INFORMATICS PROFESSIONALS. LEADING THE WAY.

Detecting Tweets Mentioning Drugs with a Deep Neural Network Ensemble

Deep Learning

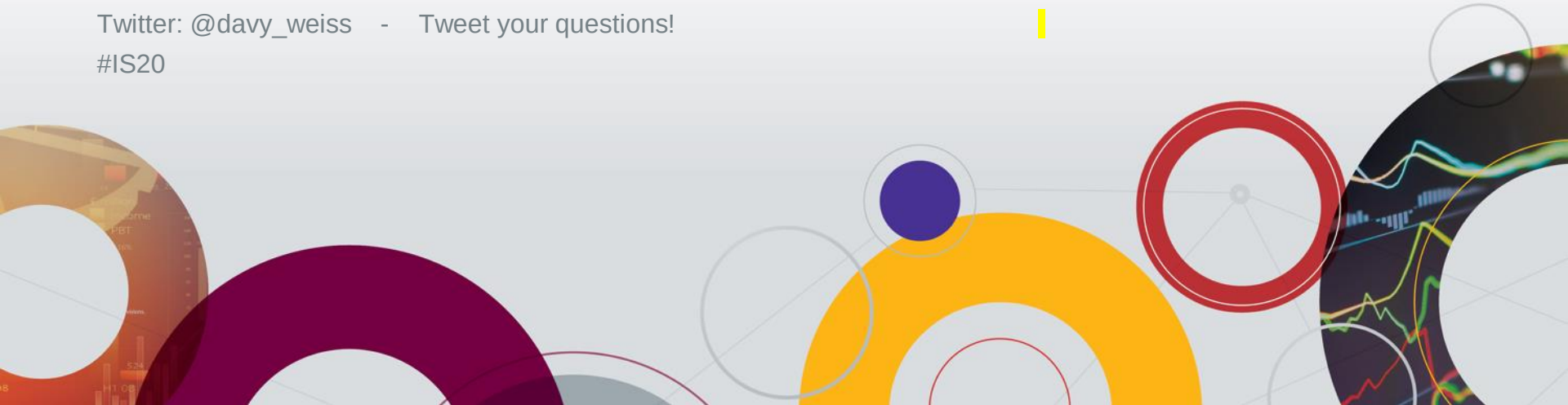
VS07

Davy Weissenbacher

University of Pennsylvania

Twitter: @davy_weiss - Tweet your questions!

#IS20



Disclosure

I and my wife have NO relevant relationship with commercial interest to disclose.

Learning Objectives

After participating in this session the learner should be able to:

- Define the problem of drug names detection in tweets
- Explain the limitation of the lexical search method
- List the main methods for detecting drug names
- Explain the best approach based on deep neural networks
- Report on the current performance of drug names detection when applied on 112 home timelines

Why extracting tweets mentioning drugs?

- **Tweets mentioning drugs** useful for:
  - Syndromic surveillance
  - Pharmacovigilance
  - Monitoring drug abuse
- Typically, collections use drug names as keywords – inherent bias
- In a ‘natural’ timeline, these tweets **are rare**, only 0.26% in our corpus
- Lexical searches fail:
  - Misspellings
  - Ambiguity
- **Goal: Extract drug mentions from user timelines (an extremely sparse corpus) using Deep Neural Networks**

Detecting Tweets Mentioning Drugs

- Drug definition
  - Drug product as defined by the FDA (prescription & over the counter)
  - Dietary supplements
- Resolving a **binary classification** task
  - Lola may has a stye, or pink eye. Doc recommends warm compress => 0 (not mentioning a drug)
  - Meet Mr and Ms Lexapro… Garanteed fidelity => 1 (mentioning a drug)

⇨ Kusuri, 薬, demo:
https://hlp.ibi.upenn.edu/kusuri/home/

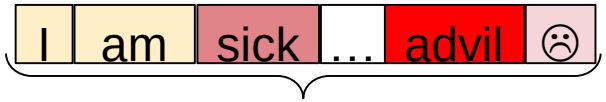
⇨ The competition is now open @ #SMM4H, task 1:
https://healthlanguageprocessing.org/smm4h-sharedtask-2020/

UPennHLP Timeline Corpus

- **Corpus**: 112 timelines (98,959 tweets) of women twitting during their pregnancy
- **Annotation**: 282.5 hours of annotation, IAA k=0.88
- **Corpus split**:
  - 70% for training (181 positive, 69,091 negative examples)
  - 30% for evaluation (77 positive, 29,610 negative examples)
- Only 258 tweets mentioned a drug, *i.e.* 0.26% of the tweets

