

# Vergleich zwischen geordneten und ungeordneten Regelmengen



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Amir Naseri



# Inhaltsübersicht

---



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

1. Einleitung
2. Erweiterung des SeCo-Frameworks
3. Evaluation
4. Konklusion

# 1. Einleitung

Gegeben sei eine Trainingsmenge

Ein Regellern-Algorithmus sucht nach einer Regelmenge, die die gesuchte Zieltheorie ist, oder dieser sehr nahe kommt

Regellern-Algorithmen als eine Klasse von Basisalgorithmen

Separate-and-Conquer Regellern-Algorithmen als eine Variante der Regellern-Algorithmen

# 1. Einleitung

Separate-and-Conquer Regellern-Algorithmen:  
Zuerst lernen wir eine Regel, die ein Teil von Instanzen abdeckt  
Die abgedeckten Instanzen werden entfernt  
Das Verfahren wird fortgesetzt

SeCo-Framework (Separate-and-Conquer): eine  
Implementierung einiger dieser Algorithmen

Entscheidungsliste: eine geordnete Liste von Regeln

Ungeordneten Regelmengen

## 2. Erweiterung des SeCo-Frameworks

### 1. Defaultlernverfahren

Umsetzung des Lernens von Regeln

Alle Werte des Klassenattributs nach Anzahl der Instanzen sortiert

Der Wert mit der höchsten Anzahl an Instanzen bildet die Defaultregel

Klassenattributswert	positiv	negativ
c1	c1	c2, c3
c2	c2	c1, c3

## 2. Erweiterung des SeCo-Frameworks

# 2. Majority-Voting-Verfahren

Für alle Werte des Klassenattributs werden Regeln gelernt

Klassenattributswert	positiv	negativ
c1	c1	c2, c3
c2	c2	c1, c3
c3	c3	c1, c2

## 2. Erweiterung des SeCo-Frameworks

# 3. Multi-Class-Covering-Lernverfahren



$R \leftarrow \emptyset$

Repeat

$r \leftarrow \emptyset$

for each class  $c_i, i = 1$  to  $C$  do

$P_i \leftarrow \{\text{subset of examples in } \varepsilon \text{ with class label } c_i\}$

$N_i \leftarrow \{\text{subset of examples in } \varepsilon \text{ with other class labels}\}$

$\acute{r} = \text{lernOneRule}(c_i, P_i, N_i)$

if  $\acute{r}$  better than  $r$

$r = \acute{r}$

$R = R \cup r$

$\varepsilon = \varepsilon \setminus \text{Covered}(r, \varepsilon)$

until  $R$  satisfies a quality threshold or  $\varepsilon$  is empty

$R = R \cup \text{default rule}$

## 2. Erweiterung des SeCo-Frameworks

# 3. Multi-Class-Covering-Lernverfahren



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

---

Gegebene Heuristik als Qualitätskriterium für den Vergleich  
zwischen Regeln

Es ist nicht strikt festgesetzt, dass man zuerst für  $c_1$ , dann für  
 $c_2$  und danach eventuell für  $c_3$  Regeln lernen muss

## 2. Erweiterung des SeCo-Frameworks

# 1. Defaultklassifizierungsverfahren

---

Umsetzung der Klassifikation

Entscheidungslisten wurden beim Lernverfahren gelernt

Keine Kollisionen treten bei der Klassifikation auf

Kollision: zwei verschiedene Regeln haben zwei verschiedene Vorhersagen für eine neue Instanz

Die Bedeutung von einer Regel ist abhängig von allen Regeln, die vor der Regel in der Entscheidungsliste stehen

