uc3m

meaning cloud

Machine learning vs. knowledge based approaches to ADR identification

November 2017

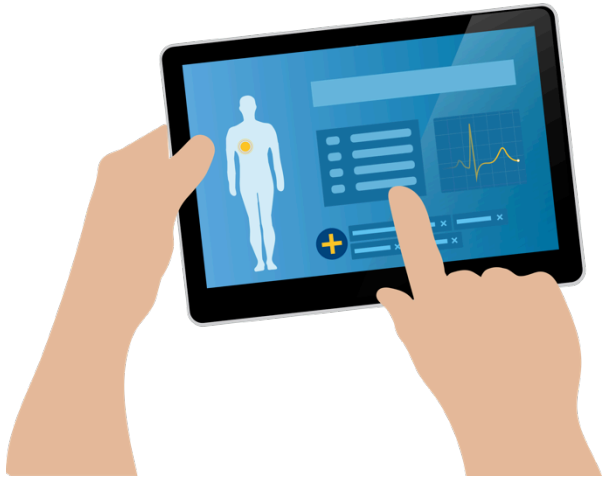
Topics

- Short about us
- Identifying ADRs
- Machine Learning for semantic relations identification
- Results
- Challenges

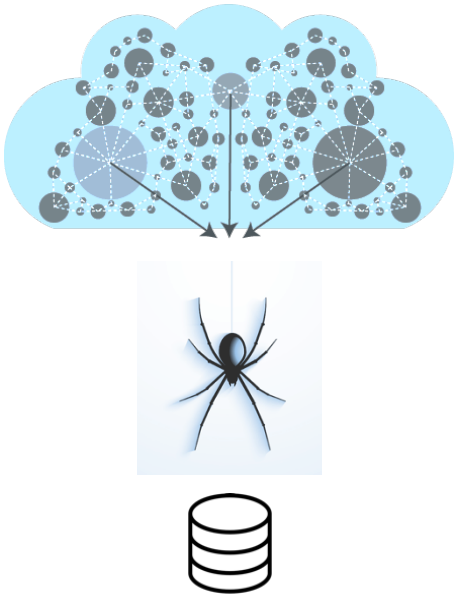
SHORT ABOUT US

Focus on text-analytics for Pharmaceuticals. Since 1998

Voice of the Patient



Electronic Health Records



Other text sources

Scientific literature – FDA –
Patents – Business opportunities

SHORT ABOUT US

Advanced Databases Group, Universidad Carlos III de Madrid

- Research lines:
  - Natural language processing
  - Accessibility
- Resources produced:
  - ***Drug-drug-interaction collection (DDI Corpus)***
  - ***DINTO ontology***

uc3m

Our goal at TAC ADR

“Combine Knowledge Based with Machine Learning Based approaches to leverage ADR identification”

Identifying ADRs

TOPIC EXTRACTION

NLP and Resource based approach

- SIDER
- UMLS
- Training corpus

Elements

- Entries
- Ontology
- Settings

Actions

Build | Delete | Import | Export | Test

Do you need more information? Check out the documentation, or just drop us a line through our support form.

Entries

Show 10 entries Search:

Actions	ID	Form	Type	Ontology Type	Last updated
	59c92278ba02d	"penile ""popping"" sound"	ENTITY	AdverseReaction	2017-09-25 17:36:24
	59c9228a27fc7	1st degree heart block	ENTITY	AdverseReaction	2017-09-25 17:36:42
	59c9228b00efd	2nd degree heart block	ENTITY	AdverseReaction	2017-09-25 17:36:42
	59c9248d2cd5f	5'nucleotidase increased	ENTITY	AdverseReaction	2017-09-25 17:45:17
	59c9247da8b7c	A-hypervitaminosis	ENTITY	AdverseReaction	2017-09-25 17:45:01
	59c923b661f2f	A-V dissociation	ENTITY	AdverseReaction	2017-09-25 17:41:42
	59c922a13f4d5	A.R.D.S.	ENTITY	AdverseReaction	2017-09-25 17:37:05
	59c92479057bb	AAION	ENTITY	AdverseReaction	2017-09-25 17:44:57
	59c9247909336	AAION	ENTITY	AdverseReaction	2017-09-25 17:44:57
	59c92484e3ed0	Abasia	ENTITY	AdverseReaction	2017-09-25 17:45:08

(21826 entries)

Previous 1 2 3 4 5 … 2183 Next

+ Add new entry

TOPIC EXTRACTION

NLP and Resource based approach

- SIDER
- UMLS
- Training corpus

sementity:

class: instance

type: Top>AdverseReaction

meddra_llt: AAION

meddra_llt_id: 10068247

meddra_pt: Optic ischaemic neuropathy

meddra_pt_id: 10030924

cui_id: C2242711

meddracui_id: C0155305

source: SIDER

sementity:

class: instance

type: Top>AdverseReaction

meddra_llt: AAION

meddra_llt_id: 10068247

meddra_pt: Arteritic anterior ischaemic optic neuropathy

meddra_pt_id: 10030924

cui_id: C2242711

meddracui_id: C2242711

source: SIDER

Dictionary	#entries
Adverse Reactions	21,826
Factor	41
Severity	158
Animal	27
DrugClass	101

