Biomedical Natural Language Processing: Applications and Software

UCL ICH Coding Club

Ferran Gonzalez Hernandez

25/03/2020

Natural Language Processing (NLP)

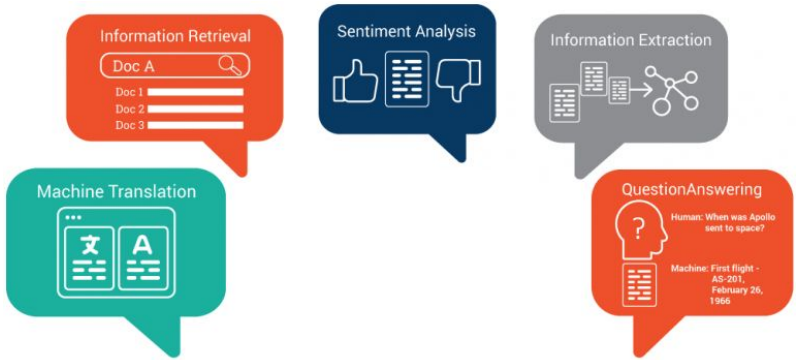
- What is Natural Language Processing?

  Terminology: text mining & NLP

- Machine Learning & AI

- Applications?

  Speech Recognition
  Email filtering
  Search engines
  Machine Translation
  Question answering
  Sentiment analysis
  ....



Biomedical NLP

- Why can NLP be helpful?

  Biomedical Literature
  30M > records in PubMed


Macedonian Biomedical Scientists Are Not Adequately Represented in BiomedExperts Database

Structure Knowledge

- Search engines, document classification, topic modelling, summarisation

Infer knowledge

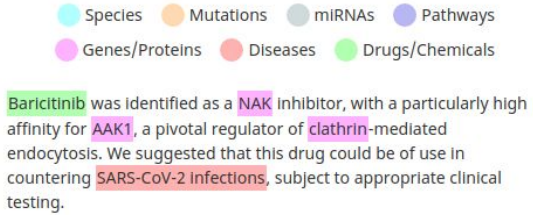
- Relationships between genes, proteins, diseases and drugs
- Hypothesis generation through knowledge graphs

Common Tasks

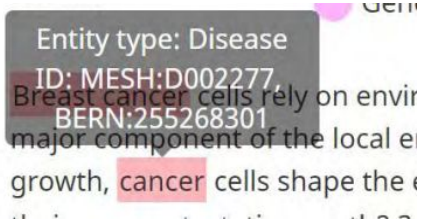
Information Retrieval

Document Classification
Topic Modelling

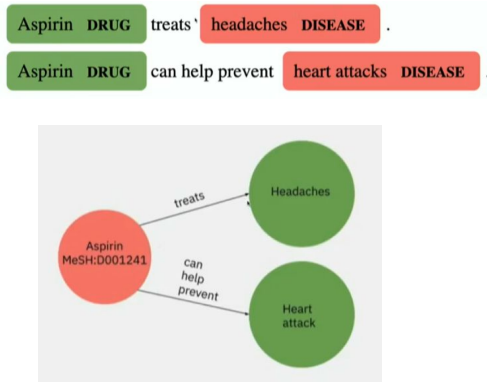
Named-Entity Recognition (NER)



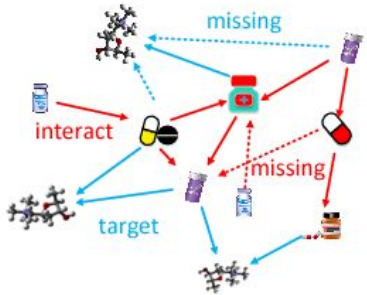
Entity Linking



Relationship extraction



Knowledge graph construction

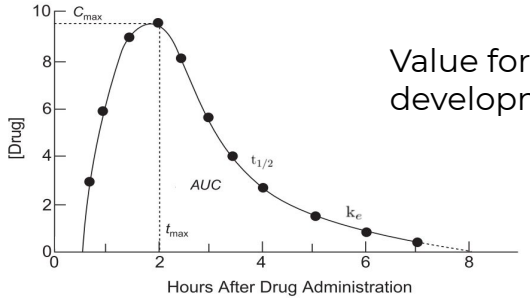


NLP & COVID-19

https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30132-8/fulltext

Own research

- Pharmacokinetics



Value for drug development?

