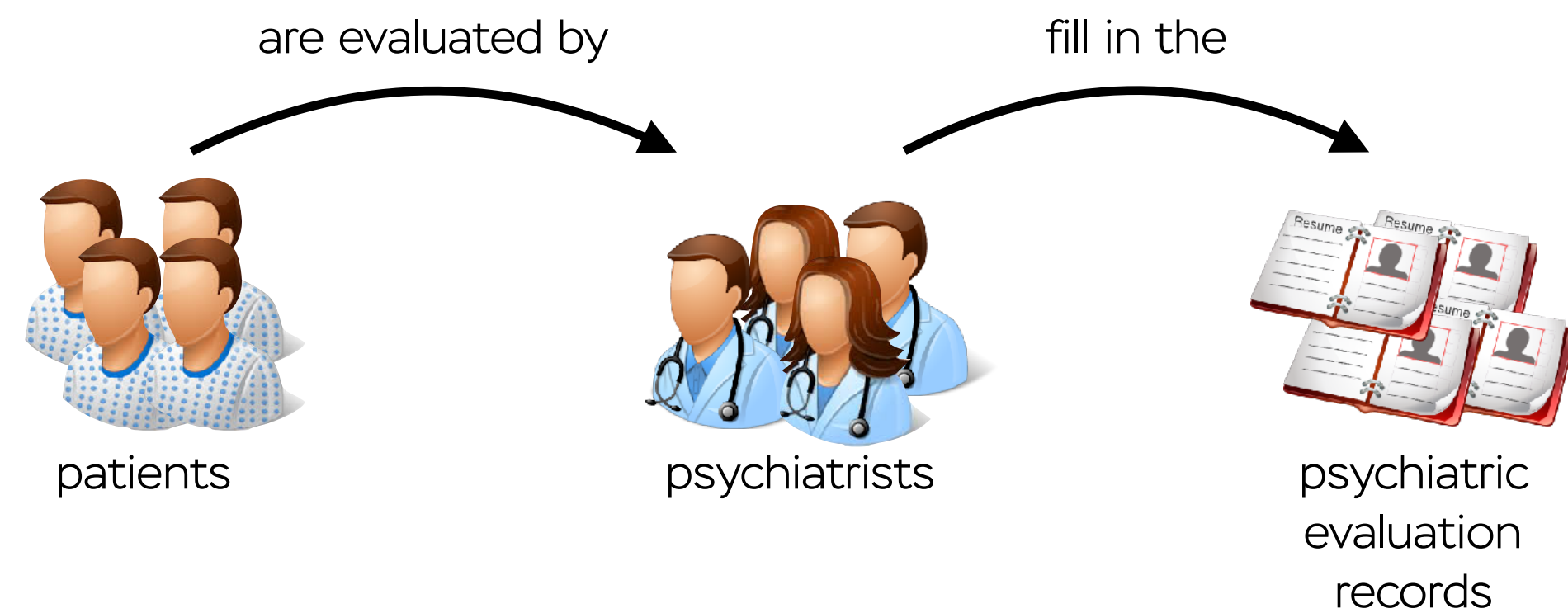# 2016 CEGS N-GRID

Shared-Tasks and Workshop on Challenges in
Natural Language Processing for Clinical Data

Michele Filannino*, Amber Stubbs, Özlem Uzuner

SUNY   MIT

data

are evaluated by — fill in the

patients → psychiatrists → psychiatric evaluation records

- 1000 initial psychiatric evaluation records
  - with surrogate personal health information (PHI)
- from CEGS N-GRID & Partners Healthcare
- mental health records
  - **1st set ever released to the research community!**

data: example

Subject: Patient Initial Visit Note -Identifying Information Date of Service:

09/14/2067CPT Code: 90792: With medical services
Age:

43Sex: Female
Interpreter Used: None needed
Chief Complaint / HPI Chief Complaint (Patients own words)

Valentina is a 43-year old female with a past psychiatric history significant for an underlying depressive disorder, anxiety disorder, binge eating disorder, no history of prior inpatient psychiatric hospitalizations, no history of prior suicide attempts, no known history of self-injurious behavior, no history of EtOH or illicit-substance use and a past medical history significant for DVT (8-years ago while on OCP) who presents to the EDCRP for continued psychopharmacologic care following her transfer from Dr. Yvonne Ellison's practice. Per Valentina, "I just need to continue these medications...they have been so helpful."

