

---

# Klassifikation der Schwierigkeitsgrade von Sudokus mit Methoden des maschinellen Lernens

---

Investigation of Sudoku difficulty levels  
Bachelor-Thesis von Michael Bräunlein

---



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Fachbereich Informatik  
Knowledge Engineering Group  
Betreuer Prof. Dr. Johannes Fürnkranz

---

---

## **Zusammenfassung**

---

Das Zahlenrätsel Sudoku ist weltweit bei Rätselliebhabern bekannt und beliebt. Seit seiner Veröffentlichung 1986 begeistern sich immer mehr Menschen für mehr oder weniger schwierige Exemplare. Sudokus finden sich im Internet, in der Rätselecke der Tageszeitungen und sogar als ganze Bücher, um nur einige Erscheinungsorte zu nennen.

Die Regeln sind einfach zu lernen und doch kann man sich sehr lange mit Sudokus beschäftigen, da die schwersten Sudokus meist nur von Profis gelöst werden können.

Der Spielspass ist sehr stark davon abhängig, dass die Schwierigkeit zum persönlichen Können passt. Ist das Sudoku zu leicht, stellt es keine Herausforderung dar. Ist es zu schwer, kommt schnell ein Gefühl der Überforderung auf.

Die ausgewiesenen Schwierigkeitsstufen von Sudokus aus verschiedenen Quellen haben zwar oft die gleichen Namen wie zum Beispiel "Mittel", unterscheiden sich aber dennoch häufig nach Meinung des Spielers.

Das Ziel dieser Bachelorarbeit ist, Merkmale aus Sudokus zu extrahieren, anhand derer die Sudokus von einem Klassifizierer möglichst zuverlässig in Schwierigkeitsstufen eingeteilt werden können.

---

## Inhaltsverzeichnis

---

<b>1</b>	<b>Aufgabenstellung und Zielsetzung</b>	<b>3</b>
<b>2</b>	<b>Einführung</b>	<b>4</b>
2.1	Die Regeln . . . . .	4
2.2	Begriffserklärung . . . . .	5
<b>3</b>	<b>Lösungsmethoden</b>	<b>6</b>
3.1	Kandidatenlisten . . . . .	6
3.2	Full House . . . . .	7
3.3	Naked Single . . . . .	8
3.4	Hidden Single . . . . .	8
3.5	Pointing Pair / Triple . . . . .	8
3.6	Box-Line Reduction . . . . .	8
3.7	Naked Subset . . . . .	8
3.8	Hidden Subset . . . . .	8
3.9	Fish . . . . .	8
3.9.1	X-Wing . . . . .	8
3.9.2	Swordfish . . . . .	8
3.9.3	Jellyfish . . . . .	8
3.10	Single Digit Patterns . . . . .	8
3.10.1	Skyscarper . . . . .	8
3.10.2	2-String Kite . . . . .	8
3.10.3	Turbot Fish . . . . .	8
3.10.4	Empty Rectangle . . . . .	8
3.11	Wings . . . . .	8
3.11.1	XY-Wing . . . . .	8
3.11.2	XYZ-Wing . . . . .	8
3.11.3	W-Wing . . . . .	9
3.12	Sue de Coq . . . . .	9
3.13	Coloring . . . . .	9
3.14	Almost Locked Set . . . . .	9
3.14.1	ALS XZ . . . . .	9
3.14.2	ALS XY Wing . . . . .	9
3.14.3	ALS Chain . . . . .	9
<b>4</b>	<b>Merkmalsextrahierung</b>	<b>10</b>
4.1	Allgemeines Vorgehen . . . . .	10
4.2	Entkopplung von konkreten Zahlen . . . . .	10

---

## 1 Aufgabenstellung und Zielsetzung

---

Diese Bachelorarbeit beschäftigt sich mit der Einteilung von Sudokus in verschiedene Schwierigkeitsstufen. Hierzu sollen Methoden des maschinellen Lernens verwendet werden.

Es soll eine Methode gefunden werden, mit der Merkmale aus Sudokus extrahiert werden können, die dann als Feature Vektoren in einer .arff Datei<sup>1</sup> gesammelt werden. Die Feature Vektoren werden anschließend mit Hilfe von Weka<sup>2</sup> klassifiziert.

Es werden verschiedene Klassifikatoren und unterschiedliche Parameter betrachtet. Ausserdem werden Optimierungen der Featurevektoren diskutiert.

---

<sup>1</sup> <http://www.cs.waikato.ac.nz/ml/weka/arff.html>

<sup>2</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

---

## 2 Einführung

---

Die Vorfahren des heutigen Sudokus waren vermutlich die lateinischen Quadrate, mit denen sich vor allem der Mathematiker Leonhard Euler befasste. Hier ging es darum, in ein Quadrat mit  $n$  Spalten und  $n$  Reihen Symbole so einzutragen, dass jedes Symbol in jeder Spalte und Zeile jeweils genau einmal vorkommt.

