

Extracting adverse drug reactions and their context from drug labels using sequence labelling ensembles

Maksim Belousov¹, Nikola Milošević¹, William Dixon^{2,3}, Goran Nenadić^{1,3}

¹School of Computer Science, The University of Manchester, ²Arthritis Research UK Centre for Epidemiology, The University of Manchester, UK, ³Health e-Research Centre

Motivation

- Adverse drug reactions (ADRs) present a challenge for drug development and administration that harms millions of patients.
- About 7% of all hospital admissions were accounted to ADRs. Moreover, quality of life and adherence to treatment is affected by adverse drug reactions
- In the United States, drug product labels are required by law to contain the information regarding clinically significant adverse drug reactions. All drug product labels are available through the DailyMed website in a Structured Product Label (SPL) format.

Drug Facts	Drug Facts (continued)	Drug Facts (continued)
Active ingredient (in each tablet) Aspirin 81 mg (NSAID).....Pain reliever "nonsteroidal anti-inflammatory drug"	Purpose Temporarily relieves minor aches and pains	Warnings Children and teenagers who have or are recovering from chicken pox or flu-like symptoms should not use this product. When using this product, if changes in behavior with nausea and vomiting occur, consult a doctor because these symptoms could be an early sign of Reye's syndrome, a rare but serious illness.
Uses Temporarily relieves minor aches and pains	Warnings (continued) Children and teenagers who have or are recovering from chicken pox or flu-like symptoms should not use this product. When using this product, if changes in behavior with nausea and vomiting occur, consult a doctor because these symptoms could be an early sign of Reye's syndrome, a rare but serious illness.	Warnings (continued) Children and teenagers who have or are recovering from chicken pox or flu-like symptoms should not use this product. When using this product, if changes in behavior with nausea and vomiting occur, consult a doctor because these symptoms could be an early sign of Reye's syndrome, a rare but serious illness.
Directions Take 1 or 2 tablets every 4 to 6 hours while symptoms last. Do not exceed 48 tablets in 24 hours or as directed by a doctor. Children under 12 do not use unless directed by a doctor.	Directions (continued) Take 1 or 2 tablets every 4 to 6 hours while symptoms last. Do not exceed 48 tablets in 24 hours or as directed by a doctor. Children under 12 do not use unless directed by a doctor.	Directions (continued) Take 1 or 2 tablets every 4 to 6 hours while symptoms last. Do not exceed 48 tablets in 24 hours or as directed by a doctor. Children under 12 do not use unless directed by a doctor.
Other information Store at 20°-25° (68°-77°F).	Other information (continued) Store at 20°-25° (68°-77°F).	Other information (continued) Store at 20°-25° (68°-77°F).
Questions? 1-800-719-9260		

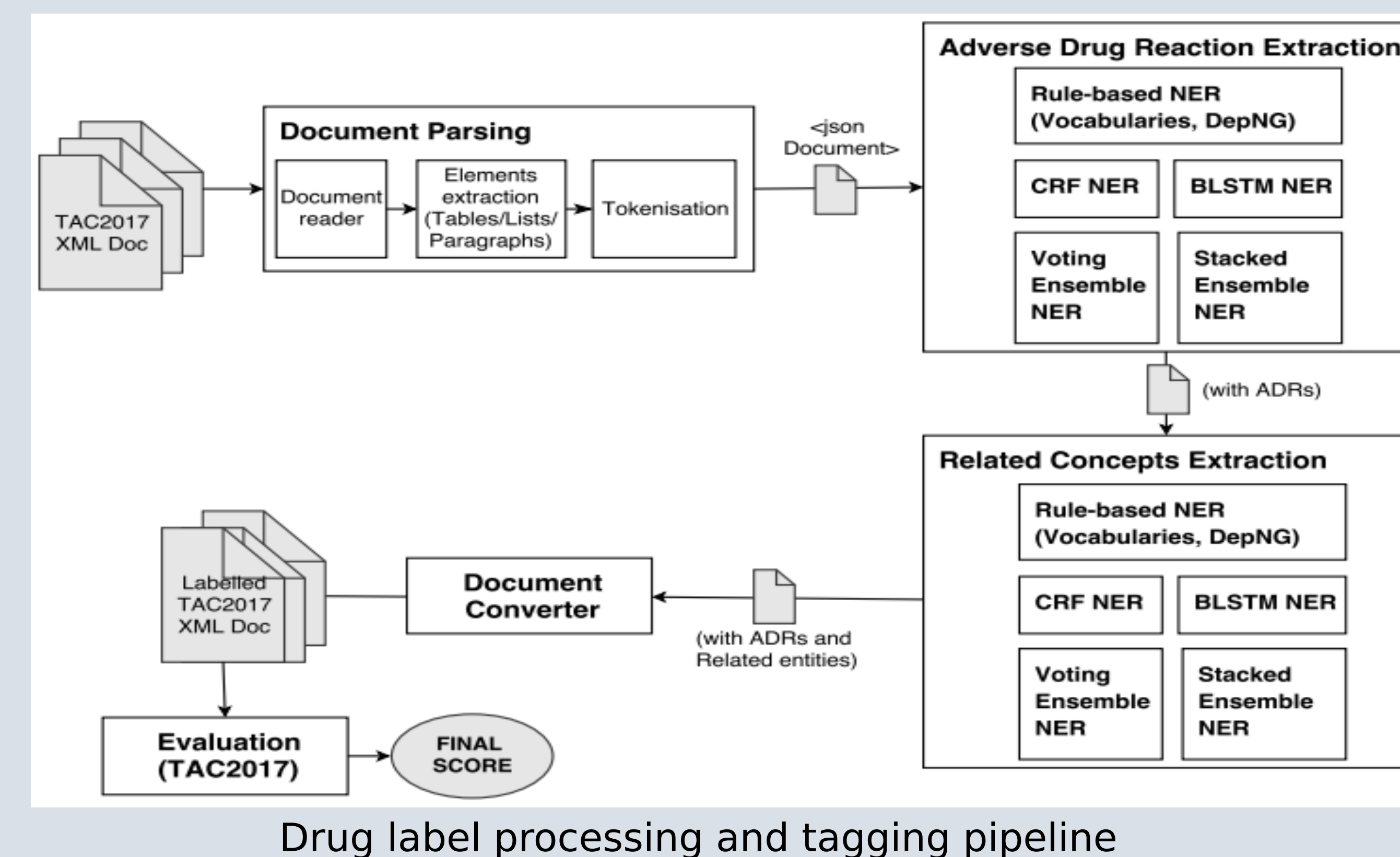
- This work presents the output of the shared task (TAC 2017 ADR Track) which objective was to extract adverse drug reactions from drug labelling text documents.
- The dataset contains 101 annotated and 2,208 unannotated drug labels (documents).

