

2018 n2c2 Shared Tasks

Track 2: Adverse Drug Events and Medication Extraction in EHRs

Amber Stubbs, Kevin Buchan, Michele Filannino,
Ozlem Uzuner, PhD, FACMI

Associate Professor, George Mason University
Research Affiliate, MIT

Visiting Associate Professor, Harvard Medical School

November 2, 2018

Track 2: Goals

- Answer the question: “Can NLP systems automatically discover drug to adverse event (ADE) relations in clinical narratives?
- Three tasks:
 - Concepts: Identifying drug names, dosages, durations and other entities.
 - Relations: Identifying relations of drugs with ADEs and other entities given gold standard entities.
 - End-to-end: Identifying relations of drugs with ADEs and other entities on system predicted entities.

Corpus

- 505 files from MIMIC
 - Pseudo query:
 - **SELECT HADM_ID (i.e., hospital visits) WHERE "adverse effects in therapeutic use" was in the ICD10 description**
 - Then manual review
- Annotated for the following information:
 - Entities: Drug, strength, dosage, duration, frequency, form, route, reason, ADE
 - Relations: strength-drug, dosage-drug, duration-drug, frequency-drug, form-drug, route-drug, reason-drug, ADE-drug

Gold Standard Creation

- Annotation
- Adjudication

Gold Standard Creation

- Annotation
 - Pre-annotation with c-TAKES
 - Double annotation
- Arbitration
- Machine based consistency checks with manual review and adjustment
- Bug fixes

Agreement Metrics

- P, R, F1, calculated at micro- and macro-levels

Inter-annotator Agreement

***** TRACK 2 *****							***** TRACK 2 *****						
	strict			lenient				strict			lenient		
	Prec.	Rec.	F(b=1)	Prec.	Rec.	F(b=1)		Prec.	Rec.	F(b=1)	Prec.	Rec.	F(b=1)
Drug	0.9351	0.8753	0.9042	0.9806	0.9179	0.9482		0.8255	0.9274	0.8735	0.8498	0.9546	0.8991
Strength	0.9710	0.9636	0.9672	0.9869	0.9794	0.9831		0.8974	0.9358	0.9162	0.9436	0.9840	0.9634
Duration	0.6298	0.5759	0.6016	0.8553	0.7821	0.8171		0.4722	0.4722	0.4722	0.8889	0.8889	0.8889
Route	0.9718	0.9299	0.9504	0.9775	0.9353	0.9559		0.8545	0.8545	0.8545	0.9212	0.9212	0.9212
Form	0.9409	0.9289	0.9349	0.9607	0.9485	0.9545		0.6370	0.8515	0.7288	0.6519	0.8713	0.7458
Ade	0.5891	0.5522	0.5701	0.6658	0.6241	0.6443		0.3158	0.2500	0.2791	0.4737	0.3750	0.4186
Dosage	0.6808	0.6703	0.6755	0.9366	0.9221	0.9293		0.7739	0.7876	0.7807	0.8261	0.8407	0.8333
Reason	0.6437	0.6366	0.6401	0.7392	0.7309	0.7350		0.3582	0.6000	0.4486	0.5075	0.8500	0.6355
Frequency	0.7730	0.7568	0.7648	0.9669	0.9465	0.9566		0.8424	0.8516	0.8470	0.9620	0.9725	0.9672
Overall (micro)	0.8727	0.8426	0.8574	0.9486	0.9159	0.9319	Overall (micro)	0.7665	0.8534	0.8076	0.8308	0.9249	0.8754
Overall (macro)	0.8657	0.8396	0.8500	0.9364	0.9086	0.9196	Overall (macro)	0.7983	0.8452	0.8177	0.8592	0.9108	0.8808
***** RELATIONS *****													
Strength -> Drug	0.9273	0.9209	0.9241	0.9808	0.9739	0.9774	Strength -> Drug	0.8802	0.8989	0.8895	0.9427	0.9628	0.9526
Dosage -> Drug	0.6360	0.6344	0.6352	0.9144	0.9120	0.9132	Dosage -> Drug	0.7434	0.7568	0.7500	0.8230	0.8378	0.8304
Duration -> Drug	0.5714	0.5324	0.5512	0.7949	0.7406	0.7668	Duration -> Drug	0.3684	0.3784	0.3733	0.7895	0.8108	0.8000
Frequency -> Drug	0.7302	0.7196	0.7249	0.9529	0.9391	0.9459	Frequency -> Drug	0.8033	0.8167	0.8099	0.9454	0.9611	0.9532
Form -> Drug	0.8907	0.8876	0.8892	0.9436	0.9404	0.9420	Form -> Drug	0.6090	0.7941	0.6894	0.6617	0.8627	0.7489
Route -> Drug	0.9261	0.8906	0.9080	0.9685	0.9313	0.9495	Route -> Drug	0.8272	0.8171	0.8221	0.9012	0.8902	0.8957
Reason -> Drug	0.5274	0.5668	0.5464	0.6315	0.6787	0.6543	Reason -> Drug	0.3053	0.5686	0.3973	0.4421	0.8235	0.5753
ADE -> Drug	0.5244	0.5029	0.5134	0.6077	0.5828	0.5950	ADE -> Drug	0.2857	0.1500	0.1967	0.3810	0.2000	0.2623
Overall (micro)	0.7786	0.7759	0.7772	0.8961	0.8931	0.8946	Overall (micro)	0.6715	0.7500	0.7086	0.7781	0.8690	0.8211
Overall (macro)	0.7554	0.7612	0.7539	0.8685	0.8736	0.8655	Overall (macro)	0.6818	0.7240	0.6972	0.8045	0.8510	0.8214