History of Present Illness and Precipitating Events

Valentina describes that she first presented for binge-eating disorder treatment approximately one-year prior to current presentation, when she notes that she was bingeing on a daily basis in the context of a number of stressors/transitions, including the birth of her two children within a year of each other, marriage to her college boyfriend at the age of 39, and ensuing conflicts with her mother as a result of the marriage. She notes that at that time, she binged daily, which steadily led to a marked increase in a 30-pound weight gain within 4-months. She states that she would binge in secret from her husband and two daughters and states that following each binge she felt "awful...ashamed." As a result, she states that her mood started to precipitously decline and states that she ultimately realized that she "could no longer fun from this...I had to seek treatment."

In an effort to seek treatment, she began seeing dietician, Frida Gibbons in Hubert, who she states was "helpful, but not enough." She subsequently purchased a diet book and worked through the exercises in the book and states that that, too, was helpful, but not enough. She ultimately contacted one of the authors of the book and arranged to undergo therapywithher via Skype. She states that that was helpful for approximately 6-months, at which time her therapist suggested that she seek local ED-specific care. As such, she underwent an evaluation with Dr. Deon Yarbrough 6/2066 and was subsequently referred to Dr. Tonya Alston for a 20-week course of CBT-E. Valentina states that her treatment with Dr. Alston was "phenomenal...it was exactly what I needed." She concurrently began seeing Dr. Yvonne Ellison for psychopharmacologic treatment of her underlying depressive/anxiety disorder and BED. She has since responded well to her current regimen of Zoloft, Vyvanse, and Topamax.

On further psychiatric review of symptoms, Valentina denies any curer nt feelings of depression and further denies all neurovegetative symptoms of depression. She denies any symptoms consistent with mania or psychosis. She denies any symptoms consistent with generalized anxiety and further denies any associated symptoms of panic. She further denies any history of EtOH or illicit-substance use.

Suicidal Behavior Hx of Suicidal Behavior: No
Hx of Non Suicidal Self Injurious Behavior: No
Violent Behavior Hx of Violent Behavior: No
-Psychiatric History Hx of Inpatient Treatment: No
Hx of Outpatient Treatment: Yes

-currently in psychodynamic therapist with Thelma Benson (Hubert), previously did course of CBT-E and booster sessions with Dr. Alston

-prior dx: ADHD , depression, binge-eating Disorder

Prior medication trials (including efficacy, reasons discontinued):

Ritalin (it made me anxious), and Celexa (no side-effects)Current Therapist Name(s) and Phone #(s):

Thelma Benson Other Agency Involvement: No
Psychiatric Review of Systems DEPRESSION: Has the patient had periods of time lasting two weeks or longer in which, most of the day on most days, they felt little interest or pleasure in doing things, or they had to push themselves to do things: Yes
DEPRESSION: Has the patient had periods of time lasting two weeks or longer in which, most of the day on most days, they felt sad, down, or depressed: Yes
BIPOLAR: Has patient ever had a period of time when he/she felt "up" or "high" without the use of substances: No
BIPOLAR: Has patient ever had periods of being persistently irritable for several days, or had verbal/physical fights that seemed clearly out of character: No
PSYCHOSIS: Has the patient had unusual experiences that are hard to explain: No
PSYCHOSIS: Does the patient often have thoughts that make sense to them, but that other people say are strange: No
GAD: Has the patient had times when they worried excessively about day to day matters for most of the day, more days than not: Yes
PANIC: Has the patient had episodes of sudden intense anxiety with physical sensations such as heart palpitations, trouble breathing, or dizziness that reached a peak very quickly and presented without warning: No
ANXIETY SPECTRUM DISORDERS: Does the patient have persistent fear triggered by specific objects (phobias) or situations (social anxiety) or by thought of having a panic attack: No
OCD: Does the patient struggle with repetitive unwanted thoughts or behaviors for at least one hour per day: No

