Deep Text Prior: Weakly Supervised Learning for Assertion Classification

Vadim Liventsev[1,2,3] Irina Fedulova[1] Dmitry Dylov[2]

[1]Work done at Philips Innovation Labs RUS
{Vadim.Liventsev,Irina.Fedulova}@philips.com

[2]Work done at Skolkovo Institute of Science and Technology
{Vadim.Liventsev,D.Dylov}@skoltech.ru

[3]Currently at Eindhoven University of Technology
v.liventsev@tue.nl

September 22, 2019

Pop quiz

INDICATION: Evaluate for *pneumonia*

Pop quiz

INDICATION: Evaluate for *pneumonia*

A. This patient has *pneumonia*
B. This patient does not have *pneumonia*
C. We don't know

Pop quiz

IMPRESSION: No evidence of *pneumonia*

A. This patient has *pneumonia*
B. This patient does not have *pneumonia*
C. We don't know

Pop quiz

IMPRESSION: No evidence of *pneumonia*

A. This patient has *pneumonia*
B. This patient does not have *pneumonia*
C. We don't know

Pop quiz

FINDINGS: Developing *pneumonia* should be excluded by repeat image.

A. This patient has *pneumonia*
B. This patient does not have *pneumonia*
C. We don't know

Pop quiz

FINDINGS: Developing *pneumonia* should be excluded by repeat image.

A. This patient has *pneumonia*

B. This patient does not have *pneumonia*

C. We don’t know

Pop quiz

IMPRESSION: Effusions represent area of atelectasis, although *pneumonia* could also have this appearance.

A. This patient has *pneumonia*
B. This patient does not have *pneumonia*
C. We don't know

Pop quiz

IMPRESSION: Effusions represent area of atelectasis, although *pneumonia* could also have this appearance.

A. This patient has *pneumonia*
B. This patient does not have *pneumonia*
C. We don't know

Assertion classification problem

Features:

$$\underbrace{\text{INDICATION}}_{\text{metadata}} : \text{Evaluate for } \underbrace{\text{pneumonia}}_{\text{concept of interest}} \tag{1}$$

Assertion classification problem

Features:

$$\underbrace{\text{INDICATION}}_{\text{metadata}} : \text{Evaluate for } \underbrace{\text{pneumonia}}_{\text{concept of interest}} \tag{1}$$

Classes:

- **positive** the author states that the patient *has* the condition
- **negative** the author asserts that the patient does *not* have the condition
- **speculative** the author mentions the condition, but *does not assert anything* as to whether the patient has it