172 files found

9 files found

Annotator Disagreements

- FORM: the most common error here is missing some instances of the word "Tablet".
 - Impacts the relations between the missed FORM and the DRUG.
- ADE: superfluous annotations
 - E.g., agitation, allergy, difficulty swallowing secretions, emesis, mucositis, severe nausea
 - Impacts the relations between ADEs and DRUGs.
- DOSAGE: some are simply missed
 - E.g., forty five, one, 6 cycles, 50,000 unit, 2, 4
- REASON: superfluous annotations
 - E.g., low grade fever, lymphedema, viral, zoster px, febrile, cellulitis, bacterial.
 - Many are adjectives or adverbs.

Pre-annotations played a role.

Gold Standard Statistics

Concepts (n = 83840)		Relations (n = 61475)	
Drug	26,803		
Strength	10,922	Strength-Drug	10,950
Form	11,006	Form-Drug	11,048
Dosage	6,900	Dosage-Drug	6,939
Frequency	10,293	Frequency-Drug	10,352
Route	8,987	Route-Drug	9,086
Duration	966	Duration-Drug	1,069
Reason	6,384	Reason-Drug	8,611
ADE	1579	ADE-Drug	1,841

Track 2: Shared Task

- Training: 303 files with annotations
 - 1.5 month development time
- Testing: 202 files, released in two forms:
 - No annotations (for Concepts and End-to-end tasks) – 3 days for system runs
 - Gold standard concept annotations (for relations task) – 2 days for system runs
 - Submit up to 3 runs for each of concepts, relations, and end-to-end

Participants

- 28 teams participated
 - Over 100 participants
 - 11 countries represented
- 158 system outputs submitted
 - Concepts: 60
 - Relations: 48
 - End-to-end: 50

Evaluation Metrics

- Evaluation using standard micro-averaged P, R, and F1
- Different levels of strictness:
 - Strict matching: first and last offset must match exactly
 - Lenient: Matches if tags overlap – this is the *primary evaluation metric*

Aggregate Statistics

	Concepts	Relations	End-to-end
Maximum	0.942	0.963	0.8905
Minimum	0.011	0.157	0.045
Median	0.903	0.892	0.800
Average	0.846	0.811	0.740
Standard deviation	0.175	0.190	0.185

Concepts – Top 10 Teams (Best Run only)

Rank	Team	Lenient Micro F1
1	Alibaba Inc.	0.9418
2	UTHealth / Dalian	0.9345
3	University of Florida	0.9287
4	The University of Manchester	0.9267
5	Medical University of South Carolina	0.9266
6	Univ. of Manchester / Toyota Technology Institute / AIST	0.9255
7	National Taiwan University Hospital/National Taitung University	0.9191
8	VA Salt Lake City / University of Utah	0.9161
9	IBM Research	0.9144
10	University of Michigan / Ramakrishna Mission	0.914

Concepts – Top 10 Teams (Best Run only)

* = Presentation
= Poster

Rank	Team	Lenient Micro F1
1	* Alibaba Inc.	0.9418
2	* UTHealth / Dalian	0.9345
3	University of Florida	0.9287
4	# The University of Manchester	0.9267
5	* Medical University of South Carolina	0.9266
6	* Univ of Manchester / Toyota Technology Institute / AIST	0.9255
7	National Taiwan University Hospital/National Taitung University	0.9191
8	* VA Salt Lake City / University of Utah	0.9161
9	* IBM Research	0.9144
10	University of Michigan / Ramakrishna Mission	0.914

Relations – Top 10 Teams (Best Run only)

Rank	Team	Lenient Micro F1
1	UTHealth/Dalian	0.963
2	VA Salt Lake City / University of Utah	0.953
3	Univ. of Manchester / Toyota Technology Institute / AIST	0.9472
4	University of Florida	0.9459
5	Med Data Quest	0.9442
6	IBM Research	0.9385
7	Medical University of South Carolina	0.9362
8	The University of Manchester	0.9175
9	Boston Children's Hospital / Harvard Medical School / Loyola University	0.9067
10	Cincinnati Children's Hospital Medical Center	0.9023

Relations – Top 10 Teams (Best Run only)

Rank	Team	Lenient Micro F1
1	* UTHealth/Dalian	0.963
2	* VA Salt Lake City / University of Utah	0.953
3	* Univ. of Manchester / Toyota Technology Institute / AIST	0.9472
4	University of Florida	0.9459
5	# Med Data Quest	0.9442
6	* IBM Research	0.9385
7	* Medical University of South Carolina	0.9362
8	# The University of Manchester	0.9175
9	# Boston Children's Hospital / Harvard Medical School / Loyola University	0.9067
10	# Cincinnati Children's Hospital Medical Center	0.9023

End-to-End – Top 10 Teams (Best Run only)

