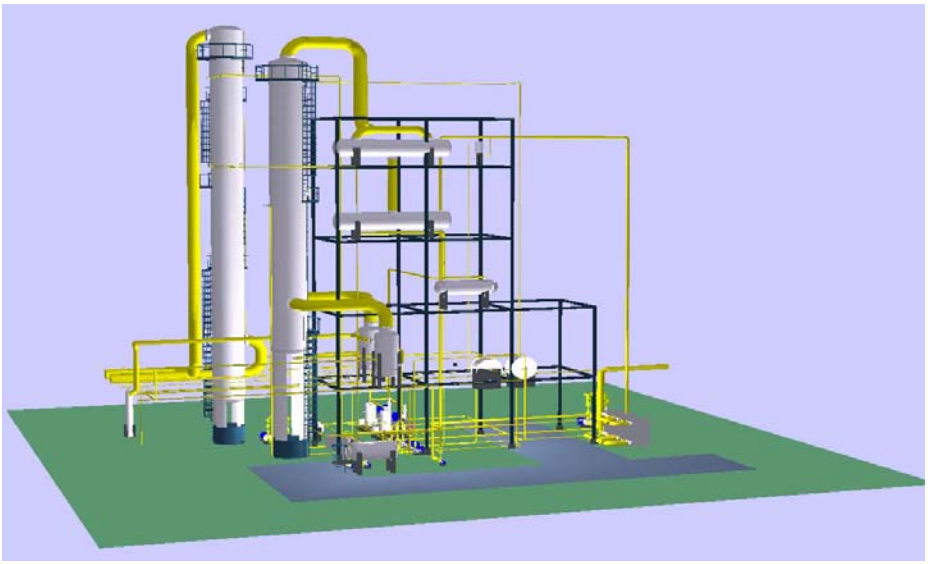


Abbildung 23. Mit dem *CAPD*-System generiertes Modell der Anlage. Zu sehen ist die realisierte Aufstellung mit dem fünfstöckigen Anlagengerüst sowie den in grau unterlegten Anlagenwegen.

bauenden Raum betrachtet werden, da Anlagenwege grundsätzlich nicht für die Platzierung von Equipments zur Verfügung stehen.

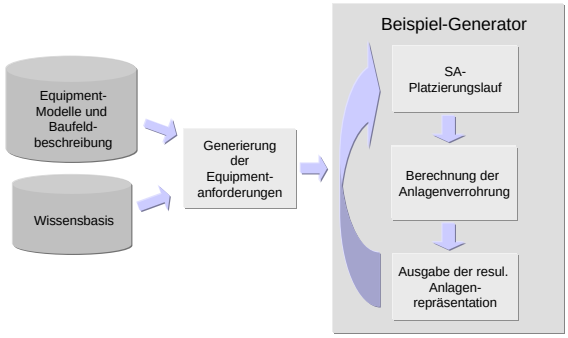


Abbildung 24. Ablaufschema zur Generierung von Beispieldaten in Form der equipmentbasierten Repräsentation für die reale Anlage

Der Ablauf der Beispielgenerierung für die Anlage II läßt sich leicht anhand der Abbildung 24 nachvollziehen. Um eine bestimmte Menge von Beispielaufstellungen, d.h. Varianten der realen Anlage, zu erzeugen wird für jedes Beispiel

ein vollständiger *SA*-Platzierungslauf angestoßen. Nach der Platzierung der 28 Equipments durch den *SA*-Algorithmus wird die Rohrleitungslänge der generierten Anlagenvariante berechnet. Aus der konkreten Aufstellung werden anschließend die benötigten Merkmale der Repräsentation berechnet. Als Label wird die resultierende Rohrleitungslänge verwendet.

Equipmentbasierte Anlagenrepräsentation

Die erste Repräsentation der realen Anlage enthält insgesamt 2030 Merkmale, verteilt auf zehn Merkmalsblöcke. Diese setzen sich aus fünf relationalen Merkmalsblöcken, die jeweils 378 Merkmale zur Beschreibung einer Equipmentkombinationen enthalten und fünf einstelligen Merkmalsblöcken mit je 28 Merkmalen (siehe Tabelle 6) zusammen.

