

Erweitern von Aqqu durch AutoSklearn

Bachelorarbeit

von Daniel Bindemann

Übersicht

- Motivation
- Funktionsweise von Auto-Sklearn
- Funktionsweise von Aqqu
- Ansatz
- Evaluation

Motivation

- “What is the time zone in Japan?”
- “Who plays Mary Jane in the Spiderman Movies?”
- “What currency does France accept?”

Motivation

- “What is the time zone in Japan?”
- “Who plays Mary Jane in the Spiderman Movies?”
- “What currency does France accept?”
- Wissendatenbanken enthalten Antwort
 - Entitäten (z.B. “*Japan*” und “*Mary Jane*”)
 - Relationen (z.B. “*country.currency_used*”
und “*location.time_zones*”)

Motivation

- SPARQL-Anfrage, welche Antwort findet auf “Who plays Mary Jane in the Spiderman Movies?”:

```
SELECT DISTINCT ?1 WHERE {  
  "Mary Jane"  
    film.film_character.portrayed_in_films ?0 .  
  ?0 film.performance.film "Spiderman" .  
  ?0 film.performance.actor ?1  
}
```

Motivation

- SPARQL-Anfrage, welche Antwort findet auf “Who **plays** Mary Jane in the Spiderman Movies?”:

```
SELECT DISTINCT ?1 WHERE {  
  "Mary Jane"  
    film.film_character.portrayed_in_films ?0 .  
  ?0 film.performance.film "Spiderman" .  
  ?0 film.performance.actor ?1  
}
```

→ Sehr umständlich!

Motivation

Aqqu

- Beantwortet natürlichsprachliche Fragen
- Maschinelles Lernen bildet zentralen Teil

Motivation

Aqqu

- Beantwortet natürlichsprachliche Fragen
- Maschinelles Lernen bildet zentralen Teil

Auto-sklearn

- System für automatisiertes maschinelles Lernen
- Optimiert Lernprozess ohne Parameterangabe

Motivation

Aqqu

- Beantwortet natürlichsprachliche Fragen
- Maschinelles Lernen bildet zentralen Teil

Auto-sklearn

- System für automatisiertes maschinelles Lernen
- Optimiert Lernprozess ohne Parameterangabe

➔ Ziel: Verbessern von Aqqu durch Auto-Sklearn

Auto-Sklearn

- Findet Konfiguration einer Pipeline:
 - Bis zu 3 Daten-Präprozessoren (von 4)
 - (Optional) Merkmalspräprozessor (von 14)
 - Ein Klassifikator (von 15)

Auto-Sklearn

- Findet Konfiguration einer Pipeline:
 - Bis zu 3 Daten-Präprozessoren (von 4)
 - (Optional) Merkmalspräprozessor (von 14)
 - Ein Klassifikator (von 15)
- ➔ Mehrdimensionales Optimierungsproblem (aus 110 Hyperparametern),
gelöst durch *Bayesian Optimization* (SMAC)

Auto-Sklearn

- Findet Konfiguration einer Pipeline:
 - Bis zu 3 Daten-Präprozessoren (von 4)
 - (Optional) Merkmalspräprozessor (von 14)
 - Ein Klassifikator (von 15)
- ➔ Mehrdimensionales Optimierungsproblem (aus 110 Hyperparametern),
gelöst durch *Bayesian Optimization* (SMAC)
- *Meta-learning*: Findet Start-Konfigurationen
- *Ensemble-Selection*: Kombinationen aus Modellen

Aqqu

- Natürlichsprachliche Frage → SPARQL-Anfrage

Schritte:

- 1) Finden von Entitäten (→ Punktzahlen)

Aqqu

- Natürlichsprachliche Frage → SPARQL-Anfrage

Schritte:

- 1) Finden von Entitäten (→ Punktzahlen)
- 2) Generierung von **Antwort-Kandidaten**
(= SPARQL-Anfrage)

Aqqu

- Natürlichsprachliche Frage → SPARQL-Anfrage

Schritte:

- 1) Finden von Entitäten (→ Punktzahlen)
- 2) Generierung von **Antwort-Kandidaten**
(= SPARQL-Anfrage)
- 3) Bewertung der Relationen (→ Punktzahlen)

Aqqu

- Natürlichsprachliche Frage → SPARQL-Anfrage

Schritte:

- 1) Finden von Entitäten (→ Punktzahlen)
- 2) Generierung von **Antwort-Kandidaten**
(= SPARQL-Anfrage)
- 3) Bewertung der Relationen (→ Punktzahlen)
→ Menge von Kandidaten mit jeweils Punktzahlen
- 4) Kandidaten → **Merkmalsvektoren**