NLP and Resource based approach

- And some rules to identify negation:

```
<<without|exclude|decrease|reduce>> :-
ENTITY{"type":"NegationLeft", "label":"$1"}
+ CONSUME{}
<<{AFFECTEDADR}>> :-
ENTITY{"type":"AffectedAdr", "label":"$1"}

"...without associated bleeding events. ..."
```

- MeaningCloud Insights Engine API supports this rule syntax

Machine Learning for semantic relations identification

Machine learning for semantic relations identification

Representing ADR mention context through a set of features:

ADRMention – Other pairs (where Other is Severity, DrugClass, Negation, Animal or Factor)

- M1TXT, M2TXT, BWTXT: the text of both/between mentions.
- C1BOW, C2BOW: bag-of-words of both mentions.
- C1POS, C2POS: part of speech of both mentions.
- PB1POS, PA1POS, PB2POS, PA2POS, PWPOS: the PoS tags of the two tokens before/ after/between both mentions.
- WA1TXT, WB2TXT, WA2TXT, WB1TXT: the two tokens after/before the mention.
- LA1LEM, LB2LEM, LA2LEM, LB1LEM: the lemmas of the two tokens after/before both mentions.
- LWLEM: the lemmas between of the two mentions
- NTOKB: the number of tokens between the two mentions.

Machine learning for semantic relations identification

And the algorithm?

- SVM, support vector machines (using scikit-learn on Python)

- Specifically, SVC implementation:
  - Default parameter values
  - Linear kernel

But, no deep learning??!!

Of course (CNN), but not in the official runs.

Results

Results

- Task 1. ADR and related entities

Type	P	R	F1
Exact (+type)	54.79	66.33	60.01
Exact (-type)	55.78	66.34	60.60

Table 3: Task 1 results on the test set.

Type	P	R	F1
AdverseReaction	63.82	70.77	67.12
Severity	37.13	49.52	42.44
Factor	4.05	7.65	5.3
Negation	10.59	53.76	17.7
DrugClass	19.23	39.63	25.9
Animal	76.56	56.98	65.33
Macro	54.79	66.33	60.01

Table 4: Task 1 results by type of mention on the test set.

Low precision!!

- Task 2. Relations between ADRs and entities

Type of Relation	P	R	F1
Negated	8.43	4.86	6.17
Hypothetical	5.95	9.56	7.34
Effect	24.94	25.74	25.33
Macro	12.19	15.59	13.68

Table 5: Task 2 results on the test set.

Type of Relation	P	R	F1
Negated	1.12	27.59	2.15
Hypothetical	35.5	52.49	42.36
Effect	24.93	48.77	32.99
Macro	46.7	49.97	47.32

Table 6: Task 2 results using correct mentions on the test set.

Oh, oh!!

Results

- Task 3. Positive ADRs

	P	R	F1
Micro	70.03	71.42	70.71
Macro	69.23	72.93	70.13

Table 7: Task 3 results on the test set.

Pretty good!!
Only a few negated mentions?

- Task 4. Normalization through MedDRA

	P	R	F1
Micro	73.40	80.25	76.67
Macro	72.10	80.38	75.29

Table 8: Task 4 results on the test set.

Using dictionaries with semantic information produces nice results

Challenges

Challenges

- Negation identification requires more effort (not only on the ADRs field). Some weird things found in the test set:

  Eg.: The most frequently observed malignancies other than non-melanoma skin cancer …

  Negation?

- CNNs and the use of syntactic features improves results

	P	R	F1
Other	0.71	0.81	0.76
Negated	0.72	0.40	0.51
Hypothetical	0.75	0.75	0.75
Effect	0.76	0.61	0.68
Avg / total	0.73	0.73	0.73

Challenges

- Recall must be improved:
  - separated multiword mentions
  - ADRs with no MedDRA code, enough lexical resources?

- How to approach errors when applying deep learning?

- Enough accuracy for practical applications? What does FDA think?

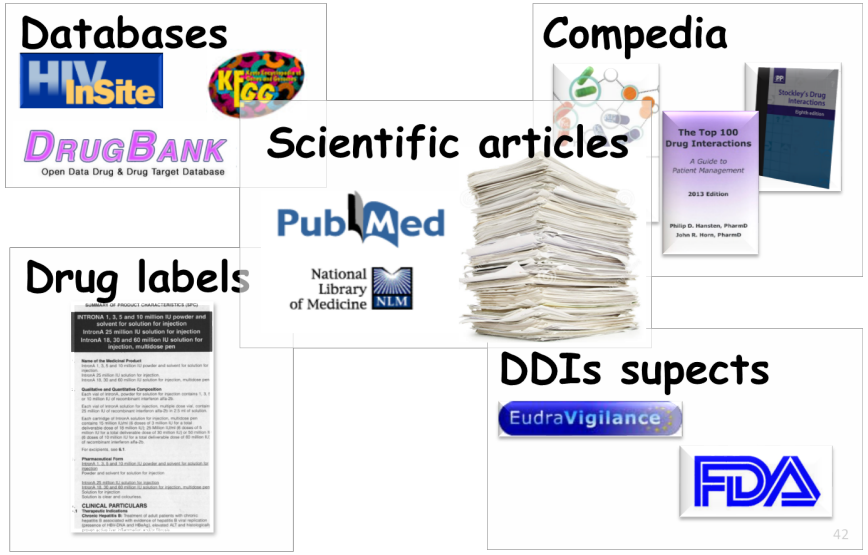
Thanks

QUESTIONS?