1	2	3	4	5
2	5	4	1	3
3	4	5	2	1
4	3	1	5	2
5	1	2	3	4

**Abbildung 2.1:** Lateinisches Quadrat

Daraus hat sich das heutige Sudoku entwickelt, das sich nicht nur bei Mathematikern großer Beliebtheit erfreut.

---

### 2.1 Die Regeln

---

Diese Arbeit beschäftigt sich nur mit der meist verbreiteten Art von Sudokus. Dabei spielt man auf einem 9x9 Felder großen Spielfeld, das wiederum in neun 3x3 Felder große Blöcke eingeteilt ist. Weiter handelt es sich nur dann um ein Sudoku, wenn genau eine Lösung vorhanden ist. Ein Sudoku gilt dann als gelöst, wenn jede Zeile, jede Spalte und jeder Block die Ziffern 1 bis 9 genau einmal enthält.

4	8	1	5	6		3		
7	6	9	3	2	4		5	
3	5			7		6		
	9	7	2	8	5	2	1	
1			1	3	6	5	4	
5	4		1	3	2	8	6	
	7	6	9	5	3		2	
2		5		4			3	
9	3	4				6	5	

**Abbildung 2.2:** Sudoku

---

## 2.2 Begriffserklärung

---

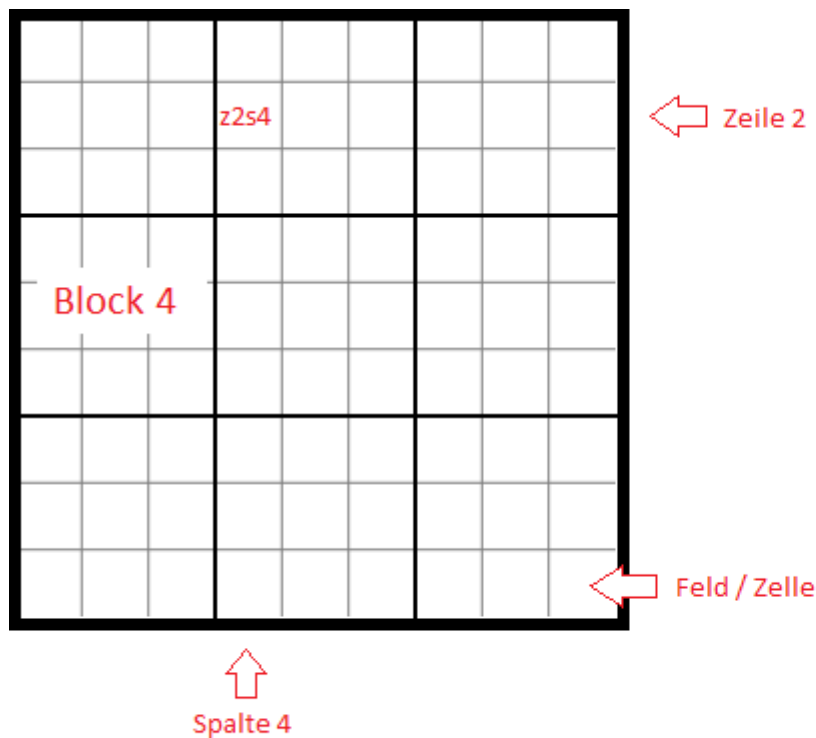
Ein Sudoku besteht aus 81 *Feldern* oder *Zellen*. Diese bilden ein Quadrat der Größe 9x9, das *Grid*. Aufgrund dieser Aufteilung hat ein Sudoku 9 *Zeilen* und 9 *Spalten*. Das Grid wird in 9 Unterquadrate geteilt, die jeweils 3x3 Felder groß sind. Diese werden *Blöcke* genannt. Zeilen, Spalten und Blöcke werden unter dem Begriff *Figur* zusammengefasst. Die Nummerierung der Blöcke erfolgt zeilenweise von links oben nach rechts unten.

*Vorgaben* sind Zahlen, die schon von Anfang an gegeben sind.

In **Abbildung 2.2** sieht man im mittleren Block sogenannte *Kandidaten*. Ein Kandidat ist eine Zahl, die in der Zelle noch möglich ist. Jede Zelle hat ihre eigene Liste mit Kandidaten.

In der Beschreibung der Lösungstechniken ist es notwendig bestimmte Felder zu betrachten. Hierzu wird eine Abkürzung verwendet, die Zeile und Spalte enthält und somit eine Zelle eindeutig indentifiziert. z2s3 meint zum Beispiel die Zelle in Zeile 2 und Spalte 3.

In der folgenden Abbildung sind die erläuterten Begriffe zum besseren Verständniss eingetragen.



**Abbildung 2.3:** Begriffe

---

### **3 Lösungsmethoden**

---

Alle in dieser Bachelorarbeit beschriebenen Techniken sind nicht im Rahmen dieser Arbeit entwickelt worden, sondern wurden aus verschiedenen Quellen zusammengetragen. Die Beschreibung der Lösungstechniken lehnt sich an die Beschreibung der Quellen an. Teile der Beispiele wurden aus den Quellen entnommen, dies ist entsprechend gekennzeichnet.

