Clinical data successes using machine learning

A survey by Joseph Paul Cohen, PhD
Montreal Institute for Learning Algorithms

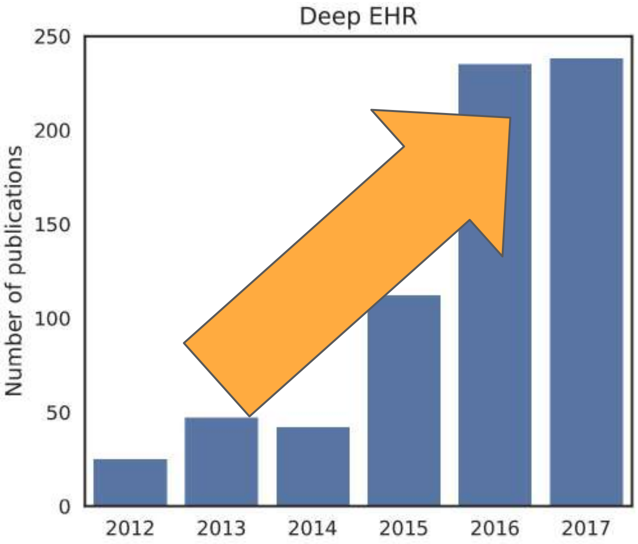
+Tutorial

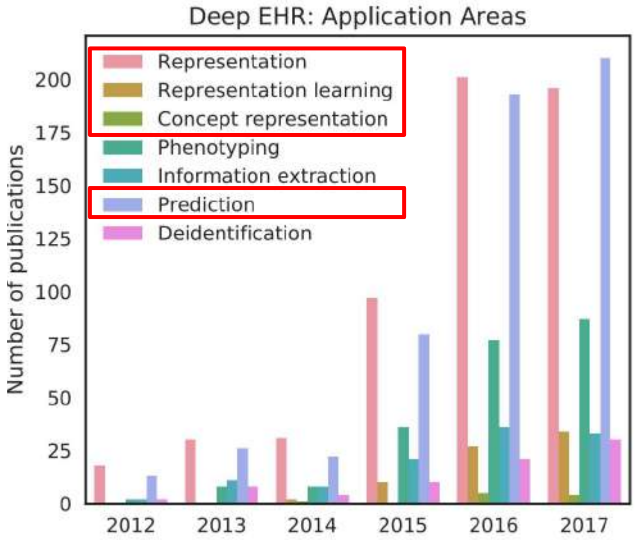
Topics:

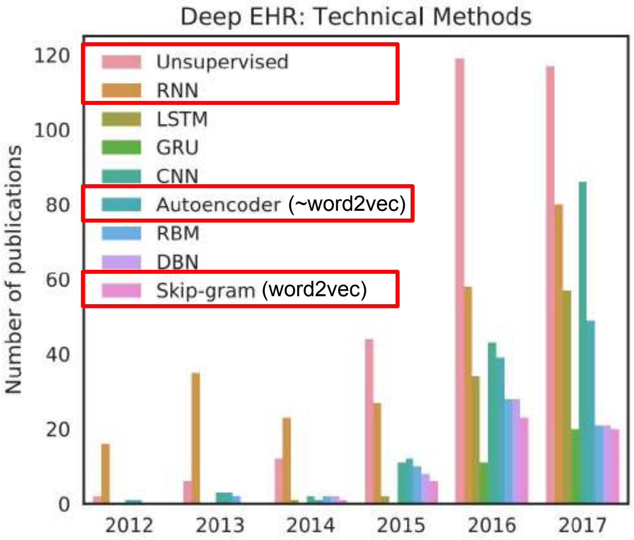
1. Medical Concept Representation
2. Clinical Event Prediction

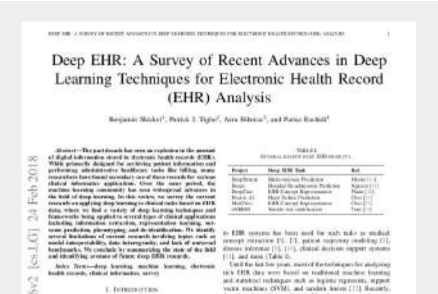
Mila Medical

Where is the Deep Learning research?








Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis

Topic of today!

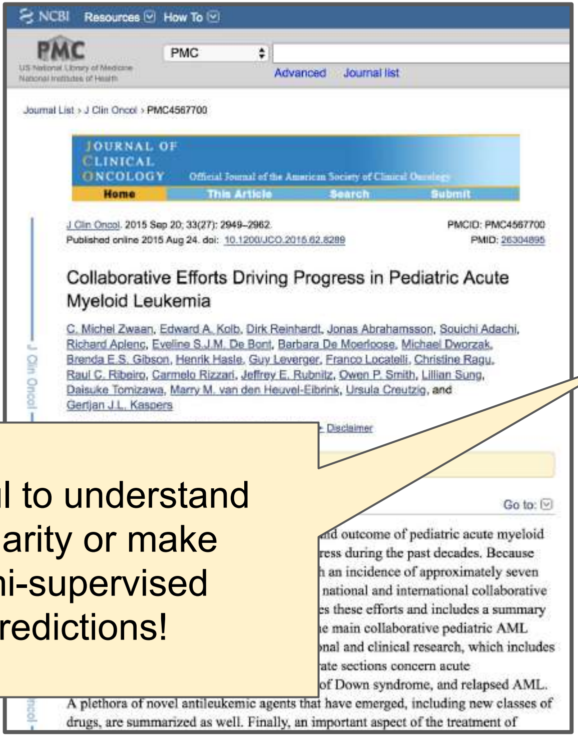
[Shickel, Deep EHR : A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record, 2018]

Concept Representation

Clinical Note

Mr. Smith is a 63-year-old gentleman with coronary artery disease, hypertension, hypercholesterolemia, COPD and tobacco abuse. He reports doing well. He did have some more knee pain for a few weeks, but this has resolved. He is having more trouble with his sinuses. I had started him on Flonase back in December. He says this has not really helped. Over the past couple weeks he has had significant congestion and thick discharge. No fevers or headaches but does hav diffuse upper right-sided teeth **He denies any chest pains, palpitations, PND, orthopnea, ed syncope.** His breathing is doing No cough. He continues to smoke half-a-pack per day. He plans or **the patches again.**