LabDA Resources

Corpus DDI (Drug-Drug Interactions)

- 1,025 annotated documents, 18,502 entities and 5,028 DDIs (by expert pharma)
- MedLine and DrugBank texts
- Annotatins guidelines and interannotator agreement.
- Available at labda.inf.uc3m.es
- Used at DDIExtraction 2011 and DDIExtraction 2013 Semeval Tasks



/ddi2013/DrugBank/DrugBank_d59

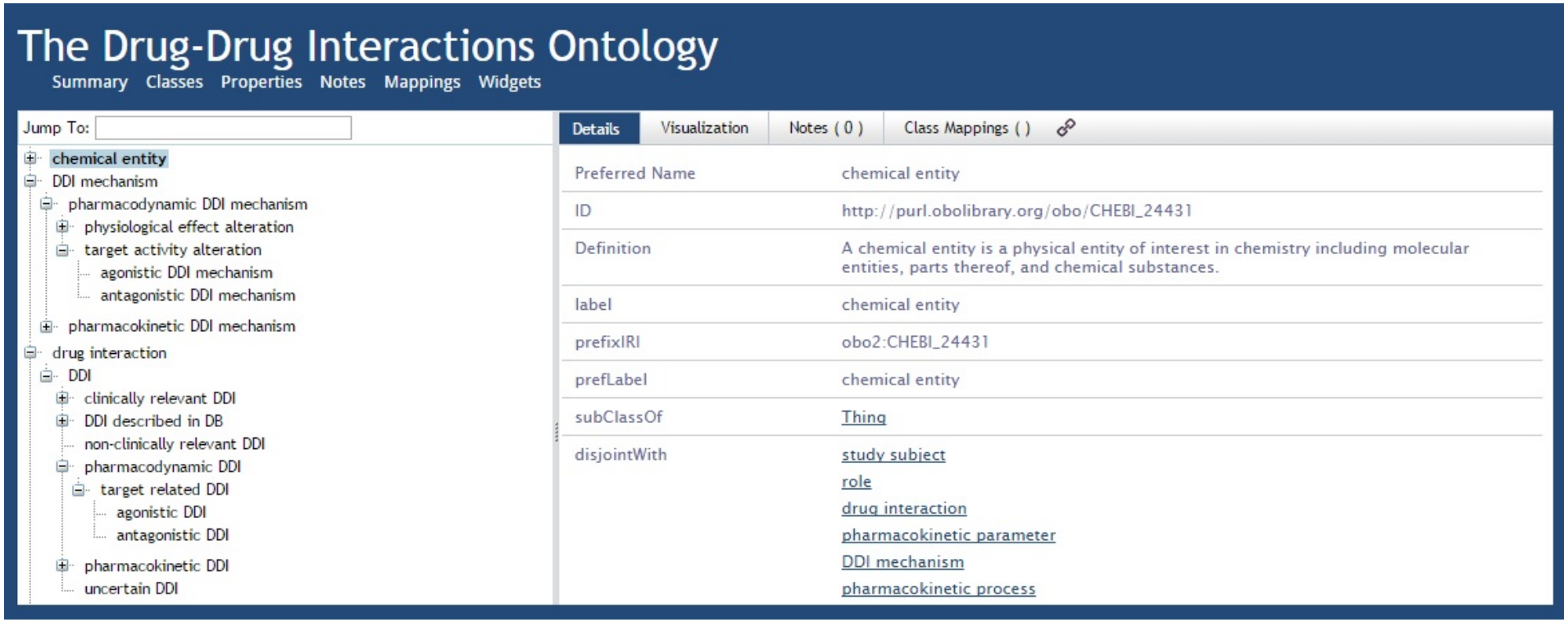
1 The effects of ERGOMAR (brand) may be potentiated by triacetyloleandomycin (drug) which inhibits the metabolism of ergotamine (drug).The pressor effects of ERGOMAR (brand) and other vasoconstrictor drugs (group) can combine to cause dangerous hypertension.

DDI: ERGOMAR – triacetyloleandomycin
DDI: triacetyloleandomycin – ergotamine
DDI: ERGOMAR – vasoconstrictor drugs

LabDA Resources

DINTO Ontology- knowledge about drugs and interations (11,555 DDIs and 8,786 pharmacological entities). Available at OBO Foundry

Application to Information Extraction and Prediction



MeaningCloud LLC

Automating the extraction of Meaning from any information source.

Address

35-37 36th Street
11106 Astoria NY

Contact Info

jmartinez@meaningcloud.com

Telephone

Phone: +1 (646) 403-3104

meaningcloud.com