- Parameters derived from Clinical Trials:

1. Identify papers reporting this information (Document Retrieval)
2. Locate mentions of parameters (NER) and contextual information (sentence classification)
3. Extract associated values (Relationship extraction)
4. Normalisation

Pharmacokinetics and Tissue Penetration of Tazobactam and Piperacillin in Patients Undergoing Colorectal Surgery

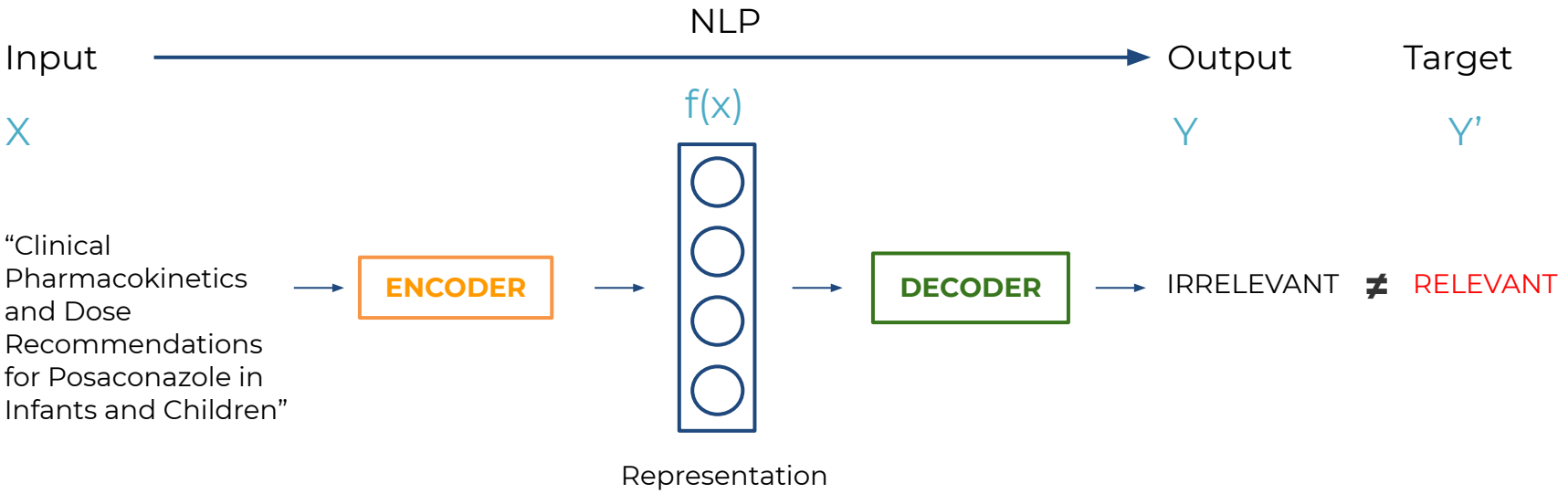
M. KINZIG,[1] F. SÖRGEL,[1] B. BRISMAR,[2] AND C. E. NORD[3]*

IBMP-Institute for Biomedical and Pharmaceutical Research, Nuremberg-Heroldsberg, Germany,[1] and Departments of Surgery[2] and Microbiology,[3] Huddinge University Hospital, Karolinska Institute, Stockholm, Sweden[2]

Received 22 August 1991/Accepted 18 June 1992

The pharmacokinetics of tazobactam and piperacillin in plasma and different tissues after a 30-min intravenous infusion of 4 g of piperacillin and 0.5 g of tazobactam were investigated in 18 patients who underwent elective colorectal surgery. Serial blood samples were collected for up to 6 h after the initiation of the infusion. The types of tissue collected were fatty tissue, muscle, skin, appendix, and intestinal mucosa (proximal and distal). On the basis of concentrations in plasma, the following pharmacokinetic parameter values were obtained (values are means ± standard deviations): maximum concentration of drug in serum, tazobactam, 27.9 ± 7.67 μg/ml; piperacillin, 259 ± 81.8 μg/ml; time to maximum concentration of drug in serum, azobactam, 0.51 ± 0.03 h; piperacillin, 0.51 ± 0.03 h; area under the concentration-time curve, tazobactam, 47.6 ± 13.3 μg · h/ml; piperacillin, 361 ± 80.3 μg · h/ml; clearance, tazobactam, 188 ± 52.3 ml/min; piperacillin, 194 ± 42.9 ml/min; half-life, tazobactam, 1.42 ± 0.32 h; piperacillin, 1.27 ± 0.24 h; apparent volume of distribution, tazobactam, 0.31 ± 0.07 liter/kg of body weigh ; piperacillin, 0.29 ± 0.06 liter/kg; volume of distribution at steady state, tazobactam, 0.28 ± 0.04 liter/kg; piperacillin, 0.25 ± 0.05 liter/kg. The concentrations of tazobactam and piperacillin in fatty tissue and muscle tissue were 10 to 13 and 18 to 30% of the levels in plasma, respectively. In skin, the concentrations of piperacillin were 60 to 95% of the levels in plasma, whereas the concentrations of tazobactam in plasma were 49 to 93% of the levels in skin tissue. The mean concentrations of tazobactam in the investigated gastrointestinal tissues (appendix, proximal and distal mucosa) exceeded levels in plasma after 1 h, while piperacillin showed a mean penetration into these

