

Klassifikation der Schwierigkeitsgrade von Sudoku mit Methoden des maschinellen Lernens

Michael Bräunlein
mbraeunlein@gmail.com



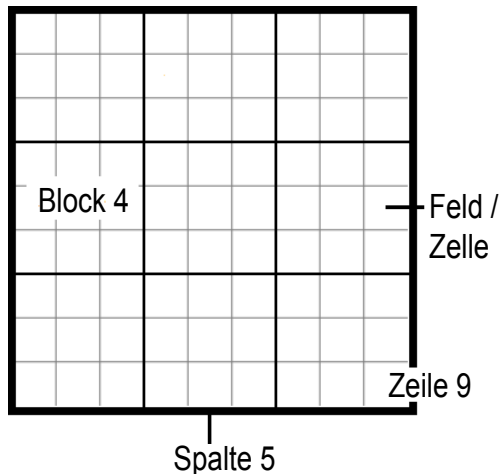
TECHNISCHE
UNIVERSITÄT
DARMSTADT



- ▶ Sudokus finden sich überall
- ▶ Unterschiedliche Bewertungsskalen
- ▶ Unterschiedliche Einteilungsverfahren
- ▶ Bisher kein Verfahren zur Einteilung mit maschinellem Lernen
- ▶ Sudokus sind zur Bearbeitung mit Computern prädestiniert



- ▶ Sudoku hat nur eine Regel
- ▶ In jeder Zeile, jeder Spalte und jedem Block muss jede Ziffer von 1 bis 9 genau einmal vorkommen
- ▶ Jedes Sudoku hat eine eindeutige Lösung
- ▶ Das Sudoku gilt dann als gelöst, wenn alle Felder ausgefüllt sind





- ▶ Jeder Spieler benutzt Lösungsmethoden
- ▶ Lösungsmethoden sagen viel über den Schwierigkeitsgrad aus
- ▶ Kandidatenlisten erleichtern das Finden von Zahlenkonstellationen, die Voraussetzung für bestimmte Lösungsmethoden sind
- ▶ Es gibt viele verschiedene Lösungsmethoden, grob werden zwei Kategorien unterschieden

Full House



TECHNISCHE
UNIVERSITÄT
DARMSTADT

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

Pointing Pair / Triple



TECHNISCHE
UNIVERSITÄT
DARMSTADT

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

Two-String-Kite



TECHNISCHE
UNIVERSITÄT
DARMSTADT

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

XY-Wing



TECHNISCHE
UNIVERSITÄT
DARMSTADT

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



- ▶ Merkmalsvektor
- ▶ n-dimensionaler Vektor
- ▶ Repräsentation eines Objekts
- ▶ Ein Eintrag steht für eine Eigenschaft des beschriebenen Sudokus
- ▶ Featurevektoren sind die Eingabe des Klassifikationsalgorithmus

Wie werden Featurevektoren erzeugt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ▶ Am Anfang bekannte Zahlen
- ▶ Einträge der Kandidatenlisten
- ▶ Hinzugefügte Zahlen
- ▶ Entfernte Zahlen
- ▶ Unterschiedliche Lösungswege für Sudokus möglich
- ▶ Einfachster Lösungsweg gesucht



- ▶ Fast gleiche Sudokus mit vertauschten Zahlen
- ▶ Gleicher Schwierigkeitsgrad
- ▶ Unterschiedliche Featurevektoren bei gleichem Lösungsweg



- ▶ Fast gleiche Sudokus mit vertauschten Zahlen
- ▶ Gleicher Schwierigkeitsgrad
- ▶ Unterschiedliche Featurevektoren bei gleichem Lösungsweg
- ▶ Lösung?



- ▶ Fast gleiche Sudokus mit vertauschten Zahlen
- ▶ Gleicher Schwierigkeitsgrad
- ▶ Unterschiedliche Featurevektoren bei gleichem Lösungsweg
- ▶ Sortierung der Features nach Häufigkeit
- ▶ Kein relevanter Informationsverlust
- ▶ Gleicher Featurevector auch bei vertauschten Ziffern

Entkopplung von konkreten Zahlen (Beispiel)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ▶ Beispiel des Featurevectors einer Methode
 $(1, 0, 4, 15, 3, 0, 9, 2, 0)^T$
- ▶ Vertauchte Ziffern 7 und 8
 $(1, 0, 4, 15, 3, 0, 2, 9, 0)^T$
- ▶ Nach der Sortierung nach der Häufigkeit
 $(15, 9, 4, 3, 2, 1, 0, 0, 0)^T$



- ▶ Fremdsoftware für Klassifizierer und Lösungsmethoden
- ▶ Für den Klassifizierer: Weka¹
- ▶ Für die Lösungsmethoden: Hodoku²
- ▶ Beide Projekte stehen unter der GPLv3 Lizenz
- ▶ Eigene Software in Java
- ▶ Extrahierung der Featurevektoren und Verbindung der Projekte

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

²<http://hodoku.sourceforge.net/de/index.php>