TEMPORALLY SENSITIVE PHENOTYPES: THE USE CASE OF DRUG-INDUCED LIVER TOXICITY

Chen Lin

CHIP-NLP LAB Boston Children's Hospital 7/9/2014



PUBLICATION:

Chen Lin, *Elizabeth W. Karlson, *Dmitriy Dligach, Monica P. Ramirez, Timothy Miller, Huan Mo, Natalie S. Braggs, Andrew Cagan, Vivian Gainer, Joshua Denny, Guergana K. Savova, (under review). Automatic Identification of Methotrexate-induced Liver Toxicity in Rheumatoid Arthritis Patients from the Electronic Medical Record. Journal of the American Medical Informatics Association.

* Indicates co-first authors

- FUNDING SOURCES: PGRN Pharmacogenomics Research Network
- i2b2 Informatics for Integrating Biology and the Bedside ٠ (U54LM008748 from the National Library of Medicine)
- **THYME Temporal Histories of Your Medical Events** • (R01LM010090 from the National Library of Medicine)

RHEUMATOID ARTHRITIS (RA) PATIENTS AND METHOTREXATE (MTX) -- LIVER TOXICITY

- MTX the first line therapy for RA since 1980
- MTX could cause liver problem (cumulative elevation of LFT)
- Define a large cohort of CASE patients for a genetic study
- CASE patients:
 - RA patients
 - MTX treatment
 - Elevated LFT
 - MTX within 3 month before LFT elevation
 - Not on other drugs

Temporally Sensitive

No specific comorbidities present

(Hepatitis, Cardiac/Arrhythmia, IBD, Sepsis, Infections, Trauma, Surgery, etc.)

EMR-BASED ALGORITHM

RA Cohort: Liao et al. 2010







CASE RA patients: MTX induced Liver toxicity

narrative and codified Electronic Medical Records (EMR)

Non-CASE RA patients

RULE-BASED FILTER

RA Cohort: Liao et al. 2010







Non-CASE RA patients

narrative and **codified** Electronic Medical Records (EMR) Have not taken MTX ever No elevated LFT

CASE patients: RA patients On MTX Elevated LFT MTX within 3 month before LFT elevation Not on other drugs No specific comorbidities present

TEMPORALLY SENSITIVE CLASSIFIER

RA Cohort: Liao et al. 2010



OUTLINE

- Objectives
- Temporal Module relation to document creation time (DocTimeRel)
- Method
 - Flow, data representation, feature selection and classification
- Data Sets
 - Training Set, Testing Set 1 and Testing Set 2
- Results

Document Creation Time Relation (DocTimeRel)

 DocTimeRel: the temporal relation of the event in question relative to the time when the record was written i.e. Document Creation Time (DCT).



APACHE CTAKES: DocTimeRel model

- DocTimeRel model part of cTAKES-Temporal module (3.2 release)
- cTAKES-DocTimeRel model identifies the document creation time relations for all events such as
 - if a patient is currently on a Medication/Procedure
 - if a specific Comorbidity is present
- Trained on THYME corpus (78 clinical notes across 26 patients; thyme.healthnlp.org)
 - clinical and pathology notes on colorectal cancer from Mayo Clinic
 - 9,730 annotated instances of DocTimeRels
- Performance on THYME corpus: F1-score: 0.814

(state of the art performance in general domain: 0.82)

The Patient is a 49 yo female with history of systemic infection. She was admitted yesterday for pneumonia with a blood pressure of 80/50.
Medication: BEFORE
Methotrexate 15 mg Q week
Zestril 5 mg QD
Surgical History:
Amputation of lower limb OVERLAP
OVERLAP
BEFORE

OUTLINE

- Objectives
- Temporal Module -- DocTimeRel
- Method
 - data representation, encapsulation and classification
- Data Sets
 - Training Set, Testing Set 1 and Testing Set 2
- Results

-

DOCUMENT REPRESENTATION

CASE patients:

RA patients

Elevated LFT

MTX within 3 month before LFT elevation Not on other drug No specific comorbidities present

- **Customized Dictionary:**
 - MTX
 - Leflunomide (Arava)
 - Comorbidities:
 - Hepatitis, Cardiac, IBD, Sepsis, Infections, Trauma, Surgery, ...
- Document Dictionary item vector
- Section Parsing
 - Ignore irrelevant sections, e.g. history
 - Emphasize "Medication section"

The yester	The Patient is a 49 yo female with history of systemic infection. She was admitted yesterday for pneumonia with a blood pressure of 80/50.							
Methotrexate 15 mg Q week Zestril 5 mg Q					LI OKL	Main Section	n	
Surgic Amput	al History: ation of lower	limb						
OVEF	OVERLAP Medica				1			
OVERLAP Not in Dictionary								
Surgical History Section								
Customized CUIs			MTX Signature Temporal: DocTimeRel, nearby words and POS			el, S		
rstem fection 0243026	Pneumonia C0032285	MTX C0025677	C0025677 In_Med	MTX Dosage: 15 mg	MTX Route:ora I	C0243026 BEFORE	C0032285 OVERLAP	C0025677 OVERLAI

DOCUMENT ENCAPSULATION

- Episode: a 3-month window prior to each elevated LFT
 - For all documents within an episode: Doc-level vector → episode-level vector
- Classify each episode into CASE vs. NON-CASE
- Summarize to patient-level results:
 - If one episode is CASE positive, the patient is a CASE;
 - If all episodes are NON-CASE, the patient is a NON-CASE



Patient clinical data: LFT abnormality is derived from codified data

CLASSIFIER

- L2-regularized Logistic Regression as implemented by LIBLINEAR
- Fast convergence
- Classification of thousands of episodes and then summarization to patient-level results

OUTLINE

- Objectives
- Temporal Module -- DocTimeRel
- Method

data representation, encapsulation and classification

• Data Sets

Training Set, Testing Set 1 and Testing Set 2

Results

DATA SETS

• BWH RA Sequential Study (BRASS)

Set	CASE	NON-CASE	Total	Inter-annotator agreement (kappa)
Training	132	348	480	Single labeled
Test				

• Vanderbilt University (VU)

Set	CASE	NON-CASE	Total	Inter-annotator agreement (kappa)
2nd Test	41	62	103	Single labeled

OUTLINE

- Objectives
- Temporal Module -- DocTimeRel
- Method
 - data representation, encapsulation and classification
- Data Sets
 - Training Set, Testing Set 1 and Testing Set 2
- Results

RESULT ON TRAINING SET (10 FOLD CROSS VALIDATION)

Features	PPV	Recall	F1-score
Customized CUIs		0.806	0.782
Customized CUIs+ MTX signature+ DocTimeRel		0.868	0.798

RESULT ON TEST SETS

Models	PPV	Recall	F1-score
Traditional BOW model	0.590	0.703	0.642
Our classifier	0.756	0.919	0.829

	PPV	Recall	F1-score
Vanderbilt University	0.66	0.853	0.745

DISCUSSION

- Episode-level classification problem
- Features: CUI-coded customized dictionary, temporal signals, drug signatures
- Section parsing
- Methodology for temporally sensitive phenotypes

QUESTIONS?

Thank you!