Nr.	Name des Merkmalsblocks	Bedeutung der Merkmale	Wertebereich	Anzahl der Merkmale je Merkmalsblock
1	*DifferenzInZ*	Höhenunterschied zweier Equipments	$\mathbb{R}$	378
2	*Distanz*	Abstand zweier Equipments	$\mathbb{R}^+$	378
3	*liegt über/unter/gleich*	relative Lage zweier Equipments bzgl. der Etage	{0,1,-1}	378
4	*liegt neben*	'Nähe' zweier Equipments auf derselben Etage	[0..1]	378
5	*liegt neben (benachbarte Etagen)*	'Nähe' zweier Equipments auf benachbarten Etagen	[0..1]	378
6	*liegt auf Etage*	Etagennummer eines Equipments	{0,..,4}	28
7	*Verhältnis Länge zu Breite*	Verhältnis zwischen Länge und Breite eines Equipments	$\mathbb{R}^+$	28
8	*liegt außerhalb des Stahlbaus*	prozentuale Fläche außerhalb des Stahlbaus	[0..1]	28
9	*liegt am Weg*	Entfernung zum Anlagenweg	[0..1]	28
10	*Equipmenttyp*	Equipmenttyp	{1,..,7}	28

Tabelle 6. Equipmentbasierte Repräsentation der realen Anlage

Im Fall des Merkmalsblocks *DifferenzInZ* wird der Höhenunterschied zweier Equipments in mm angegeben. *Distanz* gibt auf der Verbindungsgeraden zwischen den Mittelpunkten der beiden Modelle den Abstand (in *mm*) zwischen den Volumenmodellen der beiden Equipments an. Bei *liegt über/unter/gleich* zeigt ein Merkmalswert "1" an, dass sich das erstgenannte Equipment oberhalb des

zweiten befindet, für den Wert “-1” gilt der umgekehrte Fall, und bei “0” befinden sich die Equipments auf derselben Etage. Der Merkmalsblock *liegt neben*, der eine Nachbarschaftsrelation von Equipments auf derselben Etage beschreibt, kann am besten anhand der Abbildung 25 erläutert werden.

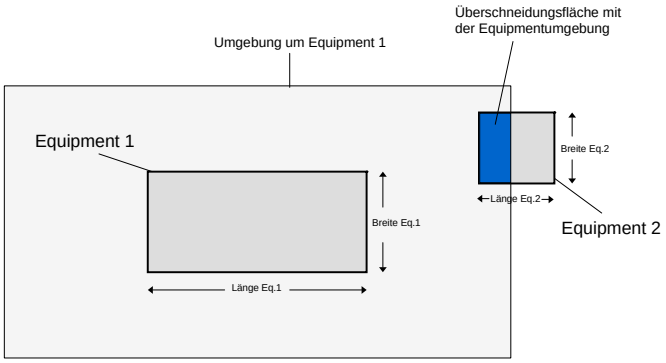


Abbildung 25. Berechnung der Nachbarschaftrelation zweier Equipments

Dort sind zwei Equipments zu erkennen, ein größeres auf der linken Seite (*Equipment 1*) sowie ein kleineres Equipment auf der rechten Seite (*Equipment 2*). Um den Grad der Nachbarschaft dieser Equipments zu messen, wird zunächst um das breitere Equipment eine nach zwei Seiten begrenzte Umgebung mit der Breite $\frac{Breite(Eq.1)+Breite(Eq.2)}{2}$ und um das längere Equipment eine nach zwei Seiten begrenzte Umgebung mit der der Länge $\frac{Länge(Eq.1)+Länge(Eq.2)}{2}$ gelegt. Ist wie in Abbildung 25 das breitere auch gleichzeitig das längere Equipment, so ergibt sich eine geschlossene Umgebung um dieses Equipment. Die Relation ist dann definiert als Überschneidungsfläche zwischen der Umgebung und der Fläche des kleineren Equipments bezogen auf die Fläche des kleineren Equipments. Liegt das kleinere Equipment vollständig innerhalb der Umgebung, ergibt sich ein Wert von “1”, liegt es vollständig außerhalb, so ergibt sich ein Wert von “0” für die Relation. Ist ein Equipment das breitere und das andere das längere, so ergibt sich jeweils eine nur nach zwei Seiten begrenzte Umgebung um das jeweilige Equipment. Der Relationswert berechnet sich dann aus der Multiplikation der in die Umgebungen hineinragenden Flächenanteile des jeweils schmaleren und kürzeren Equipments. *liegt neben (benachbarte Etagen)* ergibt nur dann einen Wert, wenn