Clinical Publication


NCBI Resources How To
PMC
US National Library of Medicine
National Institutes of Health
Advanced Journal list
Journal List › J Clin Oncol › PMC4567700
JOURNAL OF CLINICAL ONCOLOGY
Home This Article Search Submit
J Clin Oncol. 2015 Sep 20; 33(27): 2949–2962. PMCID: PMC4567700
Published online 2015 Aug 24. doi: 10.1200/JCO.2015.62.8289 PMID: 26304895

Collaborative Efforts Driving Progress in Pediatric Acute Myeloid Leukemia

C. Michel Zwaan, Edward A. Kolb, Dirk Reinhardt, Jonas Abrahamsson, Souichi Adachi, Richard Aplenc, Eveline S.J.M. De Bont, Barbara De Moerloose, Michael Dworzak, Brenda E.S. Gibson, Henrik Hasle, Guy Leverger, Franco Locatelli, Christine Ragu, Raul C. Ribeiro, Carmelo Rizzari, Jeffrey E. Rubnitz, Owen P. Smith, Lillian Sung, Daisuke Tomizawa, Marry M. van den Heuvel-Eibrink, Ursula Creutzig, and Gertjan J.L. Kaspers

Disclaimer

Go to:

...nd outcome of pediatric acute myeloid ...ress during the past decades. Because ...h an incidence of approximately seven ...national and international collaborative ...s these efforts and includes a summary ...e main collaborative pediatric AML ...onal and clinical research, which includes ...ate sections concern acute ...of Down syndrome, and relapsed AML. A plethora of novel antileukemic agents that have emerged, including new classes of drugs, are summarized as well. Finally, an important aspect of the treatment of

Helpful to understand similarity or make semi-supervised predictions!

Representations for:

Patient
Doctor
Visit
Disease
Drug
Symptom

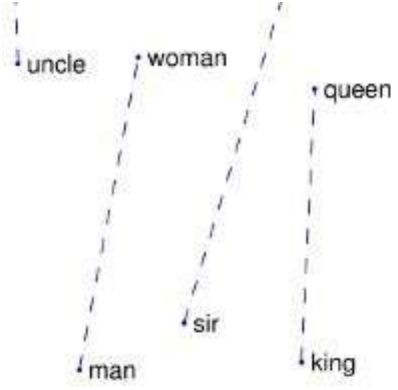
Word Embeddings for Biomedical Language

Extract relationships between words and produce a latent representation

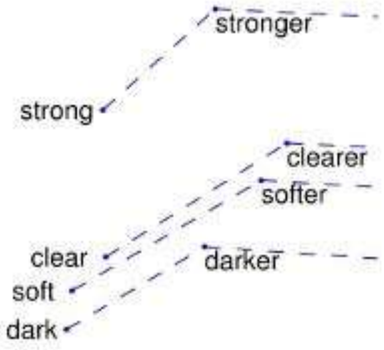
What to do with word embeddings?

- We can compose them to create paragraph embeddings (bag of embeddings).
- Use in place of words for an RNNs (More on this later!)
- Augment learned representations on small datasets

Study the compositionality of the learned latent space

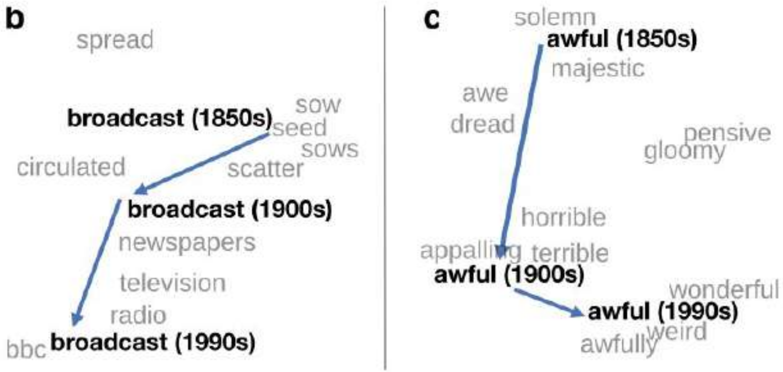


[Mikolov, 2013]



[Pennington, 2014]

Study how the meaning between two texts varies (or hospitals, or doctors)?



[Cultural Shift or Linguistic Drift, Hamilton, 2016]

Token representations

One-hot encoding: binary vector per token

Example:

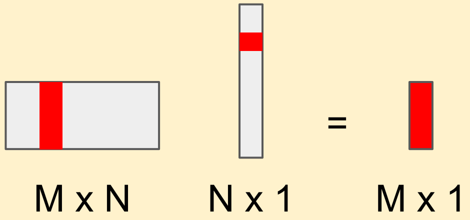
cat = [0 0 1 0 0 0 0 0 0 0 0 0 0 0 … 0]

dog = [0 0 0 0 1 0 0 0 0 0 0 0 0 0 … 0]

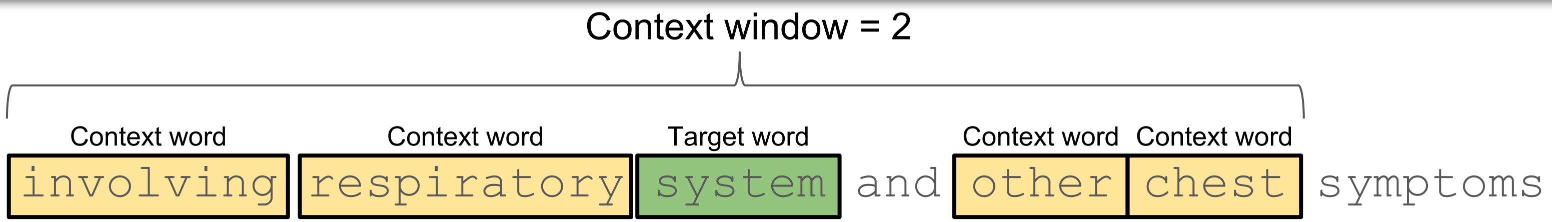
house = [1 0 0 0 0 0 0 0 0 0 0 0 0 0 … 0]

Note!

