





Healthcare Text Mining – state of the play –

Goran Nenadic













The rise of healthcare text mining





State of the art



The rise of machine learning

Methods – clinical NLP



Machine learning models

Methods – clinical NLP



Methods - NIPS



Methods - NIPS



State of the art

- Dictionary and rule-based approaches are not dead?
 - E.g. the latest n2c2 clinical trial track
 - A number of tasks in some clinical subdomains are doable via rules
 - e.g. radiology reports?
 - e.g. medication prescriptions?
 - But many are not
 - e.g. organisation names, professions?
 - e.g. patient-generated data

State of the art

- There is no a single clinical language
- Deal with very different domains
 e.g. mental health vs. cardiology notes
- Deal with very different datasets
 - Clinical notes, letters, reports, hand-over notes
 - Patient-generated data
 - Literature

State of the challenges

interpretability acvess exhaustively terminologies transfer open legal acces sensitive analysis interpretation licences privacy setting sharing of the privacy setting e D relevancy feature of sh related of sauton aring save docum nois Definition account of the second s ISSU

Which data used?

cris i2b2/n2c2 2013 dataset alberto semeval snpphena setmedline sharedvtasks da pubmed biocreative biocreative institutional mimic-iiibioasq JOSC

Using patient data?

- What do patients say?
- Citizens' jury in June 2018
 - 3 days, 18 citizens, 7 witnesses



The jury was slightly more cautious about using free-text than coded health data for research but were nevertheless broadly supportive as long as there was a transparent process for patients to easily opt out of their data being used in this way.

http://healtex.org/jury/







Other issues

- Interpretability
 - Black-box models
- Evaluation metrics
 - Intrinsic vs. extrinsic
- Provenance, sharing methods and models
 - Reproducibility
 - Reusability



HOME HEALTAC-2019 EVENTS PARTNERS NEWS ABOUT CONTACT

healtex.org