Kaggle, Avito Challenge and Deep Learning

Thiago G. Martins

01/2019

Kaggle

- Start by offering machine learning competitions.
- Now owned by Google.
  - Offers a public data platform
  - Cloud-based workbench for data science
  - AI education.
- Positive
  - Nice place to learn trends
- Negative
  - Competition often leads to over engineering

kaggle™

Avito Challenge

- Avito is the Russia's largest classified advertisements website.
- Aim is to predict demand for an online advertisement based on:
  - description (title, description, images, etc.)
  - context (geography, categories)
  - historical demand for similar ads in similar contexts.
- What makes this interesting:
  - Diversity of data type: numerical, categorical, text and image
  - Text is in russian

Text and Image

Well-Taken, Authentic Photos

Too Glossy

Authentic

Poor Quality

Believable and Informative Description Copy

Description:
AMAZING WATCH FOR SALE!!!!

DON'T MISS THIS DEAL. IT'S THE DEAL OF THE CENTURY!!

Unlikely

Description:
I have an adjustable Chaleur D'Animale Watch for sale.

It's never been worn and still in the original box. Battery included.

Informative

Description:
fancy watch for sale

no low ball offers, cash and carry

Poor Quality

Deep Learning lectures

- **Lecture 1**
  - General framework
  - Fully-connected layer
  - Good for learning global patterns
- **Lecture 2**
  - Convolutional neural network (CNN) and related
  - Hierarchies of patterns that are translation-invariant
  - Image oriented
- **Lecture 3**
  - Recurrent neural network (RNN) and related
  - Able to process a data point sequentially, keeping ordering
  - Text oriented

Avito Challenge - Winning solution