Landscape of Artificial Intelligence in Medicine

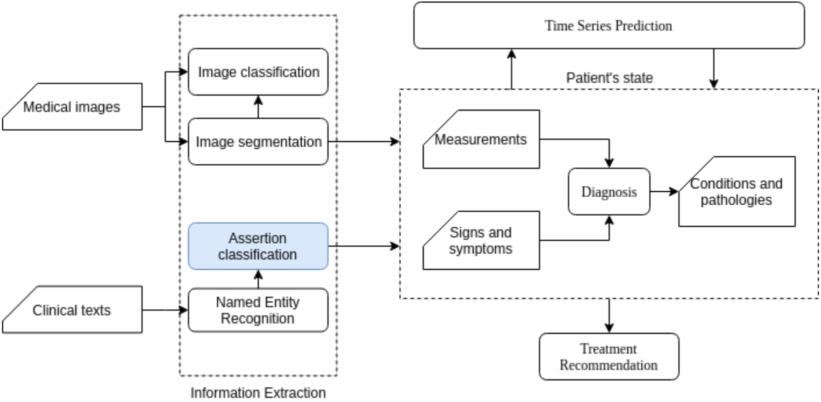


Figure: Medical AI task graph

Landscape of Artificial Intelligence in Medicine

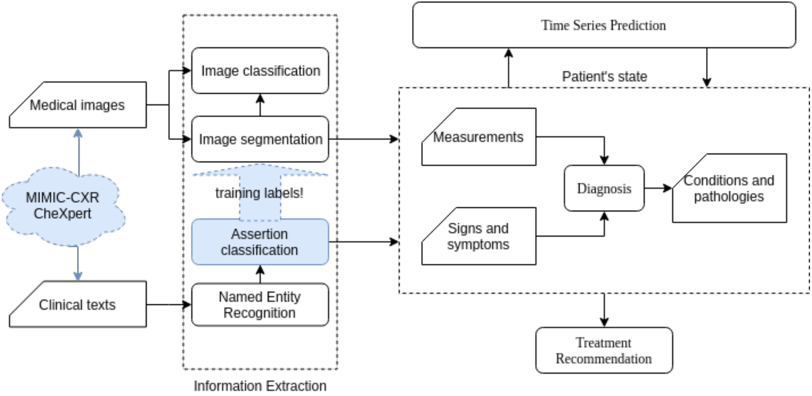


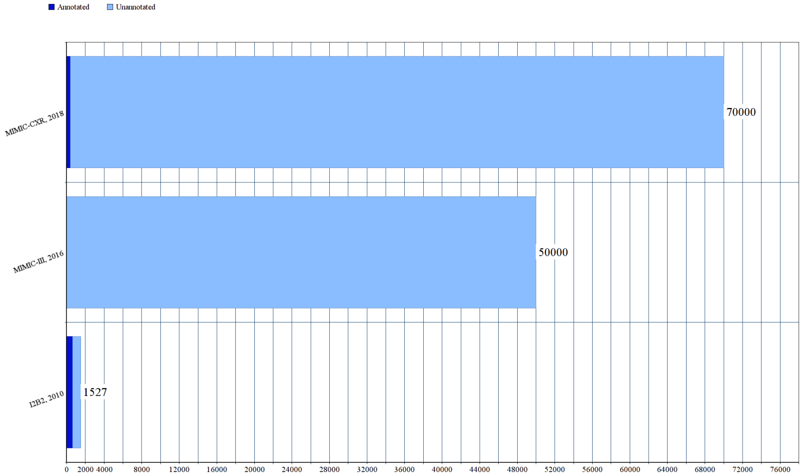
Figure: Medical AI task graph

Datasets

```
[**Hospital 9**] MEDICAL CONDITION:
  64 year old immunocompromised women with persistent cough/SOB and fluid
  overload
 REASON FOR THIS EXAMINATION:
  ?pna, pleural effusions
 ______________________________________________________________________________
                                 FINAL REPORT
 CHEST RADIOGRAPH

 INDICATION:  Immunocompromised woman, shortness of breath.

 COMPARISON:  [**2192-12-8**].

 FINDINGS:  As compared to the previous radiograph, there is no relevant
 change.  The lung volumes have increased.  The monitoring and support devices
 are all unchanged.  Unchanged scarring at the left and right lung bases but no
 newly appeared parenchymal opacity.  Unchanged size of the cardiac silhouette.
```

Figure: A sample radiology report from MIMIC-CXR

Datasets


Annotated
Unannotated
MIMIC-CXR, 2018
MIMIC-III, 2016
I2B2, 2010
70000
50000
1527
0 2000 4000 8000 12000 16000 20000 24000 28000 32000 36000 40000 44000 48000 52000 56000 60000 64000 68000 72000 76000

Rule-based approaches

```
{} <{dependency:/nmod:of|nmod:for/} ({lemma:/evidence/} >{dependency:/neg/} {})
```

Figure: A sample negation cue from NegBio. Detects the phrase ”No evidence of/for X”

Typical pipeline:

1. Part of speech tagging
2. Dependency parsing
3. Complicated feature extraction
4. Negation and speculation cues

Rule-based approaches

```
{} <{dependency:/nmod:of|nmod:for/} ({lemma:/evidence/} >{dependency:/neg/} {})
```

Figure: A sample negation cue from NegBio. Detects the phrase ”No evidence of/for X”

Typical pipeline:

1. Part of speech tagging
2. Dependency parsing
3. Complicated feature extraction
4. Negation and speculation cues

Issues:

- Generalisation issues
- Language bias
- Dataset bias

Statistical approaches

Typical pipeline:

1. Part of speech tagging
2. Dependency parsing
3. Complicated feature extraction
4. Support Vector Machines

Statistical approaches

Typical pipeline:

1. Part of speech tagging
2. Dependency parsing
3. Complicated feature extraction
4. Support Vector Machines

Issues:

- Generalisation issues
- Language bias
- Dataset bias

Our approach

Incorporate as much prior knowledge as possible into our models:

