



Machine Learning

Live Demo @ Tagung Betrieb/Arge Daten

Warum hilft Machine Learning bei der Suche nach Namen?

Eine Suche nur nach Namen bringt im Kontext bei bestimmten Namen falsche Treffer.

Beispiel 1: Monat August

„August hat im August Geburtstag“

Hier muss das neuronale Netz den Namen August vom Monat August unterscheiden können. Dies ist nur anhand der Bewertung des ganzen Satzes möglich, das neuronale Netz muss diesen interpretieren können.

Beispiel 2: Programmiersprache „Ruby on Rails“ vs. Name Ruby

„Konrad interessiert sich für Ruby on Rails.“

Hier darf Ruby nicht als Treffer ausgewertet werden, da es sich um die Programmiersprache handelt.

Ablauf Machine Learning

- 1.) Auswahl geeigneter Datenbestände mit syntaktischer Auszeichnung von Zeitungsberichten oder Literaturtexten - mind. 10.000 Datensätzen
- 2.) Aufbereitung der Trainingsdaten
 - Sentence Detection
 - Jeder Satz in einer Zeile.
 - Tokenization
 - Zerlegung des Textes (Beistriche, -, ..)
 - Part of Speech Tagging (POS)
 - Erkennung der Satzstruktur (z.B. Substantive, Verben, Adjektive etc.)
 - Name Finder
 - Erkennen von Namen
- 3.) Trainieren & Testen eines Modells
- 4.) Verwenden des trainierten Modells
- 5.) Evaluieren des Modells

Maschinen Trainieren – Beispiel August (Monat/Name)

TRAININGSDATEN (MIT TAGS):

<START:person> August <END> geht gerne in die Schule .
Im August ist es oft sehr heiß .
Wir fahren jeden Sommer im August nach Griechenland auf Urlaub .
Nächste Woche besuchen wir <START:person> August <END> zu Hause .
Der August (Erntemonat, Ährenmonat, Sichelmonat, Ernting; lateinisch augustus) ist der achte Monat des Jahres im gregorianischen Kalender .
Der August hat 31 Tage und wurde im Jahre 8 v. Chr. nach dem römischen Kaiser Augustus benannt, da er in diesem Monat sein erstes Konsulat angetreten hat .
Unter Kaiser Commodus wurde der Name des Monats ihm zu Ehren in Commodus geändert, nach dem Tod des Kaisers erhielt der Monat seinen alten Namen zurück .
<START:person> August <END> ist ein beliebter Name .
<START:person> August <END> geht gerne in die Schule .
<START:person> August <END> und <START:person> Natalie <END> gehen in die gleiche Klasse .

TESTDATEN

August ist ein guter Schüler .
Im August fahren wir oft in den Süden .
Warum nehmen wir August nicht mit auf die Klassenfahrt ?
August , du bist doch noch klein!
Im August reifen die Äpfel.

EVALUIERUNGSDATEN (MIT TAGS)

<START:person> August <END> ist ein guter Schüler .
Im August fahren wir oft in den Süden .
Warum nehmen wir <START:person> August <END> nicht mit auf die Klassenfahrt ?
<START:person> August <END> , du bist doch noch klein!
Im August reifen die Äpfel.

Maschinen trainieren

ERSTE SCHRITTE

- 1.) Testlauf mit Trainingsdaten für August
Hier wird ein Testlauf mit den o.a. Trainingsdaten durchgeführt.
Welche Wörter könnte die Maschine erkennen?
Werden alle Wörter erkannt?
- 2.) Testlauf mit Basismodell (250.000 Sätze)
Jetzt wird ein Testlauf mit dem Basismodell durchgeführt.
Werden alle Wörter erkannt?
- 3.) Testlauf mit Basismodell + angepasste Trainingsdaten
Jetzt werden die Trainingsdaten des Basismodells um die angepassten Trainingsdaten erweitert.
Werden alle Wörter erkannt?

WEITERFÜHRENDE AUFGABEN

Geben Sie weitere Testdaten ein und vergleichen Sie das Verhalten des Basismodells mit dem erweiterten Modell?

Geben Sie weitere Testdaten ein und versuchen Sie Kombinationen zu finden, die auch das erweiterte Modell nicht erkennt.

Geben Sie weitere Trainingsdaten ein und testen Sie erneut. Können Sie das Modell so anpassen, dass die Testdaten gefunden werden?

Wie verhält sich ein nur mit den Trainingsdaten trainiertes Modell?

Experimentieren Sie mit den Trainingsdaten, was passiert wenn Sie dies reduzieren oder erweitern?

Maschinen Trainieren – Beispiel Ruby (Name vs. Programmiersprache)

TRAININGSDATEN (MIT TAGS):

<START:person> Ruby <END> ist eine Programmiererin .
<START:person> Ruby <END> ist gerne vorne dabei .
<START:person> Ruby <END> geht es heute gut .
Er sagt zu <START:person> Ruby <END>, dass er nicht auf Besuch kommen kann.
Ruby on rails aber nicht .
Ruby on Rails ist eine Programmiersprache .
<START:person> Ruby <END> ist im englischen Sprachraum seit dem 19. Jahrhundert ein vorwiegend weiblicher Vorname .
<START:person> Ruby Langford Ginibi <END> (geb. 26. Januar 1934 in Coraki, New South Wales; gest. 1. Oktober 2011 in Sydney)[1] war eine australische Autorin .
Als Angehörige des Aborigines-Volkes der Bundjalung und Historikerin befasste <START:person> Ruby Langford Ginibi <END> sich insbesondere mit der Geschichte und der Kultur der ursprünglichen Einwohner Australiens.

TESTDATEN

Ruby ist eine Programmiererin .
Was ist bloß mit Ruby los ?
Ruby ist eine gute Schülerin .
Was macht Ruby heute ?
Ruby on Rails ist schon interessant .
Ruby ging mit Manuela spazieren .
Ruby Hall ist eine australische Schauspielerin.

EVALUIERUNGSDATEN (MIT TAGS)

<START:person> Ruby <END> ist eine Programmiererin .
Was ist bloß mit <START:person> Ruby <END> los ?
<START:person> Ruby <END> ist eine gute Schülerin .
Was macht <START:person> Ruby <END> heute ?
Ruby on Rails ist schon interessant .
<START:person> Ruby <END> ging mit <START:person> Manuela <END> spazieren .
<START:person> Ruby Hall <END> ist eine australische Schauspielerin .