Rank	Team	Lenient Micro F1
1	UTHealth / Dalian	0.8905
2	University of Florida	0.8778
3	Univ. of Manchester / Toyota Technology Institute / AIST	0.8765
4	Medical University of South Carolina	0.8688
5	VA Salt Lake City / University of Utah	0.8655
6	IBM Research	0.8598
7	The University of Manchester	0.8468
8	Cincinnati Children's Hospital Medical Center	0.8391
9	Boston Children's Hospital / Harvard Medical School / Loyola University	0.8249
10	Roam Analytics	0.8037

End-to-End – Top 10 Teams (Best Run only)

Rank	Team	Lenient Micro F1
1	* UTHealth / Dalian	0.8905
2	University of Florida	0.8778
3	* Univ. of Manchester / Toyota Technology Institute / AIST	0.8765
4	* Medical University of South Carolina	0.8688
5	* VA Salt Lake City / University of Utah	0.8655
6	* IBM Research	0.8598
7	# The University of Manchester	0.8468
8	# Cincinnati Children's Hospital Medical Center	0.8391
9	# Boston Children's Hospital / Harvard Medical School / Loyola University	0.8249
10	# Roam Analytics	0.8037

Time to Find out more!

Organizing Committee

- Ozlem Uzuner, co-chair, George Mason University, Harvard Medical School
- Michele Filannino, co-chair, George Mason University
- Amber Stubbs, co-chair, Simmons University
- Kevin Buchan, SUNY at Albany
- Kahyun Lee, George Mason University
- Susanne Churchill, Harvard Medical School
- Isaac Kohane, Harvard Medical School



HARVARD
MEDICAL SCHOOL

DEPARTMENT OF
Biomedical Informatics



Volgenau School
of Engineering

Thank you!

- National Library of Medicine, National Institutes of Health, NIH NLM 5R13LM011411, PI: Ozlem Uzuner.
- Department of Biomedical Informatics, Harvard Medical School.
- Annotators!
- Participants!!



HARVARD
MEDICAL SCHOOL

DEPARTMENT OF
Biomedical Informatics



Volgenau School
of Engineering

2018 n2c2 Shared Tasks

Track 2: Adverse Drug Events and Medication Extraction in EHRs

Amber Stubbs, Kevin Buchan, Michele Filannino,
Ozlem Uzuner, PhD, FACMI

Associate Professor, George Mason University
Research Affiliate, MIT
Visiting Associate Professor, Harvard Medical School

November 2, 2018

Concepts – Top 10 Teams (Best Run only)

* = Presentation
= Poster

Ran k	Team	Lenient Micro F1	Strict Micro F1
1	* Alibaba Inc.	0.9418	0.8956
2	* UTHealth / Dalian	0.9345	0.8903
3	University of Florida	0.9287	0.881
4	# The University of Manchester	0.9267	0.8684
5	* Medical University of South Carolina	0.9266	0.8858
6	* Univ of Manchester / Toyota Technology Institute / AIST	0.9255	0.8805
7	National Taiwan University Hospital/National Taitung University	0.9191	0.8704
8	* VA Salt Lake City / University of Utah	0.9161	0.8556
9	* IBM Research	0.9144	0.8676

Concepts – Top 10 Teams (Best Run only)

Rank	Team	Lenient Micro F1	Strict Micro F1
1	Alibaba Inc.	0.9418	0.8956
2	UTHealth / Dalian	0.9345	0.8903
3	University of Florida	0.9287	0.881
4	The University of Manchester	0.9267	0.8684
5	Medical University of South Carolina	0.9266	0.8858
6	Univ. of Manchester / Toyota Technology Institute / AIST	0.9255	0.8805
7	National Taiwan University Hosp./National Taitung Uni.	0.9191	0.8704
8	VA Salt Lake City / University of Utah	0.9161	0.8556
9	IBM Research	0.9144	0.8676
10	University of Michigan / Ramakrishna Mission	0.914	0.8568

Relations – Top 10 Teams (Best Run only)

Ran k	Team	Lenient Micro F1	Strict Micro F1
1	* UTHealth/Dalian	0.963	0.963
2	* VA Salt Lake City / University of Utah	0.953	0.953
3	* Univ. of Manchester / Toyota Technology Institute / AIST	0.947	0.947
4	University of Florida	0.946	0.946
5	# Med Data Quest	0.944	0.944
6	* IBM Research	0.939	0.938
7	* Medical University of South Carolina	0.936	0.936
8	# The University of Manchester	0.918	0.915
9	# Boston Children's Hospital / Harvard Medical School / Loyola University	0.907	0.907