- Incorporate metadata into assertion specification
- Use state of the art pretrained language models (ELMo) for sentence representation
- Use prototype network approach to incorporate relationships between classes
- Use specialized model architectures for the task at hand
- Use heuristic algorithms to pretrain the models with inexact supervision

Assertion representation

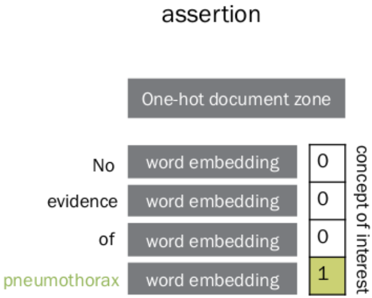


Figure: Vector representation of an assertion

Prototype networks

$$\mathrm{CMSE}(\mathrm{TC}, \mathrm{PC}, \mathrm{TR}, \mathrm{PR}) = (\mathrm{TC} - \mathrm{PC})^2 + \mathrm{TC} * (\mathrm{TR} - \mathrm{PR})^2 \quad (2)$$

- T true
- P predicted
- R reality
- C confidence

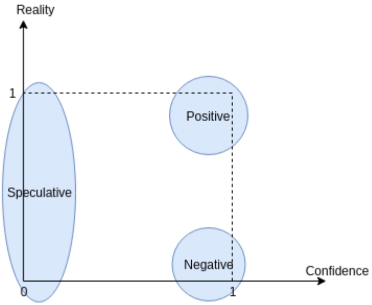


Figure: Class prototypes in reality-confidence space

Models

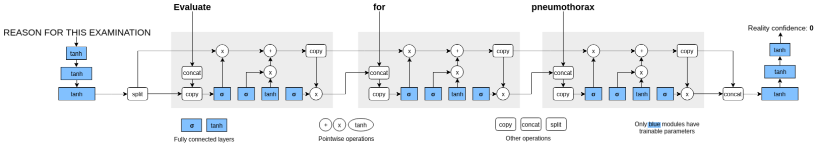


Figure: LSTM model

Models

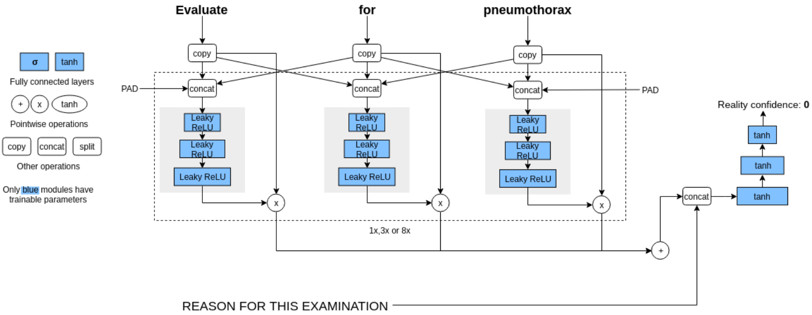


Figure: Attention-based model

NegBio+ heuristic algorithm

speculative if the assertion is in INDICATION, REASON FOR THIS EXAMINATION or similar section

negative if one of NegBio negation cues fires (”no X”, ”no evidence of X”)

positive otherwise

Weakly supervised learning schedules

1. *Classic*. Initialize weights randomly, then $\underset{\text{weights}}{\text{minimize}} ||\text{model}_{\text{weights}}(X) - y_{\text{true}}||$
2. *Deep Prior*. Initialize weights randomly, $\underset{\text{weights}}{\text{minimize}} ||\text{model}_{\text{weights}}(X) - \text{NegBio+}(X)||$.
3. *Transfer*. Use the weights obtained with *Deep Prior* as initialization, $\underset{\text{weights}}{\text{minimize}} ||\text{model}_{\text{weights}}(X) - y_{\text{true}}||$.