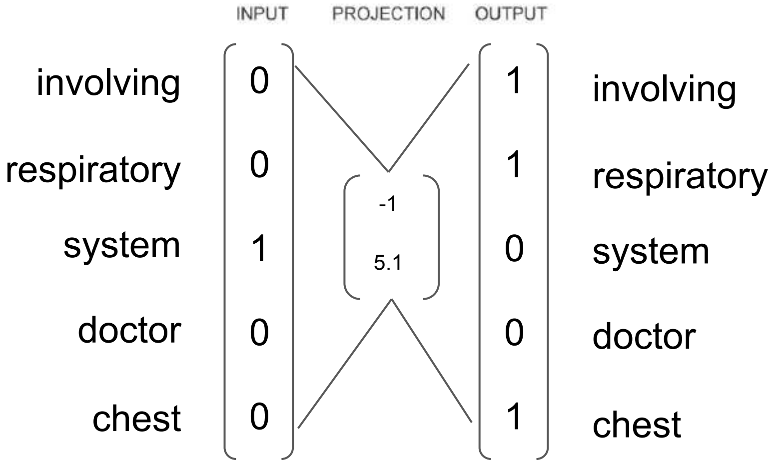
If x is one hot
The dot product of Wx
= a Single column of W

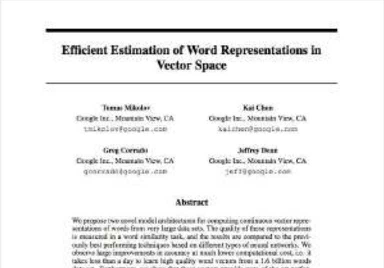


word2vec



1. Each word is a training example
2. Each word is used in many contexts
3. The context defines each word





Efficient Estimation of Word Representations in Vector Space

Tomas Mikolov
Google Inc., Mountain View, CA
tmikolov@google.com

Kai Chen
Google Inc., Mountain View, CA
kaichen@google.com

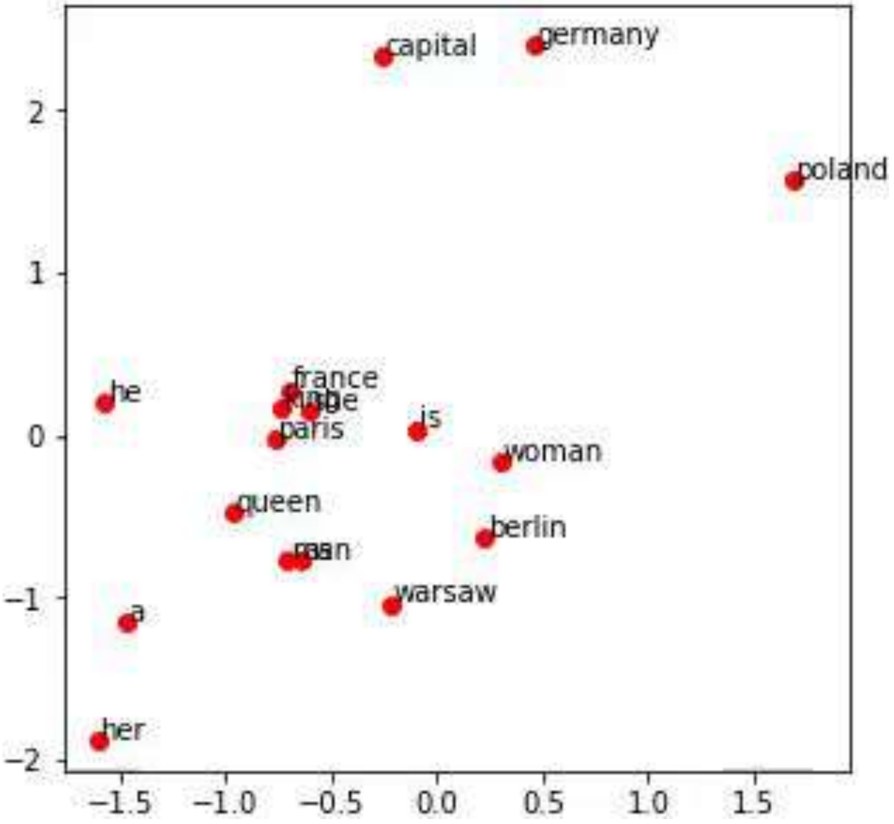
Greg Corrado
Google Inc., Mountain View, CA
gcorrado@google.com

Jeffrey Dean
Google Inc., Mountain View, CA
jeff@google.com

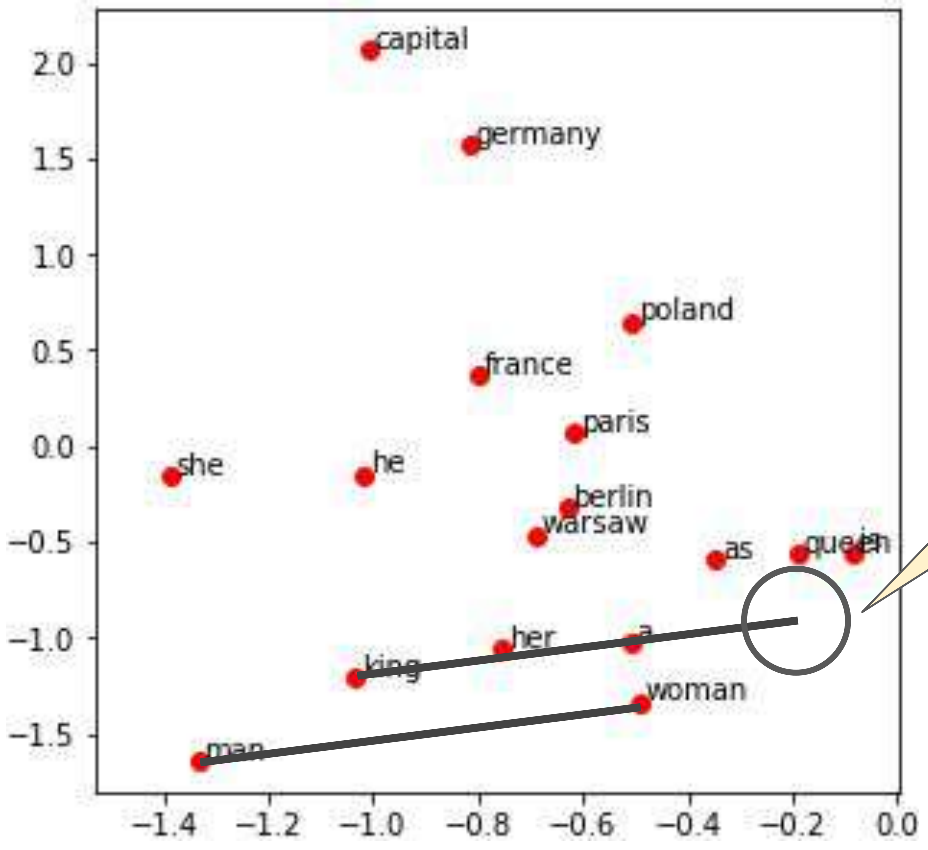
Abstract

We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e. it takes less than a day to learn high quality word vectors from a 1.6 billion words

Mikolov, Efficient Estimation of Word Representations in Vector Space, 2013



Learning in progress



The point that is closest is queen!

king + (woman - man) = ?