...

data: example

ok and states that that, too, wa
ately contacted one of the au
ndergo therapywithher via Sl
ful for approximately 6-mont

Subject: Patient Initial Visit Note -Identifying
Service:

09/14/2067CPT Code: 90792: With medical serv
Age:

43Sex: Female
Interpreter Used: None needed
Chief Complaint / HPI Chief Complaint (Patients own words)

History of self injurious behavior, no history of EtOH or illicit substance use and a past medical history significant for DVT (8-years ago while on OCP) who presents to the EDCRP for continued psychopharmacologic care following her transfer from Dr. Yvonne Ellison's practice. Per Valentina, "I just need to continue these medications...they have been so helpful."

 History of Present Illness and Precipitating Events

Valentina describes that she first presented for binge-eating disorder treatment approximately one-year prior to current presentation, when she notes that she was bingeing on a daily basis in the context of a number of stressors/transitions, including the birth of her two children within a year of each other, marriage to her college boyfriend at the age of 39, and ensuing conflicts with her mother as a result of the marriage. She notes that at that time, she binged daily, which steadily led to a marked increase in a 30-pound weight gain within 4-months. She states that she would binge in secret from her husband and two daughters and states that following each binge she felt "awful...ashamed." As a result, she states that her mood started to precipitously decline and states that she ultimately realized that she "could no longer fun from this...I had to seek treatment."

History of Present Illness and Precipitating Events

ent, she began seeing dietician, Frida
o she states was "helpful, but not enough."
hased a diet book and worked through the
nd states that that, too, was helpful, but not
enough. She ultimately contacted one of the authors of the book and arranged to undergo therapywithher via Skype. She states that that was helpful for approximately 6-months, at which time her therapist suggested that she seek local ED-specific care. As such, she underwent an evaluation with Dr. Deon Yarbrough ly referred to Dr. Tonya Alston for a 20-ntina states that her treatment with Dr. it was exactly what I needed." She seeing Dr. Yvonne Ellison for psychopharmacologic treatment of her underlying depressive/ anxiety disorder and BED. She has since responded well to her current regimen

On further psyc
curer nt feeli
neurovegetative symptoms of depression. She denies any
symptoms consistent with mania or psychosis. She denies any symptoms consistent with generalized anxiety and further denies any associated symptoms of panic. She further denies any history of EtOH or illicit-substance use.

 Suicidal Behavior Hx of Suicidal Behavior: No
Hx of Non Suicidal Self Injurious Behavior: No
Violent Behavior Hx of Violent Behavior: No
-Psychiatric History Hx of Inpatient Treatme
Hx of Outpatient Treatment: Yes

-currently in psychodynamic therapist
(Hubert), previously did course of CBT-E an
Dr. Alston

-prior dx: ADHD , depression, binge-eating D

Prior medication trials (including efficacy, reasons discontinued):

Ritalin (it made me anxious), and Celexa (no side-effects)Current Therapist Name(s) and Phone #(s):

On further psychiatric review of symptoms, Valentina denies any curer nt feelings of depression and further denies all neurovegetative symptoms of depression.   She denies any

experiences that are hard
to them, but that other people say are strange: No
GAD: Has the patient had times when they worried excessively about day to day matters for most of the day, more days than not: Yes
PANIC: Has the patient had episodes of sudden intense anxiety with physical sensations such as heart palpitations, trouble breathing, or dizziness that reached a peak very quickly and presented

Suicidal Behavior Hx of Suicidal Behavior: No
Hx of Non Suicidal Self Injurious Behavior: No
Violent Behavior Hx of Violent Behavior: No
-Psychiatric History Hx of Inpatient Treatment: No
Hx of Outpatient Treatment: Yes

Michele Filannino

■ Text - ■ Semi-structured information - ■ Typos

4

track 1:
de-identification in **2** flavors

- Track 1.A sight-unseen:
  - running existing de-id systems on the data "*as they are*"
  - how well do systems generalize to new data?