Aqqu

- Finde besten Kandidaten mit maschinellem Lernen auf Basis der Merkmalsvektoren

5) Pruning (Verwerfe irrelevante Kandidaten)

Aqqu

- Finde besten Kandidaten mit maschinellem Lernen auf Basis der Merkmalsvektoren
- 5) Pruning (Verwerfe irrelevante Kandidaten)
- 6) Ranking (Ordne Kandidaten nach Relevanz)
 - Punktweise: Bewerte jeden Kandidaten einzeln
 - Paarweise: Vergleiche Paare von Kandidaten (Trainiere auf Differenzen der Merkmalsvektoren)

Ansatz

Verbessere Ranking mit Auto-Sklearn:

- SMAC und Ensemble-Selection evaluieren Konfigurationen bzw. Ensembles nach gewählter Metrik
- Gesucht: Metrik, welche nach der Anzahl korrekt beantworteter Fragen maximiert

Ansatz

- Standard-metrik: Accuracy = Anteil korrekt beantworteter Datenpunkte
- Datenpunkte = Antwort-Kandidaten bzw.
Vergleiche zwischen Antwort-Kandidaten
- Datenpunkte gehören zu verschiedenen Fragen
→ Accuracy \neq % korrekt beantwortete Fragen

Ansatz

Punktweise:

- *Question-Accuracy-Exact*

Frage richtig gdw. korrekter Kandidat die beste Bewertung unter allen Kandidaten der Frage hat

- *Question-Accuracy*

Frage richtig gdw. alle Kandidaten dazu richtig klassifiziert

Ansatz

Punktweise:

- *Question-Accuracy-Exact*

Frage richtig gdw. korrekter Kandidat die beste Bewertung unter allen Kandidaten der Frage hat

- *Question-Accuracy*

Frage richtig gdw. alle Kandidaten dazu richtig klassifiziert

Paarweise:

- Nur *Question-Accuracy*

Ansatz

Punktweise:

- *Question-Accuracy-Exact / Question-F1-Exact*
Frage richtig gdw. korrekter Kandidat die beste Bewertung unter allen Kandidaten der Frage hat
- *Question-Accuracy*
Frage richtig gdw. alle Kandidaten dazu richtig klassifiziert

Paarweise:

- Nur *Question-Accuracy*

Evaluation

- Datensätze mit Antworten und Fragen
 - *Free917*: 917 Fragen, Accuracy
 - *WebQuestions*: 5,810 Fragen, mittlerer F1
(Hier: Teilmenge von 1059 Fragen)

Evaluation

	Free917 (Accuracy)		WebQuestions (F1)	
	Punktweise	Paarweise	Punktweise	Paarweise
Nur Aqqu	59,06	65,22	42,46	46,17
Standard-Accuracy	66,30	66,24	43,82	46,92
Question-Accuracy	62,32	66,61	43,17	46,72
Question-Acc- / F1-Exact	61,05	-	40,39	-

Evaluation

	Free917 (Accuracy)		WebQuestions (F1)	
	Punktweise	Paarweise	Punktweise	Paarweise
Nur Aqqu	59,06	65,22	42,46	46,17
Standard-Accuracy	66,30	66,24	43,82	46,92
Question-Accuracy	62,32	66,61	43,17	46,72
Question-Acc- / F1-Exact	61,05	-	40,39	-

Evaluation

	Free917 (Accuracy)		WebQuestions (F1)	
	Punktweise	Paarweise	Punktweise	Paarweise
Nur Aqqu	59,06	65,22	42,46	46,17
Standard-Accuracy	66,30	66,24	43,82	46,92
Question-Accuracy	62,32	66,61	43,17	46,72
Question-Acc- / F1-Exact	61,05	-	40,39	-

Evaluation

	Free917 (Accuracy)		WebQuestions (F1)	
	Punktweise	Paarweise	Punktweise	Paarweise
Nur Aqqu	59,06	65,22	42,46	46,17
Standard-Accuracy	66,30	66,24	43,82	46,92
Question-Accuracy	62,32	66,61	43,17	46,72
Question-Acc- / F1-Exact	61,05	-	40,39	-

Evaluation

- Paarweise ist immer besser als Punktweise
- Auto-Sklearn verbessert unter Standard-Metrik immer die Ergebnisse
- Angepasste Metriken liefern fast immer schlechtere Ergebnisse, beim bevorzugten paarweisen Ranking fast die selben