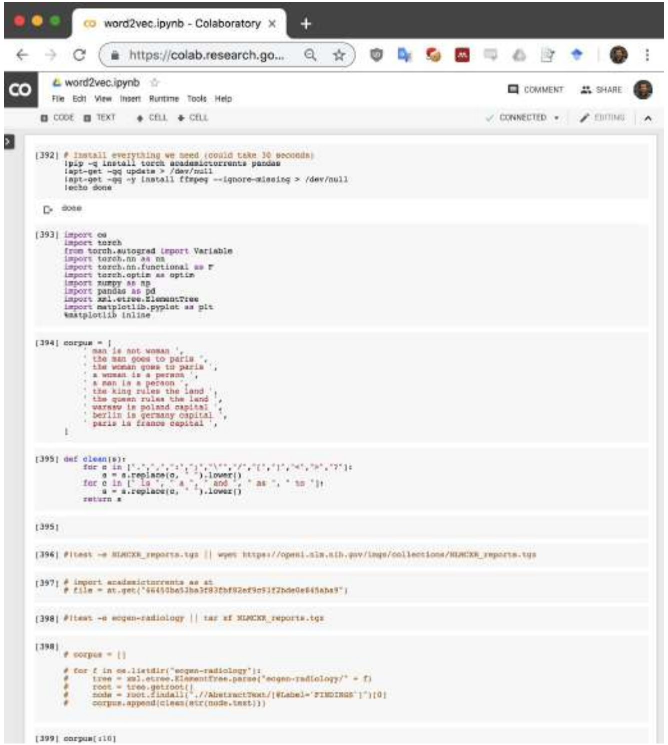
Follow along online!

https://colab.research.google.com/drive/1g4zvEg921sLQK-VsBk5mMb2-h4goCGyd

python™

jupyter

colab



```python
window_size = 2
idx_pairs = []
for sentence in tokenized_corpus:
  indices = [word2idx[word] for word in sentence]

    for center_word_pos in range(len(indices)):

      for w in range(-window_size, window_size + 1):
        context_word_pos = center_word_pos + w

        if context_word_pos < 0 or
           context_word_pos >= len(indices) or
           center_word_pos == context_word_pos:
           Continue

        context_word_idx = indices[context_word_pos]
        idx_pairs.append((indices[center_word_pos], context_word_idx))
```

```
sentence = ['paris', 'france', 'capital']
indices = [2, 5, 11]
```

for each word, treated as center word

`center_word_pos = 0, 1, 2, ...`

for each window position

`w = -2, -1, 0, 1, 2`

make sure not to jump out of the sentence

```
Output:
[[16, 12],
 [16,  1],
 ...
 [ 1, 16],
 [ 1, 12]]
```

Code by github user: mbednarski

```python
class SkipGram(nn.Module):
    def __init__(self, vocab_size, embd_size):
        super(SkipGram, self).__init__()

        self.W1 = Variable(torch.randn(embd_size, vocab_size).float(), requires_grad=True)
        self.W2 = Variable(torch.randn(vocab_size, embd_size).float(), requires_grad=True)

    def forward(self, focus):

        z1 = torch.matmul(self.W1, focus)

        z2 = torch.matmul(self.W2, z1)

        softmax = F.log_softmax(z2, dim=0)

        return softmax

model = SkipGram(vocabulary_size, 2)
```

Initialize two matrices
|E| x |V|
|V| x |E|

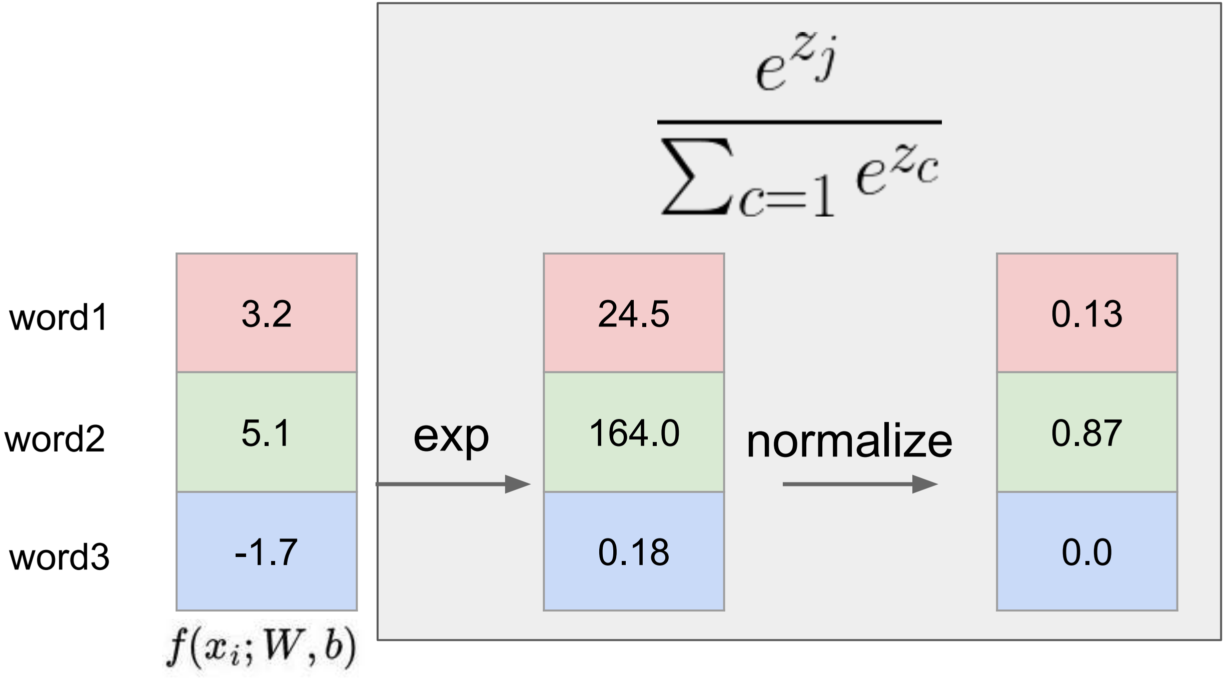
Encoder:
|E| x |V| dot |V| x 1 = |E| x 1

