

Opleiding : Tekst mining in de praktijk

Praktijkcursus - 3d - 21u00 - Ref. MMD

Prijs : 2010 € V.B.

Datamining beperkt tot tekstuele gegevens - text mining - wordt steeds vaker gebruikt in bedrijven. Het kan bijvoorbeeld worden gebruikt om producten te classificeren op basis van opmerkingen van consumenten. Je zult algoritmen en tools voor text mining toepassen op paradigmatische voorbeelden.

Pedagogische doelstellingen

Aan het einde van de training is de deelnemer in staat om:

- ✓ Tekstuele statistische methoden begrijpen
- ✓ Kenmerkenextractie implementeren uit tekstuele gegevens
- ✓ Selecties en rankings maken van grote hoeveelheden tekstuele gegevens
- ✓ Een classificatiealgoritme kiezen
- ✓ De voorspellende prestatie van een algoritme evalueren

Doelgroep

AI-engineers/projectmanagers, AI-consultants en iedereen die tekst mining voor machine learning en deep learning wil ontdekken.

Voorafgaande vereisten

Goede kennis van statistiek. Goede kennis van machine learning en deep learning. Vereiste ervaring.

Opleidingsprogramma

1 Traditionele benaderingen van text mining

- API's voor het ophalen van tekstuele gegevens.
- Tekstuele gegevens voorbereiden volgens het probleem.
- Ophalen en verkennen van het corpus van teksten.
- Accenten en speciale tekens verwijderen.
- Stemming, lemmatisering en verwijdering van verbindingswoorden.
- Alles samenbrengen om gegevens op te schonen en te standaardiseren.

Praktisch werk

Documenten zoeken, voorbereiden, transformeren en vectoriseren van gegevens in DataFrame.

DEELNEMERS

AI-engineers/projectmanagers, AI-consultants en iedereen die tekst mining voor machine learning en deep learning wil ontdekken.

VOORAFGAANDE VEREISTEN

Goede kennis van statistiek. Goede kennis van machine learning en deep learning. Vereiste ervaring.

VAARDIGHEDEN VAN DE CURSUSLEIDER

De deskundigen die de cursus leiden zijn specialisten op het betreffende vakgebied. Zij werden geselecteerd door onze pedagogische teams zowel om hun vakkennis als hun pedagogische vaardigheden voor elke cursus die zij geven. Zij hebben minstens vijf tot tien jaar ervaring in hun vakgebied en oefenen of oefenden verantwoordelijke bedrijfsfuncties uit.

BEOORDELINGSMODALITEITEN

De cursusleider beoordeelt de pedagogische vooruitgang van de deelnemer gedurende de gehele cursus aan de hand van meerkeuzevragen, praktijksituaties, praktische opdrachten, ...
De deelnemer legt ook van tevoren en naderhand een test af ter bevestiging van de verworven kennis.

2 Feature engineering voor tekstrepresentatie

- De syntaxis en structuur van de tekst begrijpen.
- De modellen Bag of Words en Bag of N-Grams.
- Het TF-IDF model, Transformer en Vectorizer.
- Het Word2Vec model en implementatie met Gensim.
- Het GloVe-model.
- Het FastText-model.

Praktisch werk

Bewerkingen opzetten om kenmerken uit tekstgegevens te extraheren om classificaties uit te voeren.

3 Tekstgelijkenis en classificatie zonder toezicht

- De essentiële concepten van gelijkenis.
- Term similariteitsanalyse: Hamming, Manhattan, Euclidische en Levenshtein afstanden.
- Gelijksoortigheidsanalyse van documenten.
- Okapi BM25 en de ranglijst.
- Classificatiealgoritmen zonder toezicht.

Praktisch werk

Bouw een aanbevelingssysteem voor vergelijkbare producten op basis van de beschrijving en inhoud van de producten die je hebt gekozen.

4 Gesuperviseerde tekstclassificatie

- Gegevens voorbereiden en standaardiseren.
- Classificatiemodellen.
- Multinomiale Naïve Bayes.
- Logistische regressie. Ondersteunende vector machines.
- Random Forest. Gradient Boosting Machines.
- Evaluatie van classificatiemodellen.

Praktisch werk

Implementatie van gecontroleerde classificaties op verschillende datasets.

5 Natuurlijke taalverwerking en diep leren

- NLP-bibliotheken: NLTK, TextBlob, SpaCy, Gensim, Pattern, Stanford CoreNLP.
- Deep learning-bibliotheken: Theano, TensorFlow, Keras.
- Natuurlijke taalverwerking en terugkerende neurale netwerken.
- RNN en het lange-termijngeheugen. Bidirectionele RNN-modellen.
- Sequentie-naar-sequentie modellen.
- Vragen en antwoorden met RNN-modellen.

Praktisch werk

Bouw een RNN om een nieuwe tekst te genereren.

PEDAGOGISCHE EN TECHNISCHE MIDDELEN

- De gebruikte pedagogische middelen en cursusmethoden zijn voornamelijk: audiovisuele hulpmiddelen, documentatie en cursusmateriaal, praktische oefeningen en correcties van de oefeningen voor praktijkstages, casestudies of reële voorbeelden voor de seminars.
- Na afloop van de stages of seminars verstrekt ORSYS de deelnemers een evaluatievragenlijst over de cursus die vervolgens door onze pedagogische teams wordt geanalyseerd.
- Na afloop van de cursus wordt een presentielijst per halve dag verstrekt, evenals een verklaring van de afronding van de cursus indien de stagiair alle sessies heeft bijgewoond.

TOEGANGSMODALITEITEN EN TERMIJNEN

De inschrijving dient 24 uur voor aanvang van de cursus plaatsgevonden te hebben.

TOEGANKELIJKHEID VOOR MINDERVALIDEN

Is voor u speciale toegankelijkheid vereist? Neem contact op met mevr. FOSSE, contactpersoon voor mindervaliden, via het adres psh-accueil@ORSYS.fr om uw verzoek en de haalbaarheid daarvan zo goed mogelijk te bestuderen.