Graph Convolutional Networks for Text Classification

Liang Yao

Liang Yao, Chengsheng Mao and Yuan Luo
{*liang.yao, chengsheng.mao, yuan.luo*}*@northwestern.edu*

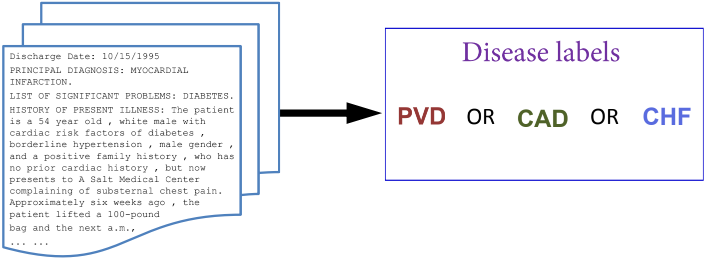
November 14, 2018

Overview

Problem

Text Classification

- A fundamental problem in Natural Language Processing (NLP)
- Many applications:
  - News filtering
  - Spam detection
  - Opinion mining
  - Computational phenotyping



Clinical Text Classification

- Relying on biomedical knowledge sources:
  - Including Unified Medical Language System (UMLS) Concept Unique Identifiers (CUIs) as features. [Garla and Brandt, 2013]
  - Projecting documents that contain related UMLS concepts closer together in a feature space. [Garla and Brandt, 2012]
- Using various types of information:
  - Regular expression [Bui and Zeng-Treitler, 2014]
  - Unlabeled corpora [Figueroa et al. 2012; Garla et al. 2013]
- Seldom using effective deep learning methods.
  - Recently widely used.
  - Powerful feature learning capabilities.
  - Successful in clinical data mining.

Deep Learning for Clinical Data Mining

- Two representative deep models:
  - Convolutional Neural Networks (CNN)
  - Recurrent Neural Networks (RNN)
- A number of tasks:
  - Entity extraction [Gehrmann et al. 2017]
  - Relations classification [Luo. 2017; Luo et al. 2017]
  - ICU time series analysis [Che et al. 2015]
  - Youth depression identification [Geraci et al. 2017]
- Not well studied in long clinical text classification.
- Failed to utilize knowledge sources.

Clinical Text Classification Challenges

- Challenge tasks:
  - Obesity Challenge
  - Smoking Challenge
  - Heart Disease Risks Challenge
  - . . .

Obesity Challenge

The i2b2 2008 obesity challenge [Uzuner, 2009]:

- To assess text classification methods for determining patient disease status with respect to obesity and 15 of its comorbidities.
- To label each document as either Present (Y), Absent (N), Questionable (Q) or Unmentioned (U) for each disease.

Obesity Challenge

The i2b2 2008 obesity challenge [Uzuner, 2009]:

- To assess text classification methods for determining patient disease status with respect to obesity and 15 of its comorbidities.
- To label each document as either Present (Y), Absent (N), Questionable (Q) or Unmentioned (U) for each disease.
- Macro F_1 score is the primary evaluation metric.
- Two tasks:
  - Textual task: identifying explicit evidences of the diseases.
  - Intuitive task: the prediction of the disease status when the evidence is not explicitly mentioned.

Obesity Challenge

The i2b2 2008 obesity challenge [Uzuner, 2009]:

- To assess text classification methods for determining patient disease status with respect to obesity and 15 of its comorbidities.
- To label each document as either Present (Y), Absent (N), Questionable (Q) or Unmentioned (U) for each disease.
- Macro F_1 score is the primary evaluation metric.
- Two tasks:
  - Textual task: identifying explicit evidences of the diseases.
  - Intuitive task: the prediction of the disease status when the evidence is not explicitly mentioned.
- The classes are distributed very unevenly.

Obesity Challenge

Table: The class distribution in the obesity challenge datasets.

Label	Training Set		Test Set	
	Textual	Intuitive	Textual	Intuitive
Y	3208	3267	2192	2285
N	87	7362	65	5100
Q	39	26	17	14
U	8296	0	5770	0

Obesity Challenge

- Most of top 10 systems are rule-based systems.
- Top 4 systems are purely rule-based.
- The overall 1st place system: Solt's system [Solt et al. 2009]. .
  - Can discover informative trigger phrases with Y, N or Q contexts:
    - *coronary artery bypass (Y)*
htn (Y)
dilated cardiomyopathy (Y)
...
    - *no evidence of cad (N)*
denied congestive heart failure (N)
w/o sob or cp chf (N)
...
    - *?dm (Q)*
presumed asthma (Q)
suggesting a recent development of chf (Q)
...