zwei Equipments auf übereinanderliegenden Etagen liegen. Dann wird der Wert so berechnet, wie wenn bei *liegt neben* die Equipments auf derselben Etage liegen. *liegt auf Etage* gibt die Etage an, in der das Equipment platziert wurde. *Verhältnis Länge zu Breite* ergibt sich aus dem Verhältnis zwischen Länge und Breite eines Equipments. *liegt außerhalb des Stahlbaus* ist definiert als der Flächenanteil eines Equipments, der außerhalb des Stahlbaus liegt. Liegt ein Equipment nicht im Erdgeschoß, liegt es immer komplett innerhalb des Stahlbaus und es ergibt sich ein Wert von "0". Bei der Berechnung des Merkmalsblocks *liegt am Weg* wird der Abstand eines Equipments zum nächstgelegenen Anlagenweg betrachtet. Grenzt ein Equipment unmittelbar an einen Weg, wird der Wert "1", liegt es drei oder mehr Meter vom nächsten Weg entfernt, ergibt sich ein Wert von "0". Bei einer Entfernung zwischen null und drei Metern, berechnet sich ein Wert zwischen "0" und "1". Die Aufschlüsselung des letzen Merkmalsblocks *Equipmenttyp* ist in Tabelle 7 aufgezeigt.

Equipmentnummer	*Equipmenttyp*	*Anzahl der Equipments je Typ*
1	Kolonne	2
2	Wärmetauscher	12
3	Behälter	2
4	Vakuumpumpe	2
5	Speisepumpe	7
6	Rücklaufbehälter	2
7	Dusche	1

Tabelle 7. Zuordnung von Equipmentnummern und Equipmenttypen für die reale Anlage

Im Vorfeld des Vergleiches der Lernsätze für die equipmentbasierte Repräsentation wurden in einem Parameteroptimierungslauf folgende optimale Einstellungen gefunden ($C = 0.1$, *epsilon* $= 1.0$ und *Größe des Samples* $= 250$).
Für die equipmentbasierte Repräsentation erhält man einen minimalen relativen Fehler von 3.05% und einen mittleren absoluten Fehler von 62582.8, also eine Abweichung von ca. 62 m bezogen auf die Rohrleitungslänge der gesamten Anlage in Höhe von ca. 2052 m (siehe Abbildung 26). Die durch vollständige Suche ermittelte optimale Merkmalsmenge enthält insgesamt 1162 Merkmale und besteht aus den relationalen Merkmalsblöcken *DifferenzInZ*, *Distanz* und *liegt über/unter/gleich* sowie dem Block *liegt am Weg*. Die restlichen Merkmalsblöcke sind für die gegebene Lernaufgabe nicht relevant. Das Optimum wird in dieser Experimentreihe allerdings von keinem der untersuchten Selektionsansätze erreicht. Das zweitbeste Ergebnis erzielt der *genetische Algorithmus* mit der Merkmalsmenge *DifferenzInZ*, *Distanz*, *liegt neben (benachbarte Etagen)*, *Verhältnis*

	absoluter Fehler	relativer Fehler	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen
Ohne Selektion	71266,7	3,47	4088	5
Merkmalsauswahl mittels Forward Selection	66942,9	3,25	784	190
Merkmalsauswahl mittels Backward Elimination	66239,1	3,32	2002	155
Merkmalsauswahl mittels GA	64165,1	3,12	1218	1950
Vollständige Suche (Brute Force)	62582,8	3,05	1162	10240
Merkmalsauswahl mittels GA (ohne MB)	68897,3	3,37	547	3000

Abbildung 26. Vergleich diverser Kennzahlen verschiedener Lernketten bei der equipmentbasiertem Repräsentation

Länge zu Breite, *liegt am Weg* und *Equipmenttyp* bei einem relativen Fehler von 3.12% und einem absoluten Fehler von 64165.1. Festzuhalten bleibt jedoch, dass der *genetische Algorithmus* nur ca. 20% der im Rahmen der vollständigen Suche durchgeführten Performanzberechnungen benötigt. Ein Untersuchung der *GA*-basierten Merkmalsauswahl ohne Verwendung von Merkmalsblöcken ergibt, dass dessen Performanz trotz eines vergleichsweise hohen Rechenaufwandes geringer als die aller anderen Auswahlverfahren mit Merkmalsblöcken ist. Diese Tatsache läßt sich insbesondere durch den deutlich größeren Suchraum (statt zehn Merkmalsblöcken mehrere hundert einzelne Merkmale) und die damit verbundene deutlich schwierigere Lernaufgabe erklären.