Cross-validation results

Table: Accuracy on I2B2 Challenge

	ctakes	0.796	0.796	0.796
Vectors	Model	Classic	Deep Prior	Transfer
elmo	attention1	0.625	0.704	0.756
elmo	attention8	0.621	0.706	0.738
elmo	lstm1024	0.798	0.755	0.866
elmo	lstm128	0.846	0.747	0.868
elmo	lstm512	**0.858**	0.733	**0.858**
elmo	lstm8	0.621	0.624	0.673

Cross-validation results

Table: Accuracy on MIMIC-CXR-FREQ

	NegBio+	0.834	0.834	0.834
Vectors	**Model**	**Classic**	**Deep Prior**	**Transfer**
elmo	attention1	0.615	0.803	0.932
elmo	attention3	0.863	0.225	0.950
elmo	attention8	0.919	0.576	0.944
elmo	lstm8	0.639	0.878	0.898
elmo	lstm128	0.927	0.873	0.967
elmo	lstm512	0.912	0.873	**0.975**
elmo	lstm1024	0.785	0.876	0.939
fasttext	attention1	0.610	0.429	0.870
fasttext	attention3	0.944	0.325	0.773
fasttext	attention8	0.778	0.441	0.838
fasttext	lstm8	0.276	0.388	0.058
fasttext	lstm128	0.705	0.914	0.929

Cross-validation results

Table: Accuracy on MIMIC-CXR-LONG

	NegBio+	0.71	0.71	0.71
Vectors	Model	Classic	Deep Prior	Transfer
elmo	attention3	0.800	0.770	0.683
elmo	attention8	0.833	0.710	**0.843**
elmo	lstm128	0.716	0.690	0.750
elmo	lstm512	0.763	0.710	0.691
elmo	lstm8	0.721	0.700	0.821
fasttext	attention3	0.722	0.680	0.821
fasttext	attention8	0.810	0.750	0.739
fasttext	lstm128	0.686	0.690	0.722
fasttext	lstm512	0.862	0.700	0.788
fasttext	lstm8	0.830	0.690	0.694

Result highlights

Table: Heuristic pretraining effect on MIMIC-CXR-FREQ

Vectors	Model	Classic	Transfer
elmo	attention1	0.615	+ 0.317
elmo	attention3	0.863	+ 0.097
elmo	attention8	0.919	+ 0.025
elmo	lstm8	0.639	+ 0.249
elmo	lstm128	0.927	+ 0.040
elmo	lstm512	0.912	+ 0.063
elmo	lstm1024	0.785	+ 0.154
fasttext	attention1	0.610	+ 0.260
fasttext	attention8	0.778	+ 0.050
fasttext	lstm512	0.748	+ 0.097
fasttext	lstm1024	0.578	+ 0.218

Result highlights

Table: Contextualized embeddings effect on MIMIC-CXR-FREQ

Model	Schedule	fasttext	elmo
lstm8	Classic	0.276	+ 0.363
lstm8	Deep Prior	0.388	+ 0.490
lstm8	Transfer	0.058	+ 0.848
lstm512	Classic	0.748	+ 0.164
lstm512	Deep Prior	0.381	+ 0.492
lstm512	Transfer	0.841	+ 0.134
lstm1024	Classic	0.578	+ 0.207
lstm1024	Deep Prior	0.309	+ 0.567
lstm1024	Transfer	0.796	+ 0.143
attention1	Classic	0.610	+ 0.164
attention1	Deep Prior	0.429	+ 0.374
attention1	Transfer	0.870	+ 0.062

Result highlights

Table: Deep prior effect on MIMIC-CXR-FREQ

Vectors	Model	Advantage over NegBio+
elmo	lstm8	+ 0.044
elmo	lstm128	+ 0.039
elmo	lstm512	+ 0.039
elmo	lstm1024	+ 0.042

Neural relaxation of algorithms

Given X and *heuristic*(X) one can outperform *heuristic*(X) by solving

$$\underset{\text{weights}}{\text{minimize}} ||\text{model}_{\text{weights}}(X) - \text{heuristic}(X)|| \tag{3}$$

Neural relaxation of algorithms

Given X and *heuristic*(X) one can outperform *heuristic*(X) by solving

$$\underset{\text{weights}}{\text{minimize}} ||\text{model}_{\text{weights}}(X) - \text{heuristic}(X)|| \quad (3)$$

Neural networks are priors!

On a different problem (mild/moderate/severe classification) we managed to achieve +30% improvement via this method

~~Commercial break~~ Acknowledgments

Organizations:

Skoltech
Skolkovo Institute of Science and Technology

PHILIPS

People:
Artem Shelmanov, Ilya Sochenkov, Francis Tyers

Ask us anything

Vadim.Liventsev@skoltech.ru
v.liventsev@tue.nl
Irina.Fedulova@philips.com
D.Dylov@skoltech.ru