- Track 1.B regular:
  - design new systems (2 months, 60/40 training/test split)
  - advancing the state of the art in medical rec. de-identification

- Ranking/error measure:
  - Precision, Recall, F1-score*

Subject: Patient Initial Visit Note -Identifying Information Date of Service:

09/14/2067 CPT Code: 90792: With medical services
Age:

43 Sex: Female
Interpreter Used: None needed
Chief Complaint / HPI Chief Complaint (Patients own words)

Valentina is a 43-year old female with a past psychiatric history significant for an underlying depressive disorder, anxiety disorder, binge eating disorder, no history of prior inpatient psychiatric hospitalizations, no history of prior suicide attempts, no known history of self-injurious behavior, no history of EtOH or illicit-substance use and a past medical history significant for DVT (8-years ago while on OCP) who presents to the EDCRP for continued psychopharmacologic care following her transfer from Dr. Yvonne Ellison's practice. Per Valentina, "I just need to continue these medications...they have been so helpful."

History of Present Illness and Precipitating Events

Valentina describes that she first presented for binge-eating disorder treatment approximately one-year prior to current presentation, when she notes that she was bingeing on a daily basis in the context of a number of stressors/transitions, including the birth of her two children within a year of each other, marriage to her college boyfriend at the age of 39, and ensuing conflicts with her mother as a result of the marriage. She notes that at that time, she binged daily, which steadily led to a marked increase in a 30-pound weight gain within 4-months. She states that she would binge in secret from her husband and two daughters and states that following each binge she felt "awful...ashamed." As a result, she states her mood started to precipitously decline and states that she ultimately realized that she "could no longer fun from this...I had to seek treatment."

In an effort to seek treatment, she began seeing dietician, Frida Gibbons in Hubert, who she states was "helpful, but not enough." She subsequently purchased a diet book and worked through the exercises in the book and states that that, too, was helpful, but not enough. She ultimately contacted one of the authors of the book and arranged to undergo therapy with her via Skype. She states that that was helpful for approximately 6-months, at which time her therapist suggested that she seek local ED-specific care. As such, she underwent an evaluation with Dr. Deon Yarbrough 6/2066 and was subsequently referred to Dr. Tonya Alston for a 20-week course of CBT-E. Valentina states that her treatment with Dr. Alston was "phenomenal...it was exactly what I needed." She concurrently began seeing Dr. Yvonne Ellison for psychopharmacologic treatment of her underlying depressive/anxiety disorder and BED. She has since responded well to her current regimen of Zoloft, Vyvanse, and Topamax.

track 2: symptom severity classification

- RDoC: framework for studying mental disorders

  - integrates many levels of information (from genomics to self-report) to understand the basic dimensions of human behavior (from normal to abnormal)

  - 5 domains: **POSITIVE VALENCE***, NEGATIVE VALENCE, COGNITIVE, SOCIAL PROCESSES, AROUSAL AND REGULATORY SYSTEMS

  - how good systems are at predicting patients' symptom severity, based on initial psychiatric evaluation records?

* Systems primarily responsible for responses to positive motivational situations or contexts, such as reward seeking, consummatory behavior, and reward/habit learning (alcohol, drink, abuse, repetitive, stereotypy, drug, gamble, count, craving, compulsive)

track 3: novel data use

- open track

- **aim:** addressing new research questions on a novel data set.

