

Deep and Statistical Learning WS 21-22

Analysis of Berlins Airbnb Data using statistical and deep learning methods.

Am besten dieser File in HTML angucken (wenn nicht in GitHub). To edit -> Rstudio.

TO DO

Orga

- Define “our” deadlines
- a concrete scope of the project (to avoid working in vain).

Data preprocesing (Sebi/Tobi)

- upload cleaned datasets

Turn to metric

- seasson (summer/winter variable).
- Tobi runs script to create custum seasson variable.
-

Bild (Tobi)

- Liste mit detected objekten erstellen und auf r-projekt

Farbe/helligkeit

- Möglichkeit mit pillow package in python
- Kelvin si colour
- Lumen unit

Bildqualität mit cnn

- Email bilder scraping mit matse klären
- drei Gruppen haben es definitiv geschafft alle Bilder zu scrapen.
- wenn dann per privaten email würde ich sagen

Picture scene analsis - das gleiche skript nutzt matse. - muss lokal aufgebaut werden dann sollte es laut matse klappen

Text i.e reviews

NLP - Word stemming von isi projekt auf rviews anwenden - Vielleicht intervale bilden - Seb macht zuerst noch weiter und deligiert dann

Analisis

Regression (Sebi)

- Basic regression with available variables.
- Update correlation plot for clarity.

Regularization/Boosting (Sebi/Tobi)

- 6. Scale up and develop a model that overfits.
 - ?
- 7. Regularize a model and tune hyperparameters
 - ? ### Report (15 Seiten) (Sebi/Tobi)

Literature (Sebi/Tobi) Citavi - Für picture detection websites direkt auf citavi - Citavi gruppieren und sortieren - Reference website und save als pdf oder über local file also pdfs runterladen - Alle angeschauten modelle zu detection und segmentation auf citavi in die gliederung die seb eröffnet - “Wissen” fenster auf citavi für cnn weitermachen

Write (Sebi/Tobi)

- .Rmd file with report structure of the project based on rene’s 7 steps. This is going to be the final version.
- .bib file export path.

Best practices

- Für Tobi: Branching in GIT
- Citavi: Add references and immediately document the corresponding PDF

Calendar

Global

- 17.01.21: QA
- 20-21.02.21 09:00-12:00: Group presentation
- 06.03.22 23:59: Hand in Term paper of 15 pages and Challenge-submittal
- 31.03.22: Hand in Revised term paper

Tobi

- 20.02.21 Deepl präsi
- 25.02.21 Spatial
- 06.03.21 DeepL paper first
- 14-18.03.21 Latex (What is this? @Sebi.15.01)
- 31.03.21 Deepl paper revised and Discrete choice

Sebi

Anmeldedaten für DeepL Videos bei Manning (DeepLearning with R in motion):

- tobirinnert@gmail.com
- DeepLearningWS2021

Data

siehe .Rmd in Data folder.

Ergebnisse QA 17.01.21

Ergebnisse Anton Treffen: Stichwörter ohne Zuordnung (Vielleicht sagen die dir was Tobi)

- Sentiment analysis - Text nicht nehmen müssen
- Alles miteinbeziehen in 1 NN = Model schlecht.
 1. Werte Strukturierter.
 2. Objekt detektieren.
- PCR
- Bilderscrapping
- Multilevel Set
- Image Regression
- Mittel CNN
 4. Fit testing
 5. Komplizität = Overfitting.

Mixed Data Multiimput CNN

- Nicht gut erklärbar.
- Boosting
- GBT2

Notizen für die QnA vom 15.11

- Ein wichtiger Teil ist Daten einlesen
- Minimum 3000 Beobachtungen verwenden
- Histogramme plotten und Korrelationen berechnen als Einstieg für Paper
- In der QnA wirkte es als wäre es gut wenn wir noch mehr Bilder aus dem Internet scrapen
- Bilder und comments analysieren wären gut
- Alle datentypen zusammen zu bringen ist die Stärke von neuronalen Netzen also sollten wir das machen
- Interpretierbarkeit der Ergebnisse soll möglich sein. Also keine reine Blackbox. Stichwort statistisches lernen
- Bis zur nächsten Session sollen wir mindestens die Daten runterladen und am besten die Bilder scrapen.
- Bilder sollen in reduzierter Pixelanzahl (64x64 oder so) runtergeladen werden
- Es gibt eine letzte Stadt mit der wir dann unsere Modelle gegeneinander antreten lassen
- Modelle sollen so geschrieben werden dass es sie auf die neue Stadt angepasst werden kann