Decoder:
|V| x |E| dot |E| x 1 = |V| x 1

Softmax over context prediction

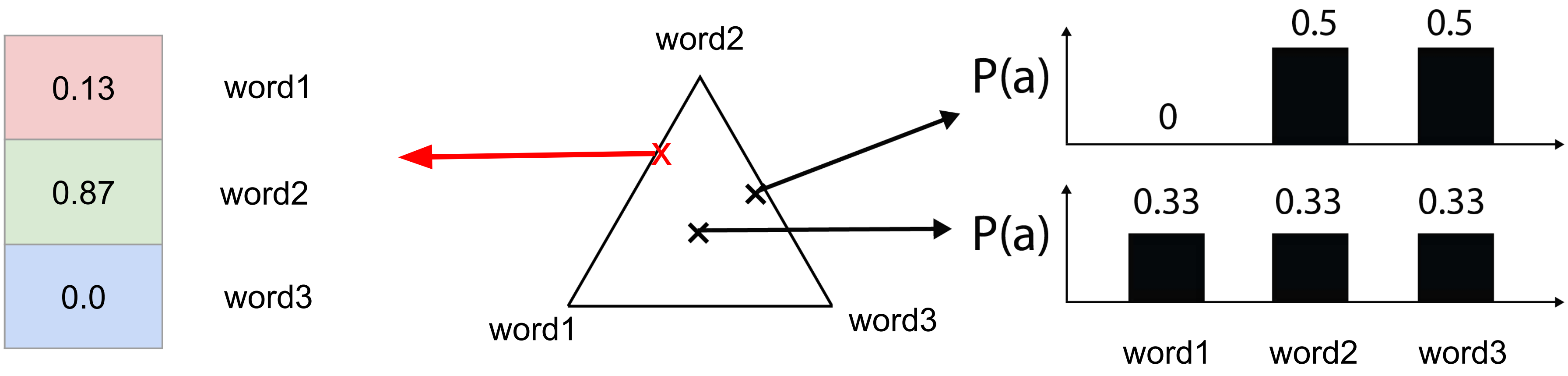
Code by github user: mbednarski

A note on the softmax function



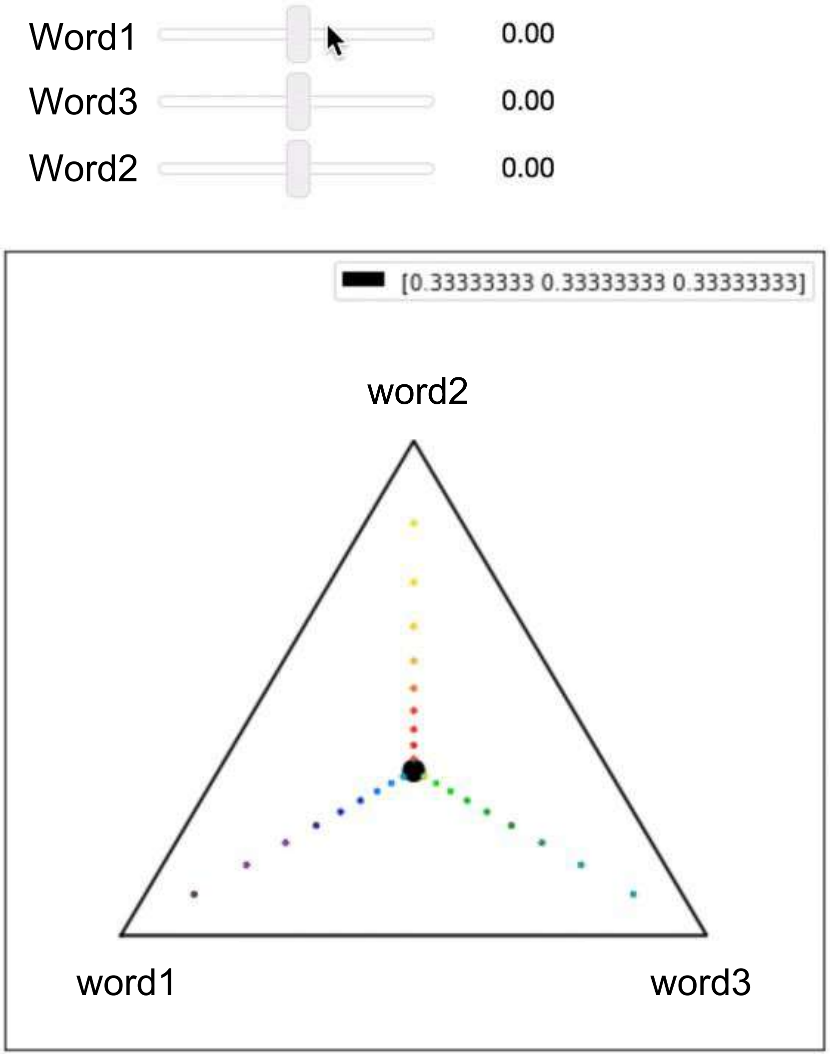
To predict multiple classes we project to a probability distribution