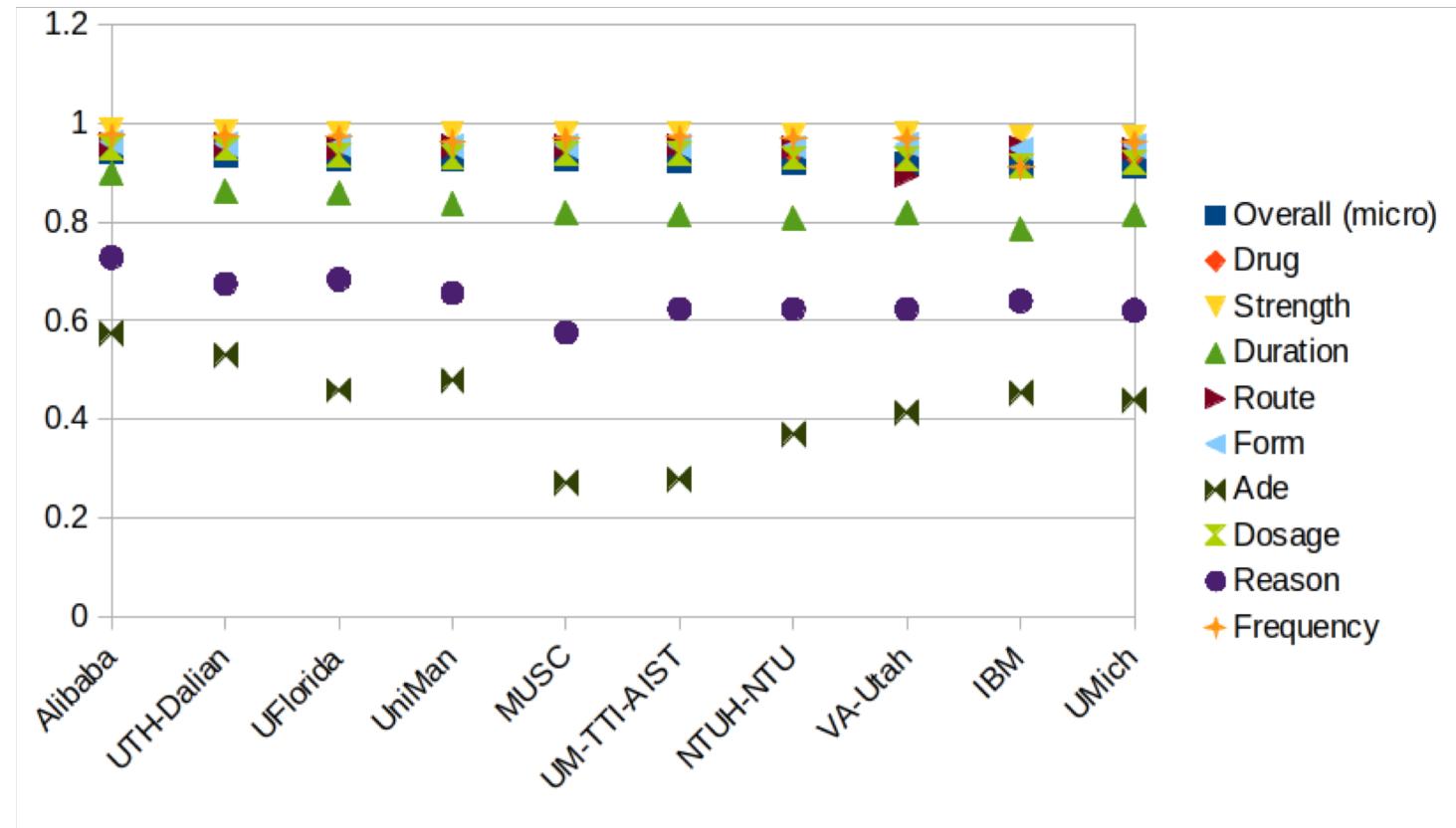
End-to-End – Top 10 Teams (Best Run only)

Ran k	Team	Lenient Micro F1	Strict Micro F1
1	* UTHealth / Dalian	0.8905	0.8197
2	University of Florida	0.8778	0.8045
3	* Univ. of Manchester / Toyota Technology Institute / AIST	0.8765	0.8043
4	* Medical University of South Carolina	0.8688	0.8051
5	* VA Salt Lake City / University of Utah	0.8655	0.7738
6	* IBM Research	0.8598	0.7852
7	# The University of Manchester	0.8468	0.7646
8	# Cincinnati Children's Hospital Medical Center	0.8391	0.7656
9	# Boston Children's Hosp / Harvard Med School / Loyola Uni	0.8249	0.7484
10	# Roam Analytics	0.8037	0.6735