participation

- 15 countries
- 50 teams
- 65 institutions
- 154 researchers
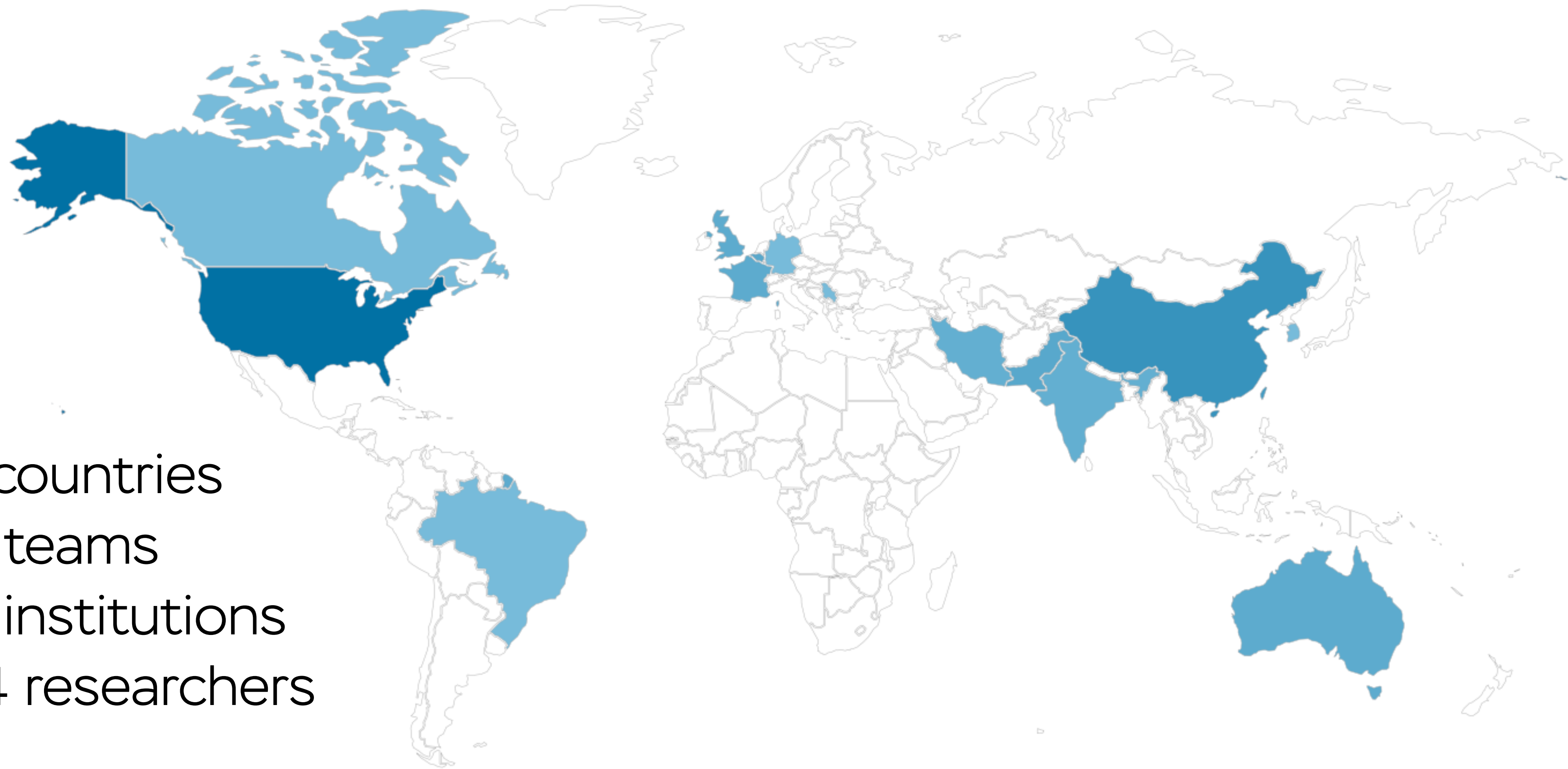
timeline

Registration:	begins in May, 2016
Data Release for Sight-Unseen Track:	6th June 2016
System Outputs Due for Sight-Unseen Track:	10th June 2016
Training Data Release:	11th June 2016
Test Data Release:	10th August 2016
System Outputs Due:	12th August 2016
Abstract Submission:	1st September 2016
Workshop:	11th November 2016 (@AMIA, Chicago)
Journal Submissions:	TBD

results

	Min	Average	Max	Standard deviation	Number of participants	Best method
Track 1.A	47.74%	56.50%	79.56%	0.20557	17	Conditional Random Fields
Track 1.B	1.89%	77.87%	91.43%	0.18326	29	Bidirectional LSTM
Track 2	52.46%	77.15%	86.30%	0.05608	65	Ensemble of 6 ML techniques