Simplex



Because it is on a simplex; the correction of one term impacts all

Image credits: http://gureckislab.org/



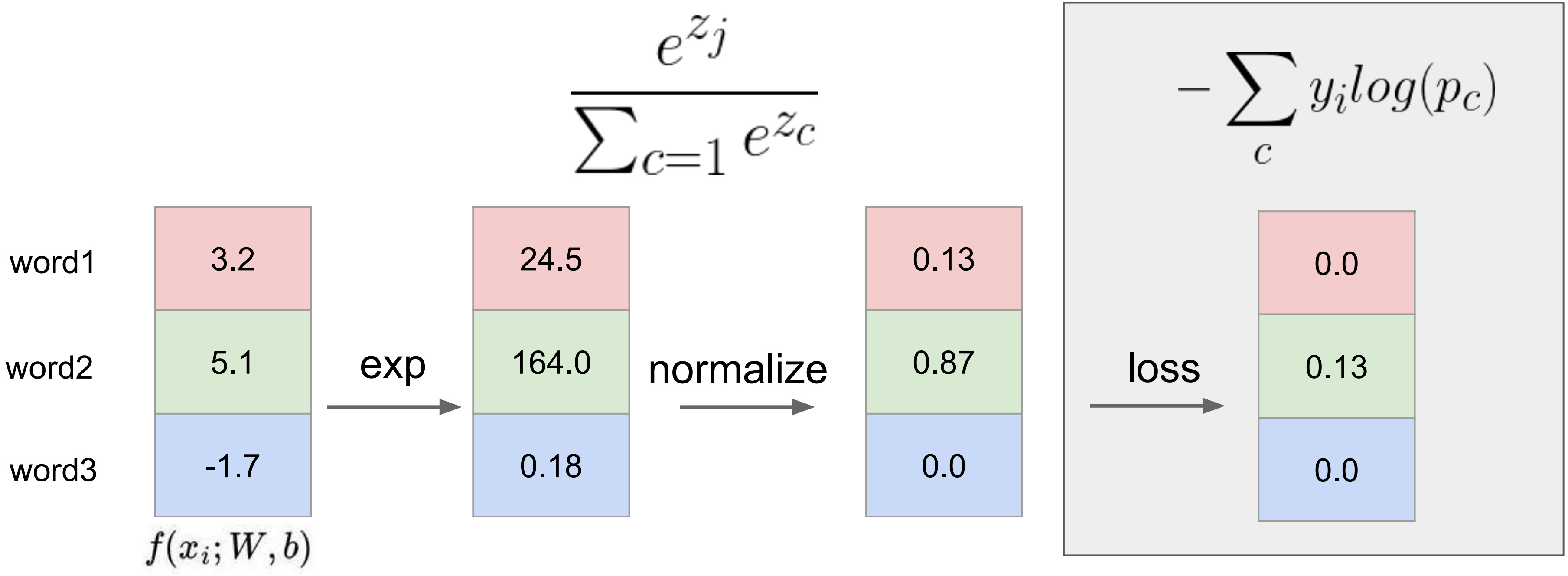
Infinite ways to generate the same output.

A correction of one sends gradients to others

We can learn unseen classes through a process of elimination.

https://github.com/ieee8023/NeuralNetwork-Examples/blob/master/general/simplex-softmax.ipynb

Softmax and Cross-entropy loss



To predict multiple class we can project the output onto a simplex and compute the loss there.

Open Access Subset of PubMedCentral

Subset of files that are available open access

- Papers available in PDF or XML
- 1.25 million biomedical articles and 2 million distinct words
- Available over FTP for bulk download (CompSci Friendly!)
- Metadata includes journal name and year


```xml
    </sec>
    <sec sec-type="methods">
      <title>Patients and methods</title>
      <sec>
        <title>Patients</title>
        <p>Between September 1995 and September 1998, 413 patients with abnormal breast findings were referred fo
histological evaluation to the Department of Gynecology of the Friedrich-Schiller University, Jena, Germany. Patients
had been selected and referred because of the presence of breast lesions detected by palpation and/or mammography
and/or sonography. In addition, MR mammography was performed in all patients. We excluded five patients with invasive
cancer who had a history of core-needle or fine-needle biopsy cancer within 2 weeks before referral, because the
presence of haematoma may mimic false-positive findings on MR mammography. In addition, five patients who did not kee
still during MR mammography were excluded.</p>
      </sec>
      <sec>
        <title>Imaging</title>
        <p>
          Analysis of the sonograms taken in patients with histologically confirmed carcinoma
          <italic>in situ</italic>
          were excluded from analysis because the value of sonography for detection of premalignant disease is
```


Example XML data

```
Breast Cancer Res.
Genome Biol.
Arthritis Res.
BMC Cell Biol.
…
```

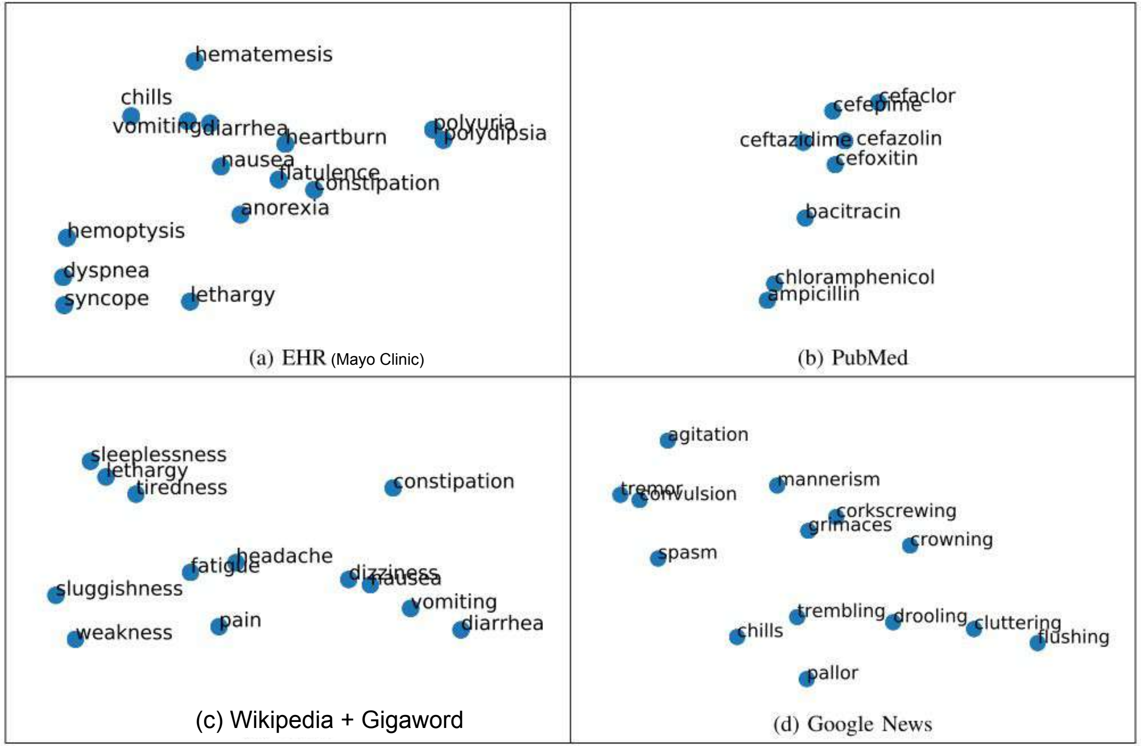
Journal names

https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/

Word Embeddings for Biomedical Language

Representations are biased by the data.

We can use this to our advantage to control the domain.



(a) EHR (Mayo Clinic)

(b) PubMed

(c) Wikipedia + Gigaword

(d) Google News

Fig. 1: Example clusters in different word embeddings.