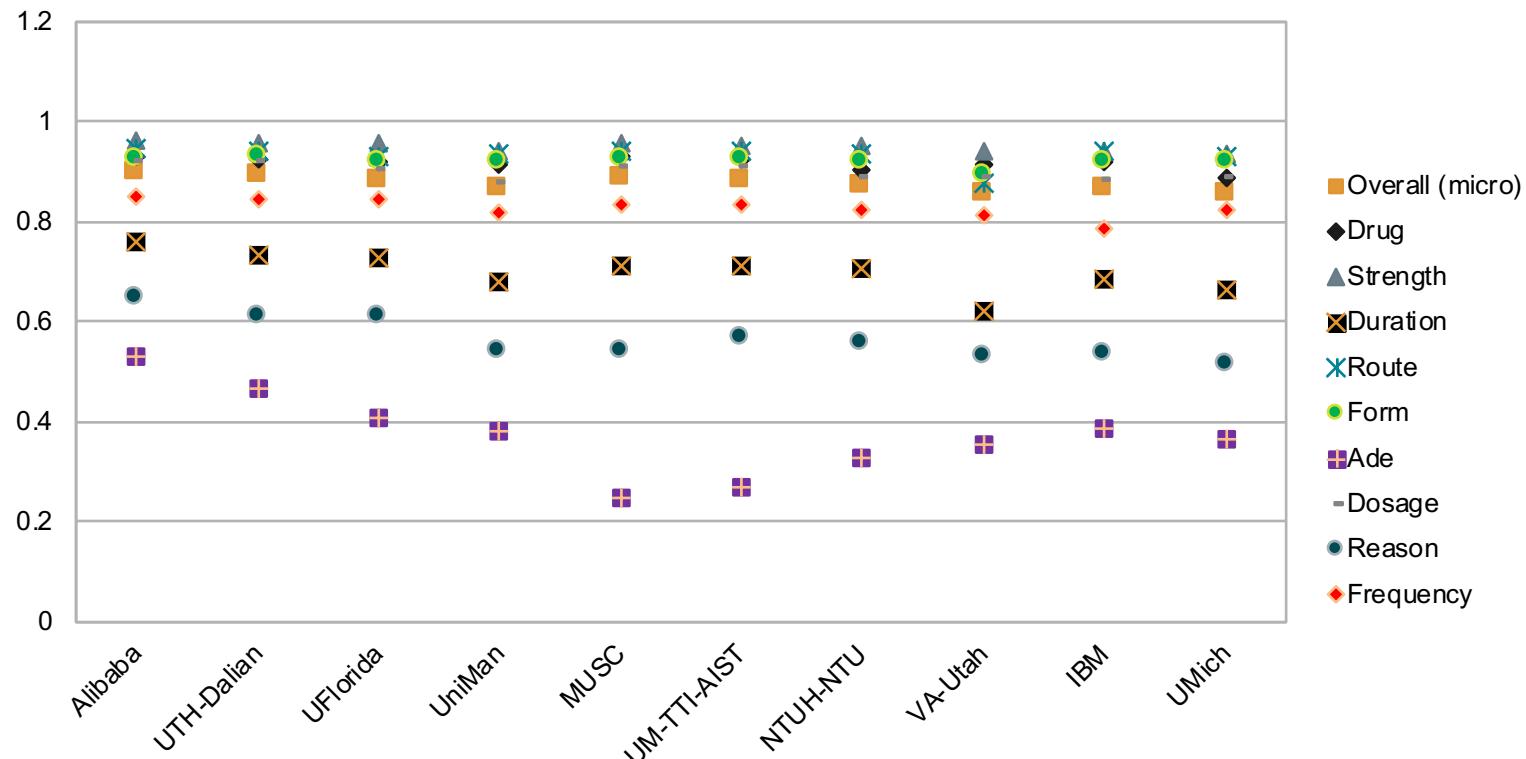
End-to-End – Top 10 Teams (Best Run only)

Ran k	Team	Lenient Micro F1	Strict Micro F1
1	UTHealth / Dalian	0.8905	0.8197
2	University of Florida	0.8778	0.8045
3	Univ. of Manchester / Toyota Technology Institute / AIST	0.8765	0.8043
4	Medical University of South Carolina	0.8688	0.8051
5	VA Salt Lake City / University of Utah	0.8655	0.7738
6	IBM Research	0.8598	0.7852
7	The University of Manchester	0.8468	0.7646
8	Cincinnati Children's Hospital Medical Center	0.8391	0.7656
9	Boston Children's H. / Harvard Med. School / Loyola Univ.	0.8249	0.7484
10	Roam Analytics	0.8037	0.6735