Typenbasierte Anlagenrepräsentation

Die bisher vorgestellten Repräsentationen (mit Ausnahme der Vergleichsrepräsentation in Absolutkooordinaten) wurden equipmentbasiert formuliert, d.h. ein Merkmal der Repräsentation bezog sich entweder auf ein bzw. zwei konkrete Equipments. Diese Repräsentation hat den Nachteil, dass die gefundenen Merkmalsmengen nicht auf andere Anlagen übertragbar sind.
Die nun vorgestellte Repräsentation löst dieses Problem durch eine equipmentunabhängige Beschreibung. Im Gegensatz zu den bisherigen Repräsentationen werden Relationen wie z.B. *liegt über* nun nur noch auf der Ebene von Equipmenttypen formuliert. Merkmalswerte sind in diesem Fall die Häufigkeit des Auftretens einer Typ-Kombination bzgl. einer gegebenen Relation (hier: *liegt über*).
Abbildung 27 zeigt einen fiktiven Ausschnitt aus einer solchen Typenkombinationsmatrix. Beispielsweise bedeutet der Inhalt des Feldes Typ_j x Typ_n, dass fünf Equipments des Typs j (für $j = 3$ ein Behälter) über Equipments des Typs n (für $n = 5$ eine Speisepumpe) platziert wurden. Diese Beschreibung erlaubt die Übertragbarkeit von gelerntem Platzierungswissen auf andere Anlagentypen.
Die Tabelle 8 enthält eine vollständige Liste der verwendeten Merkmalsblöcke, eine kurze Erläuterung der Blöcke, den Wertebereich einzelner Merkmale (hier die Häufigkeit von Typ-Kombinationen bezogen auf eine gegebene Relation) sowie die resultierende Zahl der Merkmale je Merkmalsblock. Dabei wurden die Merkmalsblöcke *liegt über* und *liegt unter* aus dem in Tabelle 6 gezeigten Merkmals-