Entity class	#mentions	# tokens	Avg. token/mention
Adverse drug reaction	12,792	21,258	1.66
Severity	863	1306	1.51
Factor	602	653	1.08
Drug class	248	518	2.09
Negation	95	109	1.47
Animal	44	44	1.00

Statistics of entity classes presented in the annotated dataset

Methods

- The system used a mix of rule-based, machine learning and deep learning models in order to recognise entities in the drug labelling documents.
- The rule-based models were developed for Negation (using a list of negation clues and dependencies from GENIA parser) and Animal (using a dictionary) classes.
- The Conditional Random Fields (CRF) model used lexical features, part-of-speech tags, grammatical relations (dependencies), vocabulary and semantic features (such as corresponding semantic types and named entity tags from various medical NER systems).
- The Bidirectional Long Short-Term Memory Networks (BLSTM) model were trained on 200-dimensional word vectors obtained by applying skip-gram model on biomedical literature and English Wikipedia dump.
- We also experimented with ensemble methods, namely voting (BLSTM+CRF) and stacking (CRF+BLSTM)



Evaluation and results

- We initially evaluated the machine learning models on the development dataset

Entity class	Method	Precision	Recall	F1-score
ADR	CRF	90	82	86
	BLSTM	86	84	85
	Voting BLSTM+CRF	91	84	87
	Stacked CRF+BLSTM	90	85	87
	Rule-based	86	100	93
Severity	CRF	67	51	58
	BLSTM	55	75	64
	Voting BLSTM+CRF	70	65	67
	Stacked CRF+BLSTM	58	71	64
	Rule-based	66	66	66
Factor	CRF	52	20	29
	BLSTM	73	46	56
	Voting BLSTM+CRF	87	36	51
	Stacked CRF+BLSTM	82	41	55
	Rule-based	66	66	66
Drug class	CRF	41	35	38
	BLSTM	57	21	31
	Voting BLSTM+CRF	62	12	20
	Stacked CRF+BLSTM	57	24	34
	Rule-based	66	66	66
Negation	CRF	25	18	21
	BLSTM	22	12	15
	Voting BLSTM+CRF	50	06	11
	Stacked CRF+BLSTM	57	24	33
	Rule-based	66	66	66
Animal	CRF	76	100	87
	BLSTM	100	46	63
	Voting BLSTM+CRF	100	38	56
	Stacked CRF+BLSTM	40	31	35
	Rule-based	86	100	93

Token-level evaluation of taggers on development dataset

- The final evaluation of 2 hybrid system was made on the unannotated (unseen) dataset.
- Run #1: Rule-based models for the Negation and Animal classes. ADR class utilised the CRF with the hand-crafted features. For all other entities we used the BLSTM tagger.
- Run #2: The rule-based tagger was applied only for the Negation class, whereas all other classes were handled with the Stacked CRF+BLSTM ensemble.

Submission	Considering entity type			Not considering entity type		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Run #1	80.19	72.23	76.00	80.19	72.23	76.00
Run #2	76.84	74.26	75.58	76.87	74.39	75.61

Performance of the submitted systems on test dataset