NLP in a nutshell



TYPES

Titles
Paragraphs
Documents
Text messages
...

Rule-based
Learnt

Symbolic
Sparse
Dense
...

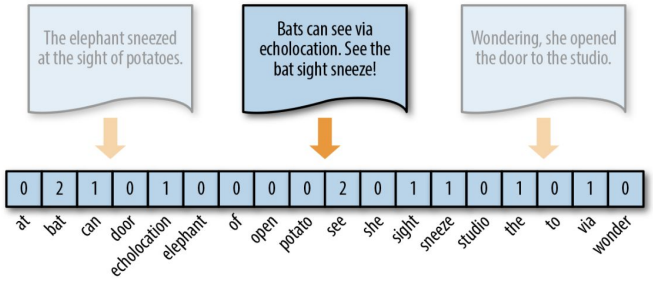
Rule-based
Logistic Regression
Decision Trees
Neural Networks
.....

Labels
Sentences
Graphs
Trees
...

Supervised
Unsupervised
Semi-supervised
...

BAG OF WORDS

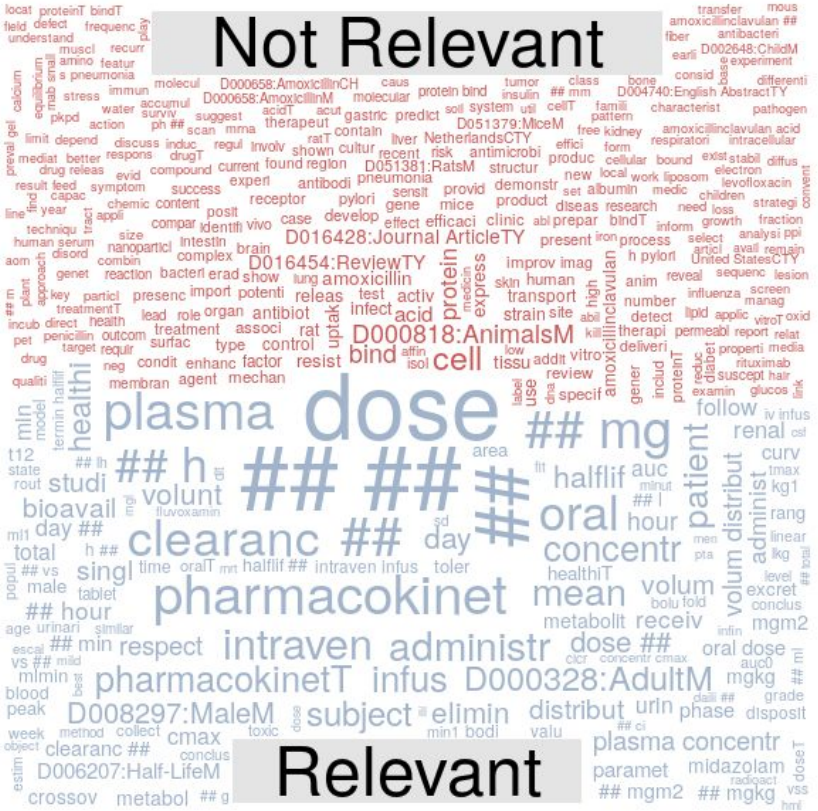
How can we represent text numerically?



Let's try it

https://github.com/UCLichCodeClub/Meeting_ICH_25Mar2020/blob/master/DocClassification.ipynb

Advantages and Limitations?