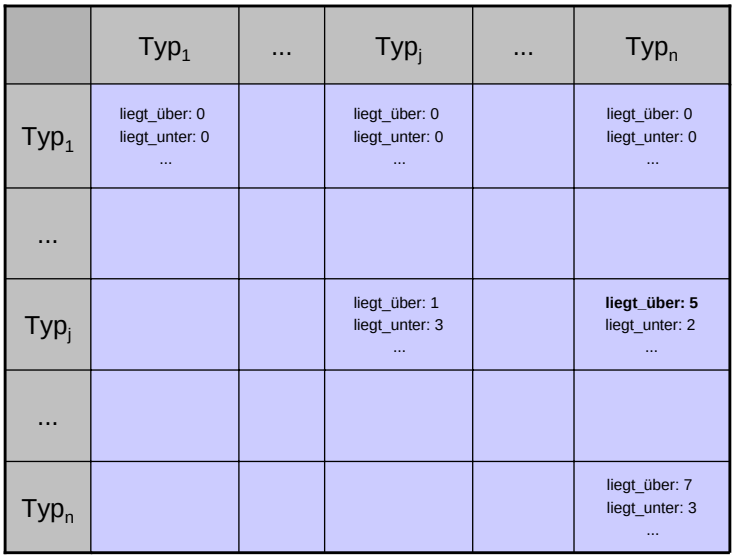


Abbildung 27. Visualisierung der typenbasierten Repräsentation

Nr.	Name des Merkmalsblocks	Bedeutung der Merkmale	Wertebereich	Anzahl der Merkmale je Merkmalsblock
1	*liegt über*	Häufigkeit, mit der ein Equipment des Typs i über dem eines Typs j liegt (ein Merkmal für jede mögliche **Typen**kombination)	$\mathbb{N}^+$	49
2	*liegt unter*	Häufigkeit, mit der ein Equipment des Typs i unter dem eines Typs j liegt (ein Merkmal für jede mögliche **Typen**kombination)	$\mathbb{N}^+$	49
3	*liegt neben*	Häufigkeit, mit der ein Equipment des Typs i neben dem eines Typs j liegt (ein Merkmal für jede mögliche **Typen**kombination)	$\mathbb{N}^+$	49
4	*liegt am Weg*	Häufigkeit, mit der ein Equipment an einem Anlagenweg liegt (ein Merkmal je Typ)	$\mathbb{N}^+$	7
5	*liegt außerhalb des Stahlbaus*	Häufigkeit, mit der ein Equipment außerhalb der Anlage liegt (ein Merkmal je Typ)	$\mathbb{N}^+$	7

Tabelle 8. Typenbasierte Repräsentation der realen Anlage

block *liegt über/unter/gleich* generiert, bei den übrigen drei Merkmalsblöcken wurde das entsprechende Merkmal bei einem Wert von größer "0" (siehe Tabelle 6) als erfüllt gewertet.
Im Vorfeld des Vergleiches der Lernansätze für die typenbasierte Repräsentation wurden in einem Parameteroptimierungslauf folgende optimale Einstellungen ermittelt ($C = 10.0$, *epsilon* $= 10.0$ und *Größe des Samples=250*).
Ein Vergleich der Lernansätze (siehe Abbildung 28) ergibt für die typenbasierte Repräsentation einen minimalen relativen Fehler von 4.02% und einen mittleren absoluten Fehler von 82122.7, also eine Abweichung von ca. 80 m bezogen auf eine Gesamt-Rohrleitungslänge von ca. 2042 m. Die durch vollständige Suche ermittelte optimale Merkmalsmenge besteht aus den relationalen Merkmalsblöcken *liegt über*, *liegt unter* und *liegt neben*. Die Merkmalsblöcke *liegt am Weg* und *liegt außerhalb* sind für die gegebene Lernaufgabe nicht relevant. Als einziges Selektionsverfahren findet der *genetische Algorithmus* das globale Optimum, obwohl dafür ein nicht unerheblicher Rechenaufwand in Kauf genommen werden muss (2760 Performanzberechnungen im Vergleich zu 30 bzw. 70 Berechnungen bei den sequentiellen Verfahren). Insgesamt beträgt der Performanzunterschied zwischen der vollständigen Merkmalsmenge und der optimalen Teilmenge 9%, d.h. durch die Selektion von Merkmalen ist nur noch ein geringer Zugewinn in der Performanz zu erzielen. Die mit der equipmentbasierten Repräsentation erzielten Ergebnisse in Bezug auf die Performanz der *GA*-basierten Merkmalsauswahl ohne Verwendung von Merkmalsblöcken sind auch auf die typenbasierte Repräsentation übertragbar. Auch hier erzielt das Verfahren ohne Merkmalsblöcke trotz eines hohen Rechenaufwandes die geringste Performanz aller untersuchten Auswahlverfahren.

	absoluter Fehler	relativer Fehler	Anzahl der selektierten Merkmale	Anzahl der benötigten Performanzberechnungen
Ohne Selektion	82854,2	4,06	161	5
Merkmalsauswahl mittels Forward Selection	82345,3	4,04	147	70
Merkmalsauswahl mittels Backward Elimination	82854,2	4,06	161	30
Merkmalsauswahl mittels GA	82122,7	4,02	154	2760
Vollständige Suche (BruteForce)	82122,7	4,02	154	160
Merkmalsauswahl mittels GA (ohne MB)	82854,2	4,06	161	3520

Abbildung 28. Vergleich diverser Kennzahlen verschiedener Lernansätze bei der typenbasierten Repräsentation