Method

Trigger Phrases Identification

Predicting Classes with Very Few Examples using Trigger Phrases

Knowledge-guided Convolutional Neural Networks

Trigger Phrases Identification

- We follow Solt's system to identify trigger phrases.
- Preprocessing: abbreviation resolution, family history removing.
- Using the disease names/alternatives, their directly associated terms and negative/uncertain words.
- The trigger phrases are disease names (e.g., Gallstones) and their alternative names (e.g., Cholelithiasis) with or without negative/uncertain words.

Predicting Classes with Very Few Examples using Trigger Phrases

- The classes in obesity challenge are very unbalanced, and some classes even don't have training examples.

Predicting Classes with Very Few Examples using Trigger Phrases

- The classes in obesity challenge are very unbalanced, and some classes even don't have training examples.
- Therefore we could not predict these classes using machine learning methods and resort to rules defined in Solt's system.

Predicting Classes with Very Few Examples using Trigger Phrases

- The classes in obesity challenge are very unbalanced, and some classes even don't have training examples.
- Therefore we could not predict these classes using machine learning methods and resort to rules defined in Solt's system.
- Excluding classes with very few examples in training sets.
  - Removing examples with Q or N label for textual task.
  - Removing examples with Q label in intuitive task.

Predicting Classes with Very Few Examples using Trigger Phrases

- Following Solt's system, we assume:
  - Positive trigger phrases > Negative trigger phrases > Uncertain trigger phrases
- Therefore:
  - If a clinical record has uncertain trigger phrases and dose not have positive/negative trigger phrases, we label it as Q.
  - If a clinical record has negative trigger phrases and dose not have positive trigger phrases, we label it as N.

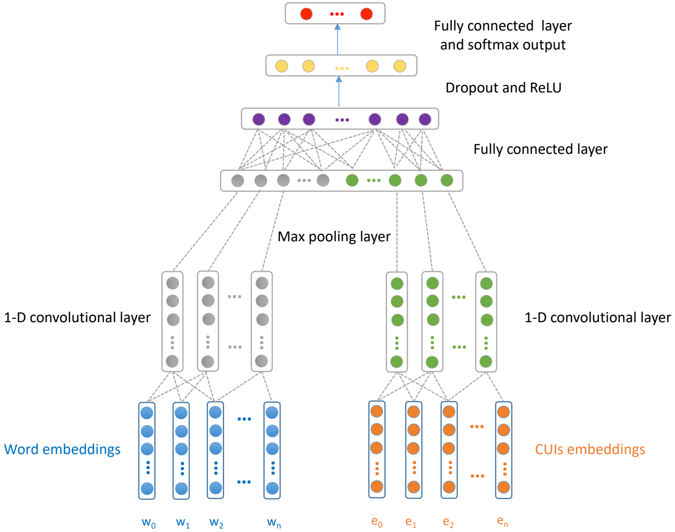
Knowledge-guided Convolutional Neural Networks

- After removing classes with very few examples, there are only two classes in the training sets:
  - Y and U for textual task.
  - Y and N for intuitive task.

Knowledge-guided Convolutional Neural Networks

- After removing classes with very few examples, there are only two classes in the training sets:
  - Y and U for textual task.
  - Y and N for intuitive task.
- We train a Convolutional Neural Network (CNN) on positive trigger phrases and UMLS CUIs of training records and predict the two remaining classes.
- CNN is a powerful deep learning model for text classification, and it performs better than recurrent neural networks in our preliminary experiment.

Knowledge-guided Convolutional Neural Networks



Knowledge-guided Convolutional Neural Networks

Words

- Positive trigger phrases with MIMIC-III word embeddings.
- We tried word embeddings with 100, 200, 300, 400, 500 and 600 dimensions.
- 200 dimensional word embeddings performs the best.

Knowledge-guided Convolutional Neural Networks

Words

- Positive trigger phrases with MIMIC-III word embeddings.
- We tried word embeddings with 100, 200, 300, 400, 500 and 600 dimensions.
- 200 dimensional word embeddings performs the best.

Entities

- We use MetaMap to link the full clinical text to CUIs in UMLS.
- We choose 13 types of CUIs which are closely related to diseases as the input entities of CNN.
- We use pre-trained CUIs embeddings made by De Vine et al.
- `https://github.com/clinicalml/embeddings`