Organizing committee:
Ozlem Uzuner, co-chair, SUNY at Albany
Amber Stubbs, co-chair, Simmons College
Michele Filannino, co-chair, SUNY at Albany

Tianxi Cai, Harvard School of Public Health
Susanne Churchill, Harvard Medical School
Isaac Kohane, Harvard Medical School
Thomas H. McCoy, MGH, Harvard
Roy H. Perlis, MGH, Harvard
Peter Szolovits, MIT
Uma Vaidyanathan, NIMH
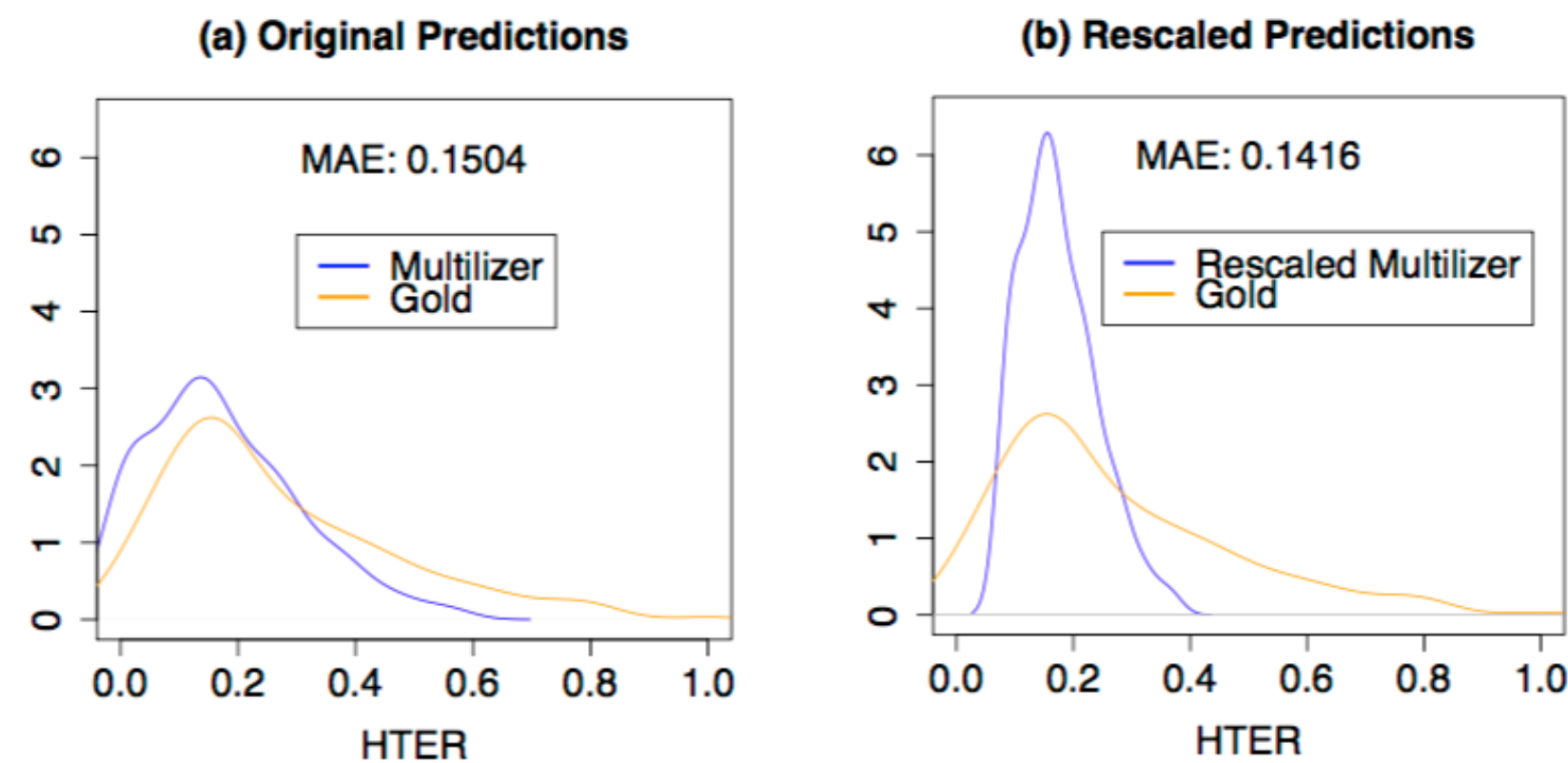Philip Wang, American Psychiatric Association