4 Zusammenfassung und Ausblick

Das Ziel dieser Untersuchungen war die Feststellung relevanter Eigenschaften der Aufstellungen von Chemieanlagen. In diesem Zusammenhang wurden zwei Zielkriterien, der Erfüllungsgrad von Platzierungsanforderungen für eine konkrete Aufstellung sowie die Gesamtlänge der resultierenden Anlagenverrohrung, betrachtet.
Im ersten Fall war das Ziel eine automatische Klassifikation von Platzierungsentwürfen anhand von Aufstellungen, die einen bestimmten Anteil an *Muss*- und *Soll*-Anforderungen erfüllen (entspricht einem positiven Beispiel) bzw. nicht erfüllen (entspricht einem negativen Beispiel). Im zweiten Fall war das Ziel eine möglichst exakte Vorhersage der zu erwartenden Rohrleitungslänge bei einer Menge gegebener Aufstellungen. In beiden Fällen wurden die Merkmalsmengen, die die Modelle mit der jeweils höchsten Vorhersagegenauigkeit auf den Beispieldaten erzeugten, ermittelt.
Generell kann festgehalten werden, dass der überwiegende Anteil der relevanten Aufstellungseigenschaften relationalen Charakter besitzt. Das bedeutet, dass Merkmale, die Beziehungen zwischen einzelnen Equipments bzw. zwischen Equipments und dem Baufeld ausdrücken, eine höhere Relevanz besitzen als Merkmale, die eine absolute Beschreibung beinhalten. Unter den relevanten relationalen Merkmalen sind wiederum solche hervorzuheben, die eine Höhenrelation zwischen Equipments beschreiben, wie insbesondere *DifferenzInZ*, *liegt über*, *liegt unter* und *liegt über/unter/gleich*.
Neben den Repräsentationen, die sich jeweils auf eine konkrete Equipmentliste und ein vorgegebenes Baufeld beziehen, wurde ein genereller Ansatz vorgestellt, der derartige Vorgaben nicht benötigt und somit prinzipiell auf beliebigen Anlagen anwendbar ist. In diesem Zusammenhang wurde mit der resultierenden Rohrleitungslänge ein Maß gewählt, das unabhängig von einer vorgegebenen Wissensbasis die Qualität einer Aufstellung beschreiben kann. Die vorgestellte typenbasierte Repräsentation ermöglichte eine sehr präzise Vorhersage (relativer Fehler in Höhe von ca. 4%) der Rohrleitungslänge unter Vorgabe der realen Anlage.
Es konnte gezeigt werden, dass der beschriebene evolutionäre Merkmalsselektionsansatz bei allen untersuchten Repräsentationen die jeweils höchste Performanz erzielt hat. Durch eine vollständige Suche im Merkmalsraum konnte für alle betrachteten Repräsentationen nicht nur die Effizienz (im Sinne der Vorhersagegenauigkeit der gelernten Modelle), sondern auch die Effektivität der jeweiligen Lernkette (im Sinne des Auffindens einer optimalen Merkmalsmenge) überprüft werden. Im Gegensatz zu den sequentiellen Selektionsverfahren erreichte der evolutionäre Ansatz, mit einer Ausnahme, bereits nach wenigen Generationen das globale Optimum. Es konnte weiterhin gezeigt werden, dass in den Fällen, in denen die sequentiellen Verfahren lediglich eine suboptimale Merkmalsmenge zurückgaben, eine Nichtmonotonie des jeweiligen Bewertungsverbandes vorlag.
Die Verwendung der *GA*-basierten Merkmalsselektion mit dem Evaluationskriterium *Accuracy* führte, je nach untersuchter Repräsentation, zu einer Reduktion der Merkmalsmenge von ca. 5 bis 70% bei gleicher bzw. sogar höherer Perfor-