Results by Category – Concepts, Lenient F1s only



Results by Category – Concepts, Strict F1s only



Gold Standard Statistics

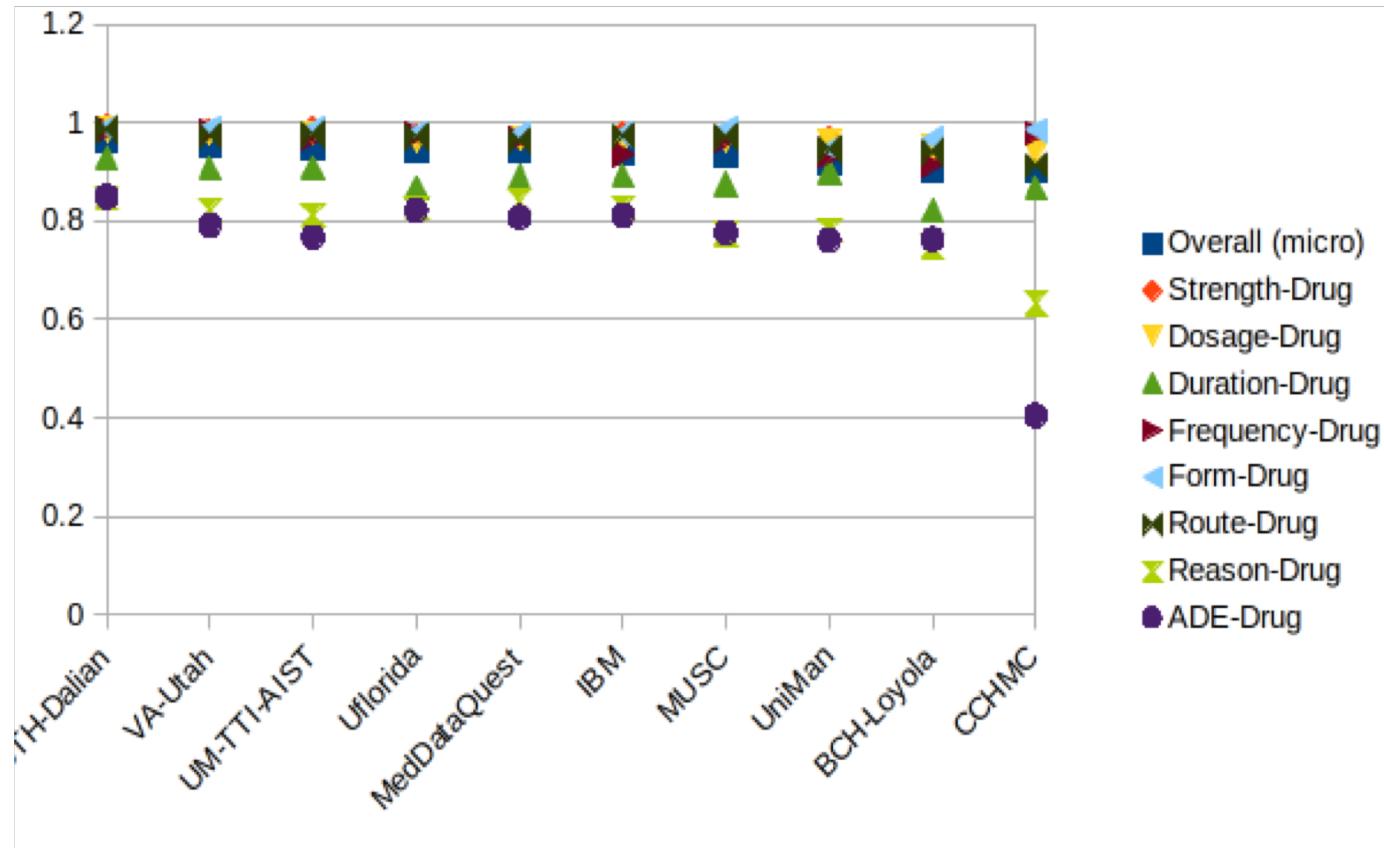
Concepts (n = 83840)	
Drug	26,803
Strength	10,922
Form	11,006
Dosage	6,900
Frequency	10,293
Route	8,987
Duration	966
Reason	6,384
ADE	1579

Annotator Disagreements

- FORM: the most common error here is missing some instances of the word "Tablet".
 - Impacts the relations between the missed FORM and the DRUG.
- ADE: *superfluous annotations*
 - E.g., *agitation, allergy, difficulty swallowing secretions, emesis, mucositis, severe nausea*
 - *Impacts the relations between ADEs and DRUGs.*
- DOSAGE: some are simply missed
 - E.g., forty five, one, 6 cycles, 50,000 unit, 2, 4
- REASON: *superfluous annotations*
 - E.g., *low grade fever, lymphedema, viral, zoster px, febrile, cellulitis, bacterial.*
 - *Many are adjectives or adverbs.*

Pre-annotations played a role.

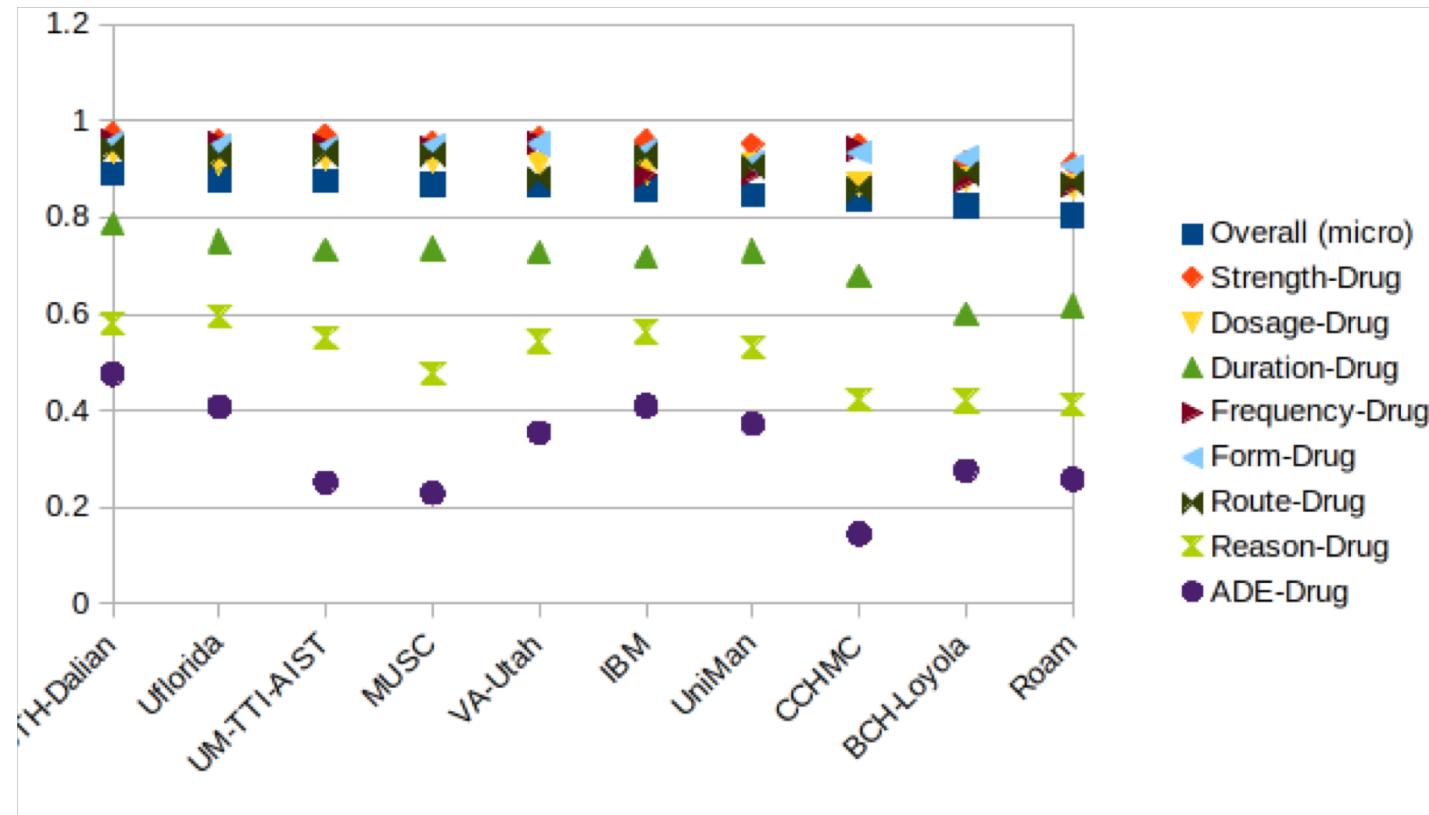
Results by Category – Relations, Lenient F1s only