thanks!

19th October 2016

MAEμ

test docs   prediction

gold

$$MAE^{\mu}(h, Te) = \frac{1}{|Te|} \sum_{\mathbf{x}_i \in Te} |h(\mathbf{x}_i) - y_i|$$

- predictions can be adjusted by guessing the central tendency

(a) Original Predictions

MAE: 0.1504

— Multilizer
— Gold

HTER

(b) Rescaled Predictions

MAE: 0.1416

— Rescaled Multilizer
— Gold

HTER

1. Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Evaluation Measures for Ordinal Regression. In 2009 Ninth International Conference on Intelligent Systems Design and Applications (pp. 283–287). IEEE. http://doi.org/10.1109/ISDA.2009.230
2. Graham, Y., 2015. Improving evaluation of machine translation quality estimation. In 53rd Annual Meeting of the Association for Computational Linguistics and Seventh International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (pp. 1804-1813).
3. Moreau, E. and Vogel, C., 2014, August. Limitations of MT Quality Estimation Supervised Systems: The Tails Prediction Problem. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics (pp. 2205-2216). Dublin City University and Association for Computational Linguistics.

4. Celli, F., Pianesi, F., Stillwell, D. and Kosinski, M., 2013, June. Workshop on computational personality recognition (shared task). In Proceedings of the Workshop on Computational Personality Recognition.
5. Machácek, M. and Bojar, O., 2014, June. Results of the WMT14 metrics shared task. In Proceedings of the Ninth Workshop on Statistical Machine Translation (pp. 293-301).
6. P. Nakov and A. Ritter and S. Rosenthal and V. Stoyanov and F. Sebastiani (2015). SemEval-2016 Task 4: Sentiment Analysis in Twitter. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016).

Michele Filannino

MAE^M

$$MAE^M(h, Te) = \frac{1}{|\mathcal{C}|} \sum_{j=1}^{|\mathcal{C}|} \frac{1}{|Te_j|} \sum_{\mathbf{x}_i \in Te_j} |h(\mathbf{x}_i) - y_i|$$