manz im Vergleich zu einem Ansatz ohne Merkmalsselektion. Der Nachteil eines globalen Suchverfahrens, insbesondere wenn dieses wie hier in einen *Wrapper* eingebunden ist, ist allerdings ein teils erheblicher Rechenaufwand - in unserem Fall von mehreren hundert bis tausend Performanzberechnungen.
Neben der Vorhersagegenauigkeit wurde ein weiteres Maß für die Generalisierungsfähigkeit der Lernansätze untersucht, das die Größe der erzeugten Merkmalsmenge in die Performanzberechnunug miteinbezieht (sogenanntes *MDL*-Maß). Mit diesem Performanzkriterium konnte im Vergleich zu den mit dem *Accuracy*-Maß erzielten Ergebnissen eine weitere Reduktion der Merkmalsmenge auf ca. 3% der Originalgröße bei vergleichbarer Performanz erzielt werden.
Als weiterer Aspekt wurde die Auswirkung von Merkmalsblöcken auf unterschiedliche Kennzahlen betrachtet. Dabei ergab die Verwendung von Merkmalsblöcken neben einer deutlich verringerten Komplexität des Suchraumes (statt mehrerer hundert einzelner Merkmale nur noch maximal zehn zusammenhängende Merkmalsblöcke) auch eine, teilweise erhebliche, Verbesserung der Vorhersagegenauigkeit der entsprechenden Modelle. Neben dieser eher empirischen Motivation ermöglicht die Verwendung von Merkmalsblöcken die Beantwortung der Fragestellung nach der Relevanz von Eigenschaften einer Aufstellung (ist die Eigenschaft *liegt über* **generell** von Bedeutung) im Gegensatz zu der konkreten Fragestellung welche spezifischen Ausprägungen einer Eigenschaft (z.B. "$Equipment_i$ *liegt über* $Equipment_j$") sich auf die Aufstellungsgüte auswirken.
Insgesamt kann also festgestellt werden, dass für die gegebene Aufgabenstellung eine typenbasierte Repräsentation gefunden wurde, die prinzipiell für *beliebige* Anlagen, ohne vorherige Festlegung von Komponentenliste und Baufeldplan, anwendbar ist und auf der untersuchten realen Anlage einen sehr geringen Vorhersagefehler aufwies.
Die bisher durchgeführten Untersuchungen lassen bereits erste Schlüsse bzgl. der Relevanz bestimmter Platzierungseigenschaften zu. Um noch aussagekräftigere Informationen ermitteln zu können, wäre jedoch die Untersuchung weiterer Chemieanlagen sinnvoll.

5 Danksagung

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Sonderforschungsbereiches Computational Intelligence (SFB 531) der Universität Dortmund gefördert.

Literatur

[Aar92] E. Aarts. Simulated annealing. *UMAP Journal*, 13(1):79–90, 1992.

[AB96] D.W. Aha und R.L. Bankert. A comparative evaluation of sequential feature selection algorithms. In D. Fisher und H.-J. Lenz, Herausgeber, *Learning from Data*, chapter 4, 199–206. Springer, New York, USA, 1996.

[ACM77] L. Amorese, V. Cena, und C. Mustacchi. A heuristic for the compact location of process components. *Chemical Engineering Science*, 32, 1977.

[Bur98] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[Che90] P. Cheeseman. *Computational Models of Scientific Discovery and Theory Foundation*. Morgan Kaufmann, Los Altos, CA, USA, 1990.

[DDE+02] G. Daniel, J. Dienstuhl, S. Engell, S. Felske, K. Goser, R. Klinkenberg, K. Morik, O. Ritthoff, und H. Schmidt-Traub. *Advances in Computational Intelligence - Theory and Practice*, Chapter 8 - Novel Tasks, Optimization, and Their Application. Natural Computing Series. Springer Verlag, 2002.

[FKMR02] S. Fischer, R. Klinkenberg, I. Mierswa, und O. Ritthoff. Yale: Yet Another Learning Environment - Tutorial. Technical Report CI-136/03, SFB 531, University of Dortmund, 2002. http://yale.cs.uni-dortmund.de/.

[Flo95] C.A. Floudas. *Nonlinear and Mixed Integer Optimization : Fundamentals and Applications*. Oxford University Press, New York, 1995.

[Fre01] A.A. Freitas. Understanding the crucial role of attribute interaction in data mining. *Artificial Intelligence Review journal*, 16(3):177–199, 2001.

[Fre02] A.A. Freitas. *Data Mining and knowledge discovery with evolutionary algorithms*. Natural Computing Series. Springer, 2002.

[GSRM99] M.C. Georgiadis, G. Schilling, G.E. Rotstein, und S. Macchietto. A general mathematical programming approach for process plant layout. *Computers and Chemical Engineering*, (23):823–840, 1999.

[Has95] L. Hasenauer. *Entwicklung einer Methodik zur rechnergestützten Erstellung von Rohrleitungsstudien*. Dissertation, Düsseldorf, 1995.

[Hol00] T. Holtkötter. *Integriertes Equipment-Modelling - Ein Beitrag zur Optimierung der Aufstellungsplanung von Chemieanlagen*. Dissertation, Düsseldorf, 2000.