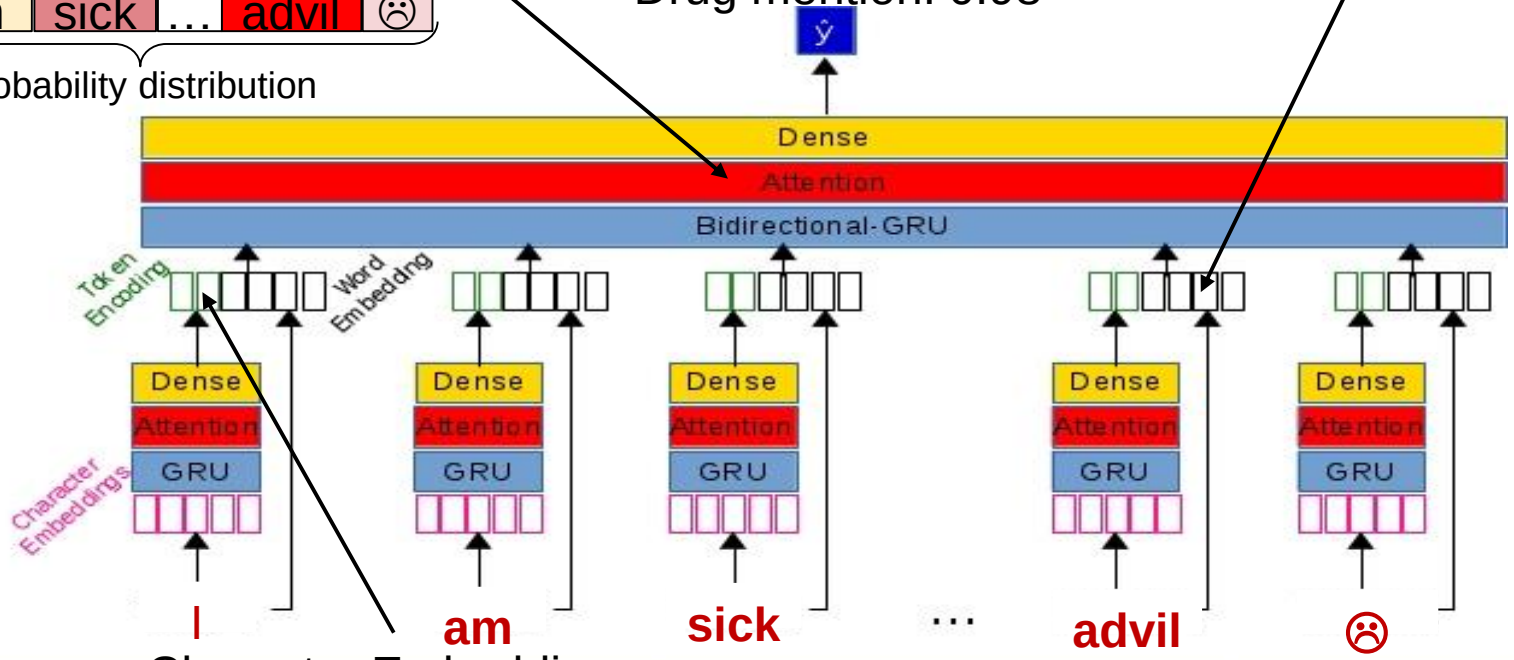
A DNN for Drug Detection

Attention layer helps
To focus on important words



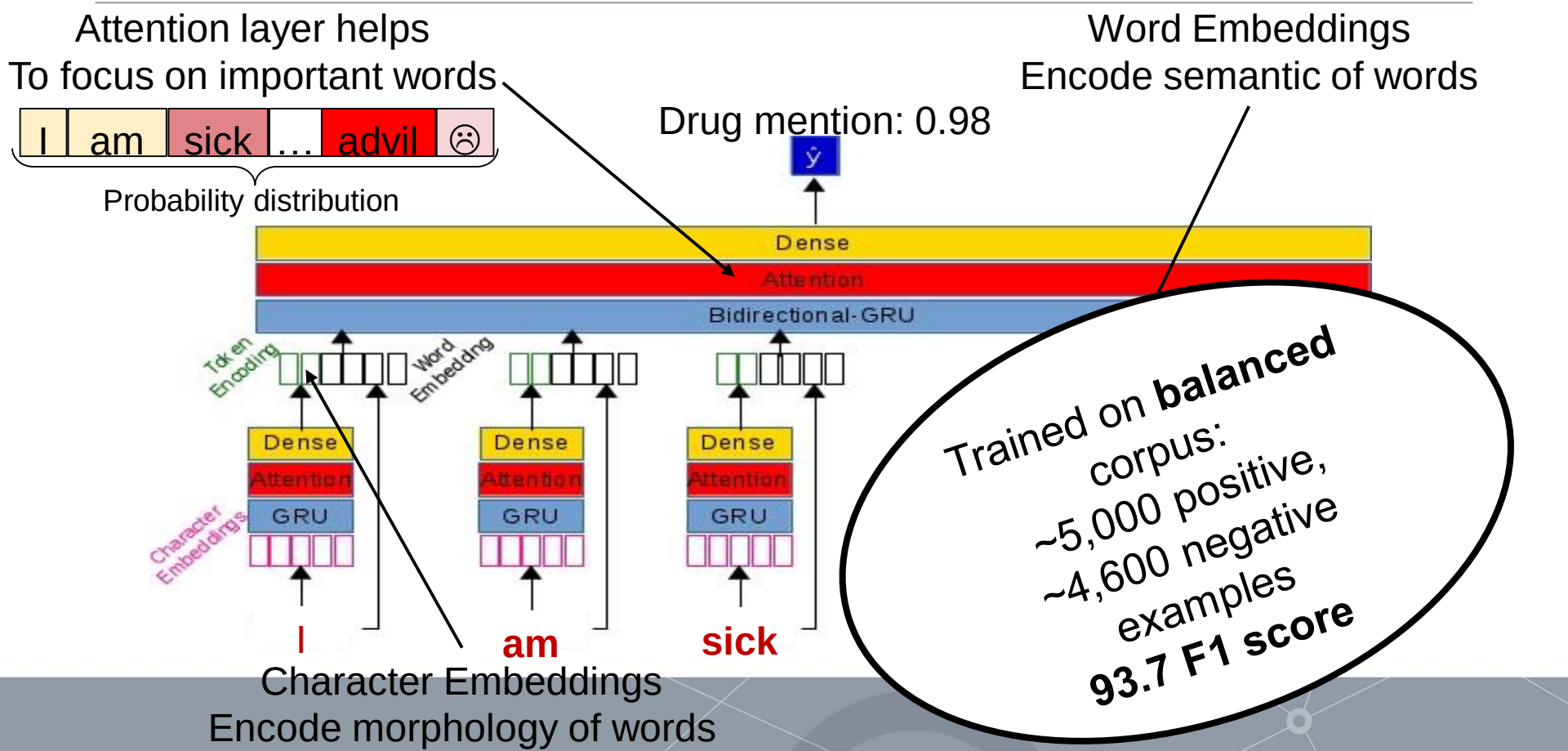
Word Embeddings
Encode semantic of words

Drug mention: 0.98



Character Embeddings
Encode morphology of words

A DNN for Drug Detection



On unbalanced data, require pre-filtering

112 Timelines

Selecting tweets potentially mentioning drugs

Lexicon Filter
RxNorm

Select tweets mentioning drug names in the lexicon

T1: I love Lyrica's new song!!!
T2: Lyrica experience? I was on Gabapentin.

Variant Filter
RxNorm

Select tweets mentioning generated variants of drug names

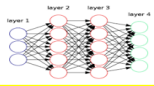
T3: My child likes her tylonol...
T4: im Allergic to all Tynenol, Asprin, n ibuprofen.
T5: Aderal always motivates me

Pattern Filter
Regular Expressions

Select tweets containing phrases manually selected

T6: my OB doctor prescribed me promethazine
T7: I'm on Gabapentin which doesn't help much.
T8: I took a nap and I'm still sleepy