test docs with class j

class

prediction
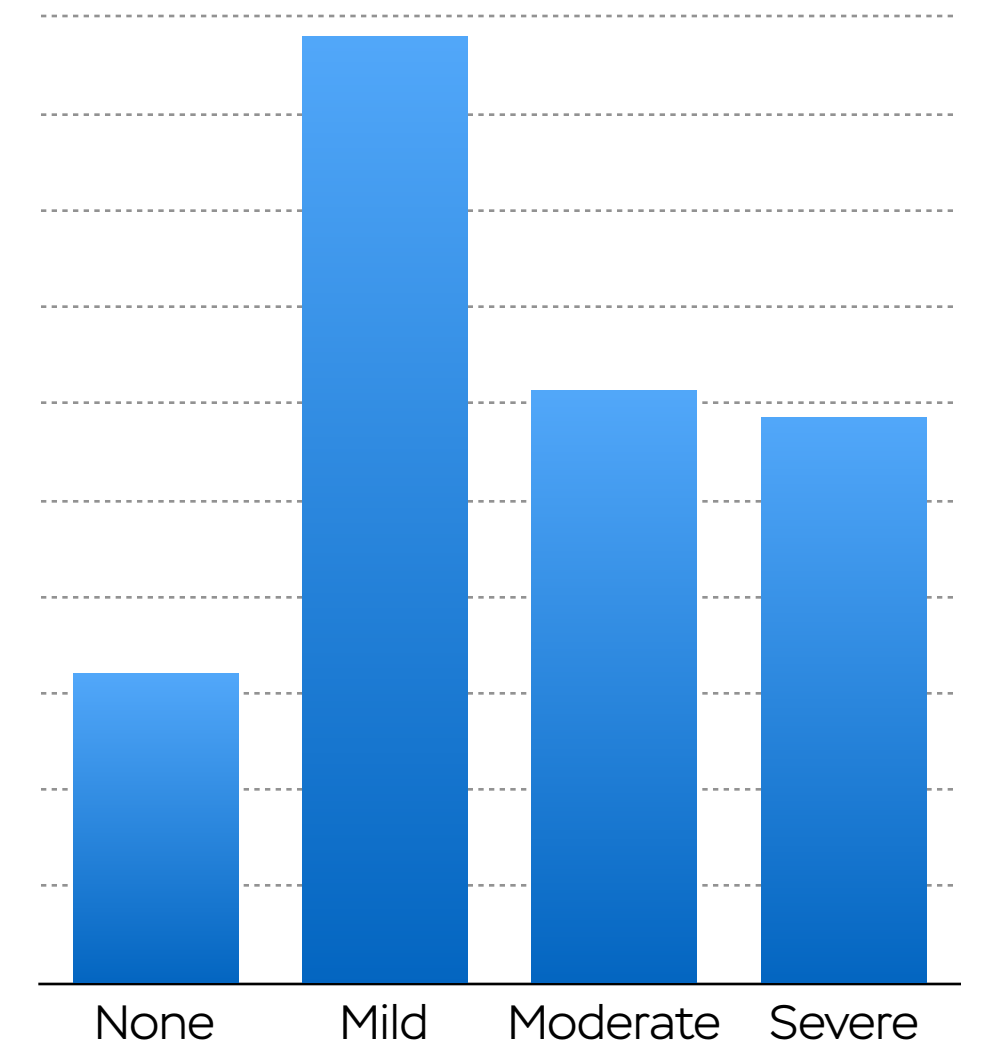
gold

- copes with imbalanced data
- the trivial class counts as any other class, instead of proportionally to its frequency

task 2: ranking examples

	none			mild			moderate					severe		
gold standard	0	0	0	1	1	1	2	2	2	2	2	3		
system A	0	0	0	1	1	1	2	2	2	2	2	3	*no errors*	1st
system B	0	0	0	1	1	1	2	2	0	2	2	3	*1 error*	3rd
system C	0	0	0	1	1	1	2	2	1	2	2	3	*1 error*	2nd
system D	0	1	2	2	2	2	2	1	2	1	2	3	*7 errors*	4th
system E	0	1	2	2	2	2	2	1	2	2	2	2	*7 errors*	6th
system F	2	2	2	2	2	2	2	2	2	2	2	3	*6 errors*	5th
system G	2	2	2	0	0	0	0	0	0	0	0	0	*totally wrong*	7th

task 2: ranking examples

	P	R	F1	Acc.	Kappa	Scott's Pi	Median Abs. E.	Mean Absolute Error		Mean² E.	R²	Pearson's
								micro-avg.	macro-avg.			
	1	1	1	1	1	1	1	1	1	1	1	1
	3	3	3	3	3	3	3	2	2	2	2	2
	2	2	2	2	2	2	2	3	3	3	3	3
	4	5	5	5	5	4	5	4	4	4	4	5
	6	4	4	4	4	5	4	6	5	6	6	4
	5	6	6	6	6	6	6	5	6	5	5	6
	7	7	7	7	7	7	7	7	7	7	7	7
	X	X	X	X	X	X	X	X	V	X	X	X
								biased		*biased*	*continuous*	*continuous*