CUIs Types

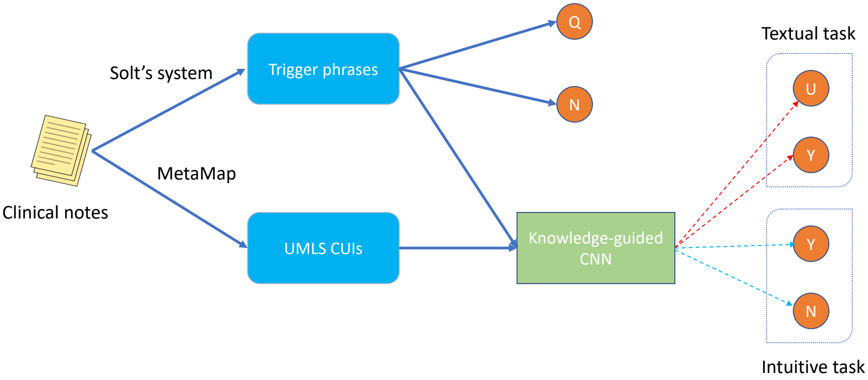
Table: The types of CUIs we used.

TUI	Semantic type description
T023	Body Part, Organ, or Organ Component
T033	Finding
T034	Laboratory or Test Result
T047	Disease or Syndrome
T048	Mental or Behavioral Dysfunction
T049	Cell or Molecular Dysfunctions
T059	Laboratory Procedure
T060	Diagnostic Procedure
T061	Therapeutic or Preventive Procedure
T121	Pharmacologic Substance
T122	Biomedical or Dental Material
T123	Biologically Active Substance
T184	Sign or Symptom

CNN Details

- Framework: TensorFlow.
- Parameters:
  - The number of convolution filters: 256
  - The convolution kernel size: 5
  - The dimension of hidden layer in the fully connected layer: 128
  - Dropout keep probability: 0.8
  - Learning rate: 0.001
  - Batch size: 64
  - The number of learning epochs: 30
- We also tried other parameters but do not find much difference.
- Softmax cross entropy loss as the loss function.
- Adam algorithm as the optimizer.

Prediction Process of Our Method



Results

TABLE III: Macro F_1 scores and Micro F_1 scores of Solt's system [5] (paper) and our method with word and entity embeddings. Scores in bold font means they are higher than the corresponding scores of the paper and Perl implementation.

Disease	Solt's paper [5]				Our method with word & entity embeddings			
	Textual		Intuitive		Textual		Intuitive	
	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1
Asthma	0.9434	0.9921	0.9784	0.9894	0.9434	0.9921	0.9784	0.9894
CAD	0.8561	0.9256	0.6122	0.9192	0.8551	0.9235	**0.6233**	**0.9345**
CHF	0.7939	0.9355	0.6236	0.9315	0.7939	0.9355	0.6236	0.9315
Depression	0.9716	0.9842	0.9346	0.9539	0.9716	0.9842	**0.9602**	**0.9727**
DM	0.9032	0.9761	0.9682	0.9729	0.9056	0.9801	0.9731	0.9770
Gallstones	0.8141	0.9822	0.9729	0.9857	0.8141	0.9822	0.9689	0.9837
GERD	0.4880	0.9881	0.5768	0.9131	0.4880	0.9881	0.5768	0.9131
Gout	0.9733	0.9881	0.9771	0.9900	0.9733	0.9881	0.9771	0.9900
Hypercholesterolemia	0.7922	0.9721	0.9053	0.9072	0.7922	0.9721	**0.9113**	0.9118
Hypertension	0.8378	0.9621	0.8851	0.9283	0.8378	0.9621	**0.9240**	**0.9484**
Hypertriglyceridemia	0.9732	0.9980	0.7981	0.9712	0.9434	0.9961	0.7092	0.9630
OA	0.9594	0.9761	0.6286	0.9589	0.9626	0.9781	0.6307	0.9610
Obesity	0.4879	0.9675	0.9724	0.9732	0.4885	0.9696	0.9747	0.9754
OSA	0.8781	0.9920	0.8805	0.9939	0.8781	0.9920	0.8805	0.9939
PVD	0.9682	0.9862	0.6348	0.9763	0.9682	0.9862	0.6314	0.9742
Venous insufficiency	0.8403	0.9822	0.8083	0.9625	**0.8816**	**0.9882**	0.8083	0.9625
Overall	0.8000	0.9756	0.6745	0.9590	**0.8016**	**0.9763**	**0.6768**	**0.9624**

Results

TABLE IV: Macro F_1 scores and Micro F_1 scores of Solt's system [5] (code) and our method with word embeddings only. Scores in bold font means they are higher than the corresponding scores of the paper and Perl implementation.