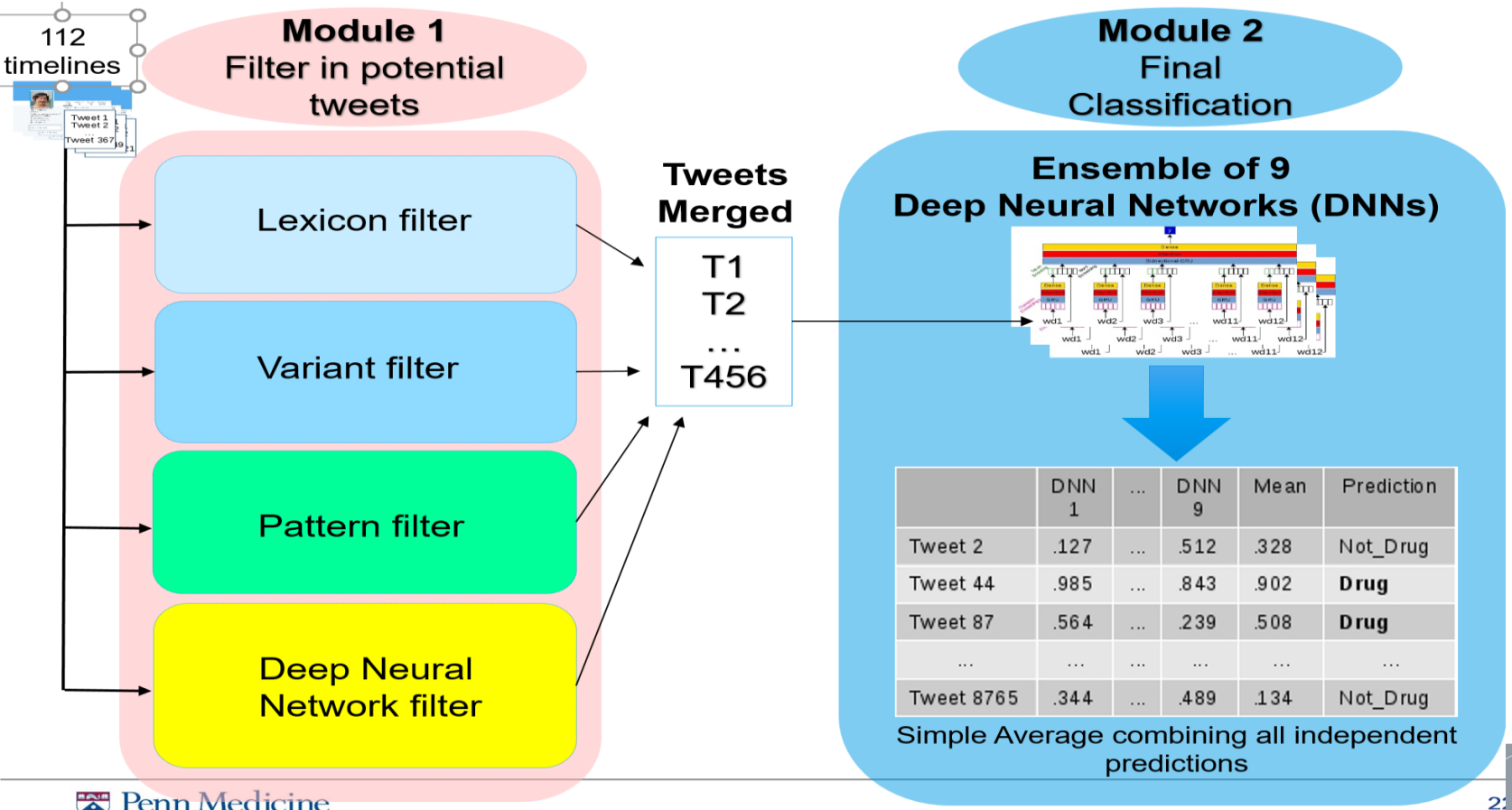
Deep Neural Network Filter



Select tweets recognized by a DNN
(DNN trained on tweets mentioning unambiguous drugs)

T9: I took 600 mg of lyrica am I going to be ok?
T10: snickers bars are my drug of choice
T11: Nasal allergies flare up during monsoon

Kusuri overview



Performance on Natural Corpus

Systems	Precision	Recall	F1
Lexicon+Variant classifier	55.0	**71.4**	62.1
Ensemble Bidirectional-GRUs (trained on unbalanced corpus)	87.5	63.6	73.7
Kusuri	**94.55**	67.5	**78.8**

Conclusion and Future Work

- On balanced corpus, performance is close to human annotators'
- On natural corpus, our system outperforms existing performances in well-established challenges [Limsopatham and Collier, 2016]
- Extreme imbalanced data is the main challenge:
  - Replacing the 4 filters by active learning
  - Exploiting unlabeled data with Generative Adversarial Networks

[Weissenbacher et al., Deep neural networks ensemble for detecting medication mentions in tweets. 2019. JAMIA.

Thank you!

Email me at:
dweissen@pennmedicine.upenn.edu