## 2. Erweiterung des SeCo-Frameworks

### 2. Majority-Voting-Klassifizierungsverfahren



R1:  $B1 \wedge B2 \rightarrow \text{Klasse} = c1$

R2:  $B2 \wedge B3 \rightarrow \text{Klasse} = c2$

R3:  $B4 \wedge B5 \rightarrow \text{Klasse} = c3$

R4:  $B6 \rightarrow \text{Klasse} = c3$

Defaultregel:  $\text{True} \rightarrow \text{Klasse} = c3$

<Klasse; Frequenz> Liste:      <c1; 0>, <c2; 0>, <c3; 0>

Die Klassenattributswerte sind eindeutig

Annahme: das neue Beispiel von R2, R3 und R4 abgedeckt wird

<Klasse; Frequenz> Liste:      <c1; 0>, <c2; 1>, <c3; 2>

## 2. Erweiterung des SeCo-Frameworks

### 2. Majority-Voting-Klassifizierungsverfahren



Die Defaultregel gehört nicht zur Regelmenge, für die wir die Frequenzen berechnen

### 3. Gewichtetes-Majority-Voting-Klassifizierungsverfahren

Deckt die Regel das Beispiel ab, so wird die Frequenz des Klassenattributswerts um den Regelwert erhöht

Annahme: R2, R3 und R4 haben jeweils 0,4, 0,1 und 0,2 als Regelwert

$\langle \text{Klasse; Frequenz} \rangle$  Liste:  $\langle c1; 0,0 \rangle, \langle c2; 0,4 \rangle, \langle c3; 0,3 \rangle$

# 3. Evaluation

1. Lernen durch Defaultlernverfahren und klassifizieren durch Defaultklassifizierungsverfahren
  2. Lernen durch Multi-Class-Covering-Verfahren und klassifizieren durch Defaultklassifizierungsverfahren
  3. Lernen durch Majority-Voting-Verfahren und klassifizieren durch Majoritiy-Voting-Verfahren
  4. Lernen durch Majority-Voting-Verfahren und klassifizieren durch gewichtetes-Majoritiy-Voting-Verfahren
- 5 Verschiedene Heuristiken:  
Correlation, Laplace, m-Estimate, Relative-Linear-Cost (RLC),  
und Weighted-Relative-Accuracy (WRAcc)

### 3. Evaluation

## Heuristiken und Trainingsdaten



Die Heuristiken werden verwendet, um die Qualität der Regeln zu bewerten.

Die Auswahl verschiedener Heuristiken:

das Verhalten von Lern- und Klassifikationsverfahren kann abhängig von Heuristiken untersucht werden

WRAcc: allgemeinere Regelmengen

Laplace: Überanpassung ohne Pruning-Mechanismen

Trainingsdaten besteht aus 30 Datensätzen, 15 mit binärwertigen und 15 mit mehrwertigen Klassen

Nominale und numerische Attribute vorhanden

Die Datensätze haben unterschiedliche Anzahl an Instanzen

### 3. Evaluation

## Multi-Class-Covering-Verfahrensparameter

Das Multi-Class-Covering-Verfahren hat einen Parameter, der angibt, wie lange die äußere Schleife des Verfahrens fortgesetzt wird

Ein Parameterwert von 80%: die äußere Schleife wird solange fortgesetzt, bis 80% der Beispiele abgedeckt werden

Heuristik	80%	85%	90%	95%	100%
Correlation	77,818	78,260	78,469	78,515	78,519
Laplace	76,310	76,413	77,282	77,348	77,616
m-Estimate	76,186	77,089	77,429	78,160	78,347
RLC	76,485	76,717	76,922	77,025	77,186
WRAcc	76,358	76,669	76,771	77,079	77,197

### 3. Evaluation

## Ergebnisse der Evaluation

Binärwertige und mehrwertige Klassen getrennt

Bei jeder Heuristik werden zuerst die bei der Kreuz-Validierung korrekt klassifizierten Instanzen durch eine Win-Loss-Tie-Tabelle dargestellt

Win-Loss-Tie-Tabelle für Datensätze mit binärwertigen Klassen:

Correlation	Default	Multi-Class	Maj.-Vot.	Gew.-Maj.-Vot.
Default	x	8-5-2	4-9-2	5-9-1
Multi-Class	5-8-2	x	5-9-1	3-12-0
Maj.-Vot.	9-4-2	9-5-1	x	6-7-2
Gew.-Maj.-Vot.	9-5-1	12-3-0	7-6-2	x

### 3. Evaluation

## Ergebnisse der Evaluation



Die Lern- und Klassifikationsverfahren werden dann hinsichtlich der vier Kriterien, nämlich durchschnittliche Genauigkeit (Macro Average), durchschnittlicher Rank, durchschnittliche Anzahl der Regeln und Bedingungen, verglichen

Kriterium	Default	Multi-Class	Maj.-Vot.	Gew.-Maj.-Vot.
Macro AVG	82,216	81,669	82,604	83,172
AVG Rank	2,700	3,033	2,233	2,033
AVG # of rules	12,066	8,600	23,733	23,733
AVG # of conditions	38,400	29,466	73,400	73,400

### 3. Evaluation

# Ergebnisse der Evaluation

Alle 30 Datensätze zusammen untersucht

Annahme: die Heuristiken sind gleich wichtig, daher werden die durchschnittlichen Genauigkeiten über fünf Heuristiken ohne Gewichtung gemittelt

Heuristik	Default	Multi-Class	Maj.-Vot.	Gew.-Maj.-Vot.
Correlation	78,637	75,519	78,729	79,575
Laplace	76,984	77,616	77,208	78,260
m-Estimate	78,758	78,347	79,025	79,561
RLC	78,139	77,186	78,398	79,026
WRAcc	76,103	77,19	75,521	77,016
Mittelwert	77,724	77,173	77,776	78,687

## 4. Konklusion



Die besten Verfahren hinsichtlich der durchschnittlichen Genauigkeit:

Heuristik	binärwertig	mehrwertig
Correlation	gew.-Maj.-Vot.	gew.-Maj.-Vot.
Laplace	gew.-Maj.-Vot.	Multi-Class
m-Estimate	gew.-Maj.-Vot.	gew.-Maj.-Vot.
RLC	gew.-Maj.-Vot.	gew.-Maj.-Vot.
WRAcc	Multi-Class	Multi-Class

## 4. Konklusion

Die Rolle des Klassifizierungsverfahren beim gewichteten-Majority-Voting-Verfahren ist essenziell

Das Majority-Voting-Verfahren liefert bei der Correlation, Laplace und der WRAcc Heuristik für die Datensätze mit mehrwertigen Klassen das schlechteste Ergebnis, während es für die Datensätze mit binärwertigen Klassen das zweitstärkste Verfahren ist

Hinsichtlich der Anzahl der gelernten Regeln und Bedingungen erzeugt das Majority-Voting-Lernverfahren fast bei allen Heuristiken und für Datensätze mit binärwertigen und mehrwertigen Klassen Regelmengen mit der größten Anzahl an Regeln und Bedingungen

## 4. Konklusion



Die besten Verfahren hinsichtlich der Anzahl der gelernten Regeln:

Heuristik	binärwertig	mehrwertig
Correlation	Multi-Class	Multi-Class
Laplace	Multi-Class	Default
m-Estimate	Default	Default
RLC	Default	Default
WRAcc	Default	Default

## 4. Konklusion



Die besten Verfahren hinsichtlich der Anzahl der gelernten Bedingungen:

Heuristik	binärwertig	mehrwertig
Correlation	Multi-Class	Multi-Class
Laplace	Multi-Class	Multi-Class
m-Estimate	Default	Multi-Class
RLC	Default	Multi-Class
WRAcc	Default	Default

## 4. Konklusion

---

Das Multi-Class-Covering gegenüber dem Defaultverfahren den Vorteil hat, dass man noch kürzere Regelmengen generieren kann, indem man den Verfahrensparameter auf einen niedrigeren Wert setzt

Danke für Ihre Aufmerksamkeit