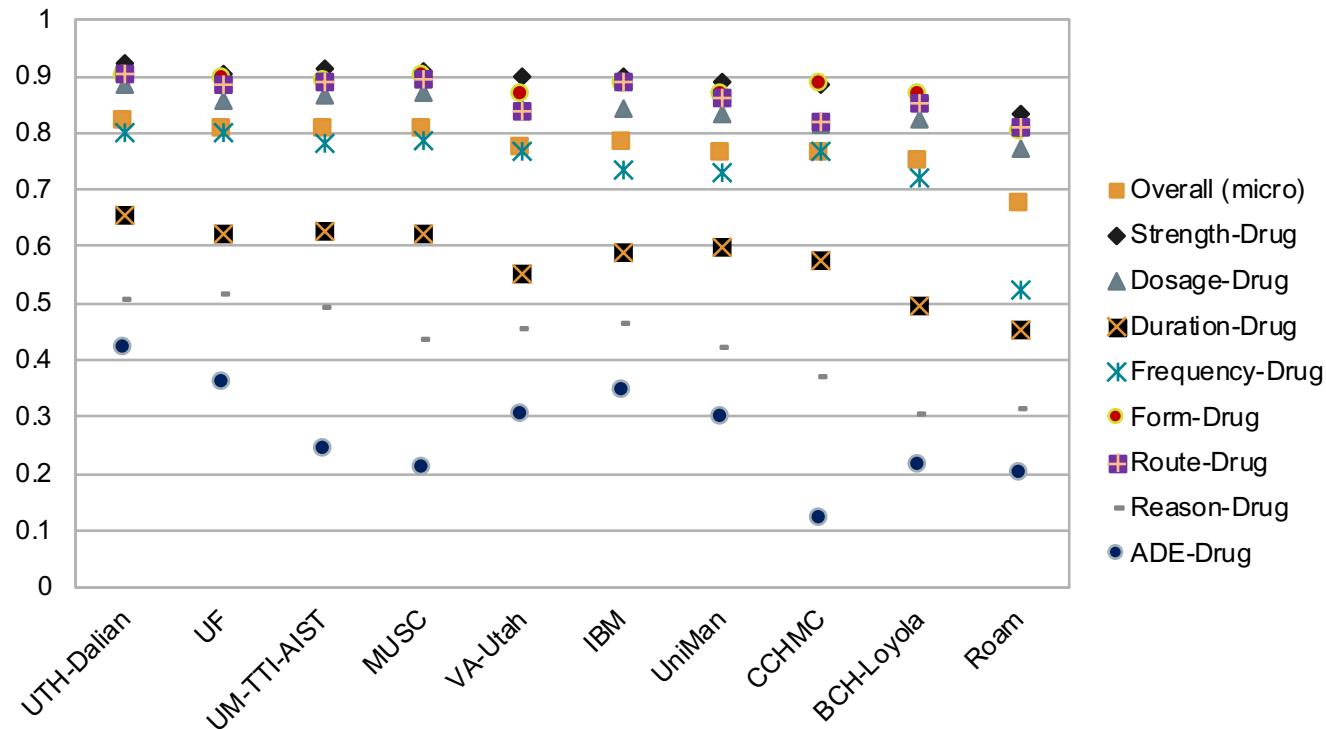
Gold Standard Statistics

Relations (n = 61475)	
Strength-Drug	10,950
Form-Drug	11,048
Dosage-Drug	6,939
Frequency-Drug	10,352
Route-Drug	9,086
Duration-Drug	1,069
Reason-Drug	8,611
ADE-Drug	1,841