A Comparison of Word Embeddings for the Biomedical Natural Language Processing

Department of Health Sciences Research

Mayo Clinic

Rochester, USA

Abstract

[cs.IR] 9 Feb 2018

Wang, A Comparison of Word Embeddings for the Biomedical Natural Language Processing, 2018

Example medical words and their five post similar words based on each training

	EHR (Mayo Clinic)	PubMed	Wikipedia + Gigaword	Google News
diabetes	mellitus,	cardiovascular,	hypertension,	diabetics,
	ntrolled,	nonalcoholic,	obesity,	hypertension,
	esterolemia,	obesity,	arthritis,	diabetic,
	pidemia,	mellitus,	cancer,	diabetes_mellitus,
	tis	polycystic	alzheimer	heart_disease
peptic ulcer disease	scleroderma,	gastritis,	ulcers,	ichen_planus,
	duodenal,	alcoholism,	arthritis,	Candida_infection,
	crohn,	rheumatic,	diseases,	vaginal_yeast_infections,
	gastroduodenal,	ischaemic,	diabetes,	oral_thrush,
	diverticular	nephropathy	stomach	dermopathy
colon cancer	breast,	breast,	breast,	breast,
	ovarian,	mcf,	prostate,	prostate,
	prostate,	cancers,	cancers,	tumor,
	postmenopausally,	tumor_suppressing,	tumor,	pre_cancerous_lesion,
	caner	downregulation	liver	cancerous_polyp

The full name of diabetes is "diabetes mellitus"

A type of peptic ulcer

Colon cancer is associated with breast cancer!

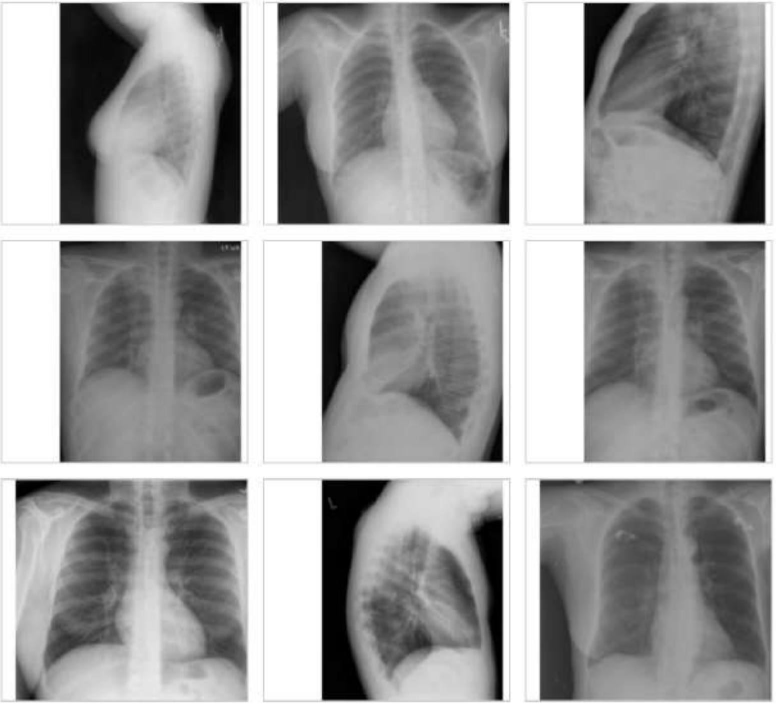
Articles say diabetics are at increased risk of hypertension (high blood pressure)

One symptom is an ulcer

Specific General

Indiana University Hospital Reports

Chest X-ray images from the Indiana University hospital network



Indiana University Chest X-ray Collection

Kohli MD, Rosenman M - (2013)

Affiliation: Indiana University

ABSTRACT

Comparison: None.

Indication: Positive TB test

Findings: The cardiac silhouette and mediastinum size are within normal limits. There is no pulmonary edema. There is no focal consolidation. There are no XXXX of a pleural effusion. There is no evidence of pneumothorax.

Impression: Normal chest x-XXXX.

NOTE: The data are drawn from multiple hospital systems.

Show MeSH

Related in: MedlinePlus Request Collection

1000 reports available in XML format!

Input:

```
<Abstract>
   <AbstractText Label="COMPARISON">None.</AbstractText>
   <AbstractText Label="INDICATION">Positive TB test</AbstractText>
   <AbstractText Label="FINDINGS">The cardiac silhouette and mediastinum size
are within normal limits. There is no pulmonary edema. There is no focal
consolidation. There are no XXXX of a pleural effusion. There is no evidence of
pneumothorax.</AbstractText>
   <AbstractText Label="IMPRESSION">Normal chest x-XXXX.</AbstractText>
</Abstract>
```

Python code to parse the XML

```
def clean(s):
    for c in [".",",",":",";","\"","/","[","]","<",">","?"]:
        s = s.replace(c, " ").lower()
    return s

corpus = []

for f in os.listdir("ecgen-radiology"):
    tree = xml.etree.ElementTree.parse("ecgen-radiology/" + f)
    root = tree.getroot()
    node = root.findall(".//AbstractText/[@Label='FINDINGS']")[0]
    corpus.append(clean(str(text)))
```

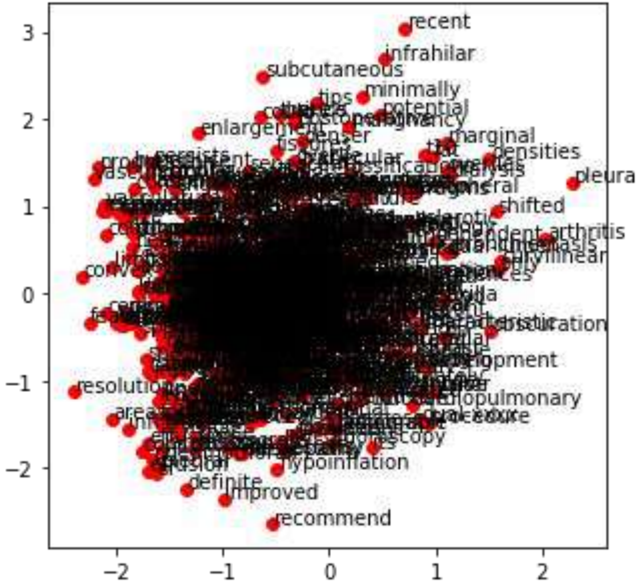
Output!:

```
['heart size normal lungs ...',
 'the heart size and ...',
 'the heart is normal in ...',
 'the lungs are clear the ...',
 'heart size normal lungs ...',
 'heart is mildly heart ...',
 'the lungs are clear there ...',
 'cardiac and mediastinal ...',
…
```