Disease	Solt's code				Our method with word embeddings only			
	Textual		Intuitive		Textual		Intuitive	
	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1	Macro F_1	Micro F_1
Asthma	0.9434	0.9921	0.9784	0.9894	0.9434	0.9921	0.9784	0.9894
CAD	0.8551	0.9235	0.6122	0.9192	0.8551	0.9235	0.6122	0.9192
CHF	0.7939	0.9355	0.6236	0.9315	0.7939	0.9355	0.6236	0.9315
Depression	0.9716	0.9842	0.9346	0.9539	0.9716	0.9842	**0.9602**	**0.9767**
DM	0.9056	0.9801	0.9731	0.9770	0.9056	0.9801	0.9731	0.9770
Gallstones	0.8141	0.9822	0.9729	0.9857	0.8141	0.9822	0.9729	0.9857
GERD	0.4880	0.9881	0.5768	0.9131	0.4880	0.9881	0.5768	0.9131
Gout	0.9733	0.9881	0.9771	0.9900	0.9733	0.9881	0.9771	0.9900
Hypercholesterolemia	0.7922	0.9721	0.9101	0.9118	0.7922	0.9721	0.9042	0.9049
Hypertension	0.8378	0.9621	0.8861	0.9283	0.8378	0.9621	**0.9240**	**0.9484**
Hypertriglyceridemia	0.9732	0.9980	0.7092	0.9630	0.9732	0.9980	0.7092	0.9630
OA	0.9626	0.9781	0.6307	0.9610	0.9626	0.9781	0.6307	0.9610
Obesity	0.4885	0.9696	0.9747	0.9754	0.4885	0.9696	0.9747	0.9754
OSA	0.8781	0.9920	0.8805	0.9939	0.8781	0.9920	0.8805	0.9939
PVD	0.9682	0.9862	0.6314	0.9742	0.9682	0.9862	0.6314	0.9742
Venous insufficiency	0.8403	0.9822	0.8083	0.9625	0.8403	0.9822	0.8083	0.9625
Overall	0.8014	0.9760	0.6745	0.9592	0.8014	0.9760	**0.6760**	**0.9612**

Results

- We released our implementation.
- `https://github.com/yao8839836/obesity`

V. N. Garla and C. Brandt, "Knowledge-based biomedical word sense disambiguation: an evaluation and application to clinical document classification," *Journal of the American Medical Informatics Association*, vol. 20, no. 5, pp. 882–886, 2013.

——, "Ontology-guided feature engineering for clinical text classification," *Journal of biomedical informatics*, vol. 45, no. 5, pp. 992–998, 2012.

D. D. A. Bui and Q. Zeng-Treitler, "Learning regular expressions for clinical text classification," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 850–857, 2014.

V. Garla, C. Taylor, and C. Brandt, "Semi-supervised clinical text classification with laplacian svms: an application to cancer case management," *Journal of biomedical informatics*, vol. 46, no. 5, pp. 869–875, 2013.

R. L. Figueroa, Q. Zeng-Treitler, L. H. Ngo, S. Goryachev, and E. P. Wiechmann, "Active learning for clinical text classification: is it better than random sampling?" *Journal of the American Medical Informatics Association*, vol. 19, no. 5, pp. 809–816, 2012.

S. Gehrmann, F. Dernoncourt, Y. Li, E. T. Carlson, J. T. Wu, J. Welt, J. Foote Jr, E. T. Moseley, D. W. Grant, P. D. Tyler *et al.*, "Comparing rule-based and deep learning models for patient phenotyping," *arXiv preprint arXiv:1703.08705*, 2017.

Y. Luo, "Recurrent neural networks for classifying relations in clinical notes," *Journal of biomedical informatics*, vol. 72, pp. 85–95, 2017.

Y. Luo, Y. Cheng, Ö. Uzuner, P. Szolovits, and J. Starren, "Segment convolutional neural networks (seg-cnns) for classifying relations in clinical notes," *Journal of the American Medical Informatics Association*, vol. 25, no. 1, pp. 93–98, 2017.

Z. Che, D. Kale, W. Li, M. T. Bahadori, and Y. Liu, "Deep computational phenotyping," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015, pp. 507–516.

J. Geraci, P. Wilansky, V. de Luca, A. Roy, J. L. Kennedy, and J. Strauss, "Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression," *Evidence-based mental health*, vol. 20, no. 3, pp. 83–87, 2017.

Ö. Uzuner, "Recognizing obesity and comorbidities in sparse data," *Journal of the American Medical Informatics Association*, vol. 16, no. 4, pp. 561–570, 2009.

I. Solt, D. Tikk, V. Gál, and Z. T. Kardkovács, "Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier," *Journal of the American Medical Informatics Association*, vol. 16, no. 4, pp. 580–584, 2009.

L. De Vine, G. Zuccon, B. Koopman, L. Sitbon, and P. Bruza, "Medical semantic similarity with a neural language model," in *Proceedings of the 23rd ACM* *international conference on conference on information and knowledge management*. ACM, 2014, pp. 1819–1822.

Thank You!