Results by Category - End-to-End, Lenient F1s only



Results by Category - End-to-End, Strict F1s only



Medication Challenge 2009 vs 2018 (end-to-end)

Rank	Team	Horizontal Micro F1
1	Patrick_Usyd.3	0.849
2	Xu_Vanderbilt.1	0.8231
3	Aronson_NLM.2	0.8131
4	Yang_Open.3	0.8116
5	Tikk_BME-Humboldt.3	0.8073
6	Manchester_Spasic.2	0.8003
7	HAMON_PARIS.2	0.779
8	Zweigenbaum_Limsi.2	0.7738
9	Yu_UWM.2	0.7651
10	Meystre_UofUtah.1	0.7593

Lenient Micro F1	Strict Micro F1
0.8905	0.8197
0.8778	0.8045
0.8765	0.8043
0.8688	0.8051
0.8655	0.7738
0.8598	0.7852
0.8468	0.7646
0.8391	0.7656
0.8249	0.7484
0.8037	0.6735

Overall Descriptions for Top 10 Systems

	# Rule-based	# Hybrid	# ML	# Used MDs
Concepts	0	3	7	1
Relations	1	2	7	0
End-to-end	0	3	7	1

External resources used: MIMIC-III, discharge-summary db, cTAKES, clamp, MedEx, Cadec, i2b2 2009, i2b2 2010, gazetteers, GloVe, nltk, sklearn, networkx, spacy, pyConText, PubMed, UMLS, SNOMED, RxNorm, word2vec

Concepts – Top 10 Teams

Team	System description
Alibaba Inc.	Language-level BiLSTM + CNN for character-level + CRF for dependencies
UTHealth / Dalian	CRF + BiLSTM-CRF for entities with category-level best prediction
UFlorida	MIMIC + word2vec ->LSTM-CRFs + RCNN-CRFs with a convolutional layer
UniMan	BiLSTM-CRF with pretrained word2vec skip-gram model on MIMIC-III
MUSC	sequential taggers: CRF, SEARN, and RNN
UniMan / TTI / AIST	CRF + stacked BiLSTM-CRF + byte pair encoding
NTUH / NTU	CRF + LSTM-LSTM-CRF + CNN-LSTM-CRF + handcrafted features
VA-Utah	Bi-LSTM-CRF + CRF model trained on MADE 1.0 data
IBM Research	BiLSTM-CRF lexical, syntactic, and morphological features
UMich / Ramakrishna	Pre-processing with cTAKES, MetaMap, MedEx, and MedXN -> LSTM + CRF

Relations – Top 10 Teams

Team	System descriptions
UTHealth/Dalian	SVM + CNN-RNN on gold entities
VA-Utah	RandomForest: candidate entities, entities between candidates, syntactic features
UniMan / TTI / AIST	walk-based neural network
UFlorida	MIMIC + word2vec ->LSTM-CRFs + RCNN-CRFs with a convolutional layer
Med Data Quest	Attention-Based BiLSTM
IBM Research	piecewise-BiLSTM with attention mechanism
MUSC	binary-class SVM classifier w/ distance, lexical, and word embedding features
UniMan	LSTMs + word embeddings from MIMIC-III + positional information of entities
BCH / HMS / Loyola	cTAKES
CCHMC	rule-based

End-to-End – Top 10 Teams

Team	System descriptions
UTHealth/Dalian	CRF + BiLSTM-CRF for entities -> CNN-RNN for relations
UFlorida	MIMIC + word2vec -> LSTM-CRFs + RCNN-CRFs with a convolutional layer
UniMan / TTI / AIST	CRF + BiLSTM-CRF + byte pairs -> walk-based neural network
MUSC	CRF + SEARN + RNN for entities -> binary-class SVM
VA-Utah	Bi-LSTM-CRF + CRF -> Random Forest
IBM Research	BiLSTM-CRF -> piecewise-BiLSTM
UniMan	word-based BiLSTM-CRF for NER + random forest for relations
CCHMC	External corpora + CRFs -> rule-based relations
BCH / HMS / Loyola	cTAKES + UMLS TUI -> cTAKES
Roam Analytics	Expert annotation + CRFs -> binary classifiers

What's Next?

- Journal special issue
 - Being planned.
- Data
 - Available for research outside of the shared task in November 2019, from
<https://n2c2.dbmi.hms.harvard.edu/track1.php>
- New shared tasks
 - Coming spring/summer 2019.

New Tasks

- Organized by you?
 - Contact ouzuner@gmu.edu if you have annotated any i2b2, CEGS-NGRID, or n2c2 data and want to make it available to the community for a shared task

Thank you!

- National Library of Medicine, National Institutes of Health, NIH NLM 5R13LM011411, PI: Ozlem Uzuner.
- Department of Biomedical Informatics, Harvard Medical School.
- Annotators!
- Participants!!



HARVARD
MEDICAL SCHOOL

DEPARTMENT OF
Biomedical Informatics



Volgenau School
of Engineering