[Joa01] T. Joachims. *The Maximum-Margin Approach to Learning Text Classifiers: Methods, Theory, and Algorithms*. Dissertation, Lehrstuhl für künstliche Intelligenz, Fachbereich Informatik, Universität Dortmund, Februar 2001.

[KJ97] R. Kohavi und G. H. John. Wrappers for feature subset selection. *Artificial Intelligence Journal, Special Issue on Relevance*, 97(1–2):273–324, 1997.

[Kös98] D. Köster. *Ein Assistenzsystem zur methodischen Unterstützung der Aufstellungsplanung von Chemieanlagen*. Dissertation, Aachen, 1998.

[Leu02] P. Leuders. *Rechnergestützte Optimierung der Layoutplanung von Chemieanlagen*. Dissertation, Aachen, 2002.

[LM98] H. Liu und H. Motoda. *Feature Extraction, Construction, and Selection: A Data Mining Perspective*. Kluwer, Dordrecht, NL, 1998.

[MHS70] R. Malingriaux, K.R. Hilbring, und L. Schuart. Zur optimalen Anordnung der Elemente in Anlagen der Stoffumwandelnden Industrie. *Wissenschaftliche Zeitung der technischen Hochschule Otto von Guericke*, 14(8), 1970.

[MKFR03] I. Mierswa, R. Klinkenberg, S. Fischer, und O. Ritthoff. A flexible platform for knowledge discovery experiments: Yale - yet another learning environment. In *LLWA 03 - Tagungsband der GI-Workshop-Woche Lernen - Lehren - Wissen - Adaptivität*, 2003. http://yale.cs.uni-dortmund.de/.

[Mut95] R. Mutschall. *Entwicklung einer Methode für die rechnergestützte Aufstellungsplanung verfahrenstechnischer Anlagen*. Dissertation, Aachen, 1995.

[Nip00] N. Nipper. *Rechnergestützte Erstellung und Bewertung von Rohrleitungsverläufen für den Chemieanlagenbau*. Dissertation, Düsseldorf, 2000.

[PC96] F. Penteado und A. Ciric. A MINLP approach for safe process plant layout. *Ind. Eng. Chem. Res.*, (35), 1996.

[Ris78] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

[RK03] O. Ritthoff und R. Klinkenberg. Evolutionary feature space transformation using type-restricted generators. In E. et al. (ed.) Cantu-Paz, Herausgeber, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2003) - Part II*, Lecture Notes in Computer Science (LNCS 2724). Springer Verlag, 2003.

[Rü00] S. Rüping. *mySVM-Manual*. Lehrstuhl für künstliche Intelligenz, Fachbereich Informatik, Universität Dortmund, 2000. http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/.

[SS98] A. J. Smola und B. Schölkopf. A tutorial on support vector regression. NeuroCOLT2 Technical Report NC-TR-98-030, Royal Holloway College, Universität London, GB, 1998.

[ST98] H. Schmidt-Traub. CAD-systems for conceptual plant layout and piperouting. In *Dechema Monographien*, Vol. 135, 361–, 1998.

[STBLL01] H. Schmidt-Traub, A. Burdorf, M. Lederhose, und P. Leuders. Systematischer Ansatz zur rechnergestützten Optimierung von Aufstellungsentwürfen von Chemieanlagen. *Chemie Ingenieur Technik*, 73(1/2):40–, 2001.

[STBMN97] H. Schmidt-Traub, M. Busch, U. Mahlfeld, und N. Nipper. Optimierung des Anlagen-Layouts mittels Aufstellungsvarianten und Piperouting. *Chemie Ingenieur Technik*, 9(69):1271–, 1997.

[STHKL98] H. Schmidt-Traub, T. Holtkötter, M. Köster, und P. Leuders. Conceptual plant layout. *Computers and Chemical Engineering*, 22(Supplement):76–, 1998.

[STHL98] H. Schmidt-Traub, T. Holtkötter, und P. Leuders. Transparente Planung rechnergestützte Aufstellungsplanung und Rohrleitungsführung. *Chemie Technik*, 27(3):76–78, 1998.

[STLL00] H. Schmidt-Traub, M. Lederhose, und P. Leuders. Rechnergestützter Rohrleitungsentwurf für die Aufstellungsplanung. *3 R International*, 39(4/5), 2000.