Grob kann man die Techniken zum Lösen von Sudokus in zwei Kategorien einteilen. Die erste Kategorie findet Zahlen heraus, die direkt in das Sudoku eingetragen werden können. Die Techniken der zweiten Kategorie entfernen Bedingungen in einzelnen Zellen des Sudokus.

---

#### **3.1 Kandidatenlisten**

---

Beim Lösen von Sudokus ist es üblich, in jedes Feld die Kandidaten einzutragen, die dort stehen können. Dabei wird vorerst nur die Sudoku Regel berücksichtigt, die besagt, dass in jeder Zeile die Zahlen 1 bis 9 vorkommen müssen. Wenn in einer Zeile nun die Zahl 3 vorkommt, dann kann sie in der selben Zeile nicht nochmal vorkommen, daher kann sie aus allen Kandidatenlisten der Zellen in der selben Zeile gelöscht werden. Dasselbe gilt für Spalten und Blöcke. Immer wenn eine Ziffer in ein Feld eingetragen wird, dann muss der Spieler die Liste der Kandidaten aktualisieren.

Kandidatenlisten sind keine eigene Lösungstechnik, sind aber wesentlicher Bestandteil vieler Techniken.

### 3.2 Full House

Wenn in einer Figur 8 Zahlen eingetragen sind, dann kann die Technik *Full House* angewendet werden. Da in jeder Figur die Zahlen 1 bis 9 stehen müssen, kann die fehlende Zahl einfach per Ausschluss ermittelt werden.

Im folgenden Beispiel fehlt in Zeile 5 nur noch eine Ziffer. Da die Zahlen 2 bis 9 bereits vorhanden sind, kann in das Feld z5s8 die Zahl 1 eingetragen werden.

2	<sup>1</sup> 4 5 6	7	<sup>1</sup> 4 5 8	<sup>1 3</sup> 5	<sup>3</sup> 8	<sup>1</sup> 6	<sup>1 3</sup> 4 5 6	<sup>1 3</sup> 5 6 9
<sup>1</sup> 4 5	8	<sup>4 5 6</sup>	<sup>1 2</sup> 4 5	9	<sup>2 3</sup> 7	<sup>1 2</sup> 6 7	<sup>1 3</sup> 4 5 6 7	<sup>1 2 3</sup> 5 6 7
<sup>1</sup> 4 5 9	3	<sup>4 5</sup>	6	<sup>1</sup> 5 7	<sup>2</sup> 7	8	<sup>1</sup> 4 5 7	<sup>1 2</sup> 5 7 9
<sup>3</sup> 5 7	<sup>5</sup> 7	8	<sup>1</sup>	6	4	9	<sup>1</sup> 7	<sup>1 2</sup> 7
6	9	2	7	8	5	3	<sup>1</sup>	4
<sup>4</sup> 7	<sup>4</sup> 7	1	3	2	<sup>9</sup>	5	<sup>6</sup> 7 8	<sup>6</sup> 7
<sup>3</sup> 4 5 7 8	<sup>4 5 6</sup> 7	9	<sup>5</sup> 8 7	<sup>5 3</sup>	1	<sup>6</sup> 7	2	<sup>3</sup> 5 6 7
<sup>1 3</sup> 5 7 8	<sup>1 2</sup> 5 6 7	<sup>3</sup> 5 6	<sup>2</sup> 5 8	4	<sup>2 3</sup> 6 7 8	<sup>1</sup> 6 7	9	<sup>1 3</sup> 5 6 7
<sup>1 3</sup> 5 7	<sup>1 2</sup> 5 6 7	<sup>3</sup> 5 6	<sup>2</sup> 5 9 7	<sup>3</sup> 5	<sup>2 3</sup> 6 7 9	4	<sup>1 3</sup> 5 6 7	8

Abbildung 3.1: Full House



---

---

### 3.3 Naked Single

---

### 3.4 Hidden Single

---

### 3.5 Pointing Pair / Triple

---

### 3.6 Box-Line Reduction

---

### 3.7 Naked Subset

---

### 3.8 Hidden Subset

---

### 3.9 Fish

---

#### 3.9.1 X-Wing

---

#### 3.9.2 Swordfish

---

#### 3.9.3 Jellyfish

---

### 3.10 Single Digit Patterns

---

#### 3.10.1 Skyscarper

---

#### 3.10.2 2-String Kite

---

#### 3.10.3 Turbot Fish

---

#### 3.10.4 Empty Rectangle

---

### 3.11 Wings

---

#### 3.11.1 XY-Wing

---

#### 3.11.2 XYZ-Wing

---

---

---

3.11.3 W-Wing

---

3.12 Sue de Coq

---

3.13 Coloring

---

3.14 Almost Locked Set

---

3.14.1 ALS XZ

---

3.14.2 ALS XY Wing

---

3.14.3 ALS Chain

---

---

## **4 Merkmalsextrahierung**

---

### **4.1 Allgemeines Vorgehen**

---

### **4.2 Entkopplung von konkreten Zahlen**

---