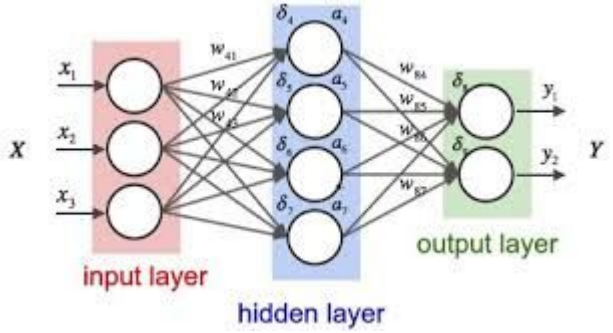


BETTER REPRESENTATIONS? WORD EMBEDDINGS

Neural Networks and Deep Learning

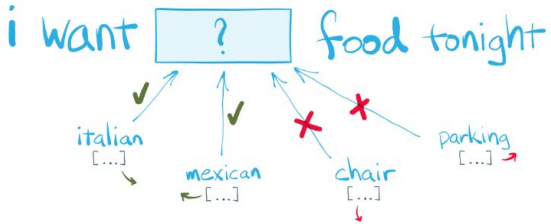
What are neural networks?

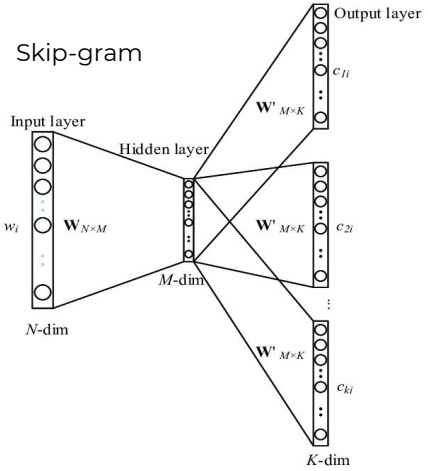
Why are they useful?



Main applications

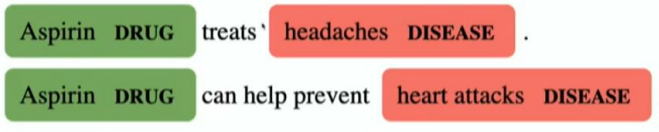
- Language modelling
- **Unsupervised** training
- Ultimate goal: learn the **weights** of the model given an **auxiliary task**





Named-Entity Recognition (NER)

Task



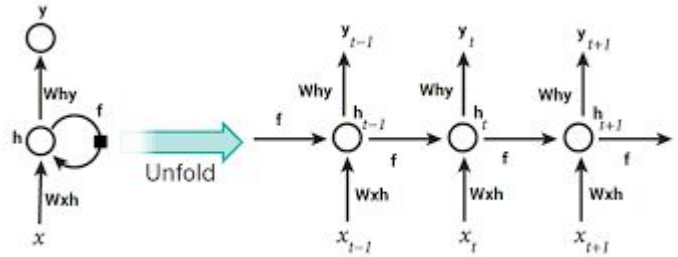
Rule/dictionary-based approaches

Bag of words + ML classifiers

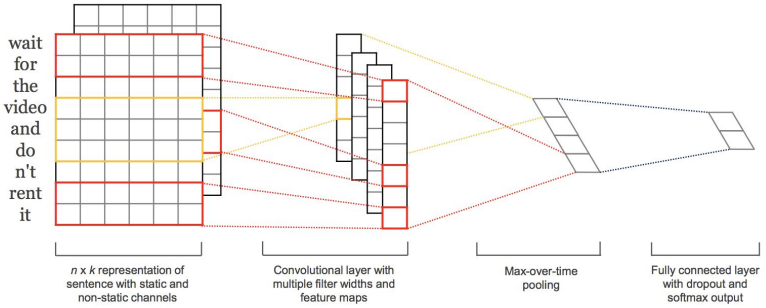
Combination of Embeddings + ML algorithms

Common neural architectures:

Recurrent Neural Networks (RNN)



Convolutional Neural Networks (CNNs)



Some interesting resources

Biomedical NLP
https://scispacy.apps.allenai.org/

State-of-the-art Language model for automatic text completion:
https://transformer.huggingface.co/doc/gpt2-large

Thank you for your attention

ferran.hernandez.17@ucl.ac.uk

fhernandez@turing.ac.uk