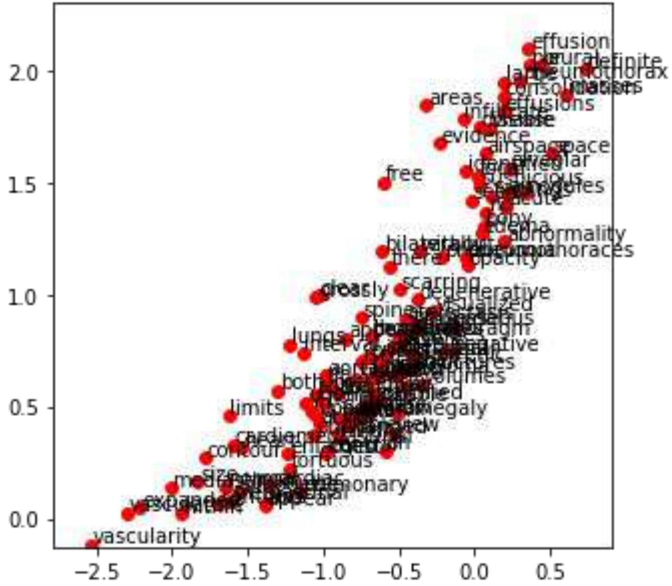
Hyperparameters!

Depending on the configuration of the model the embeddings can vary:

No information

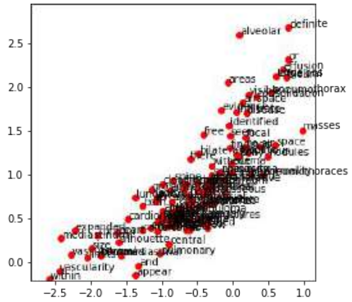


Compression



We can vary: the dimension of the embedding, learning rate, token words, window size, etc..

Size of embedding space matters

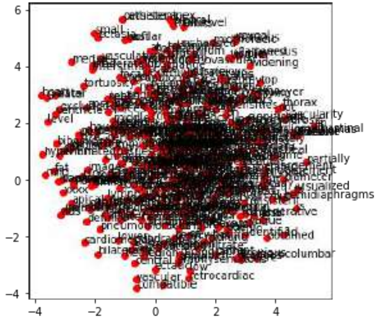


		Distance
0	spine	0.000000
1	seen	6.632637
2	evidence	8.230038
3	infiltrate	13.233962
4	findings	17.762381
5	disease	19.574038
6	focal	22.267534
7	aortic	23.346777
8	suspicious	25.558247
9	acute	29.176380
10	within	29.798399

Dim = 2

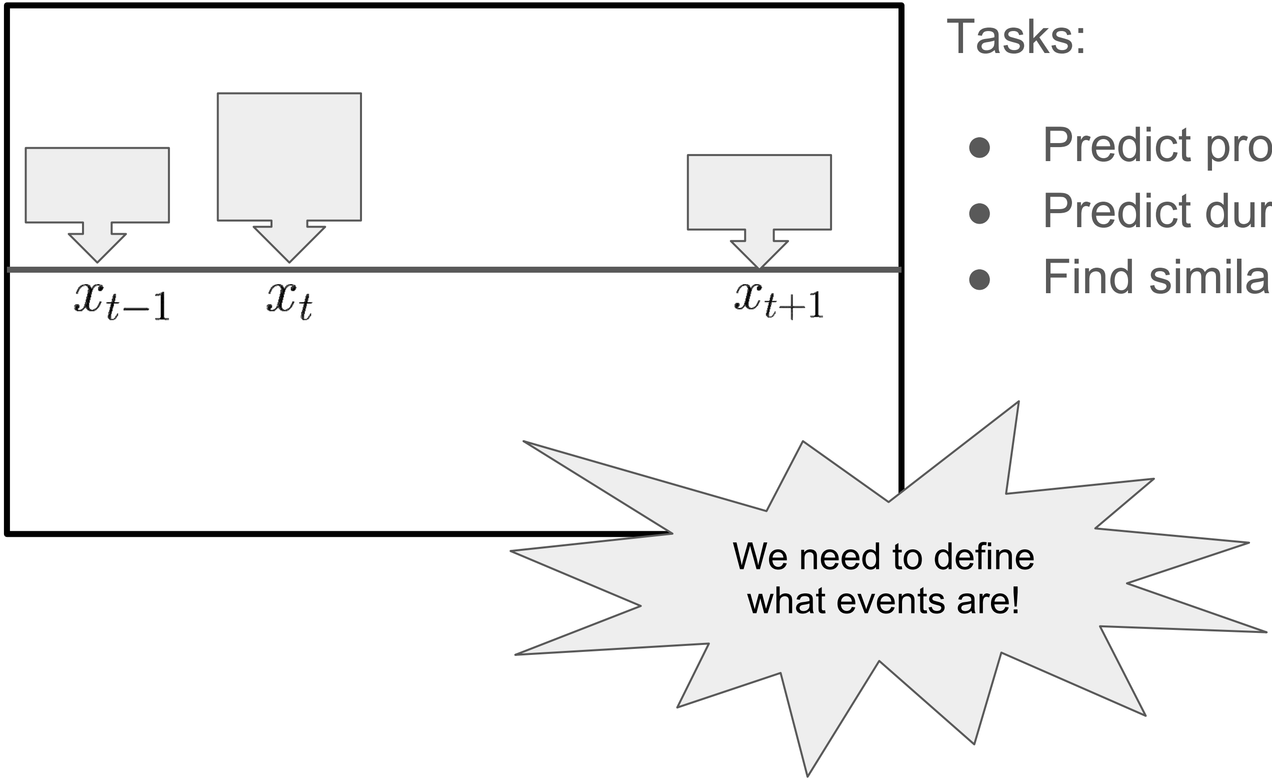
		Distance
0	spine	0.000000
1	severe	3.304547
2	changes	3.469765
3	anterior	3.474130
4	mild	3.731760
5	also	3.828588
6	thoracic	3.838329
7	noted	3.871284
8	calcifications	3.934170
9	vertebral	4.004414
10	stable	4.022851

Dim = 100



*via t-sne

Time series medical records



Tasks:

- Predict probability of event in future
- Predict duration until next visit
- Find similar patients/events/visits

ICD (International Classification of Diseases)

World Health Organization

ICD is the foundation for the identification of health trends and statistics globally, and the international standard for reporting diseases and health conditions. It is the diagnostic classification standard for all clinical and research purposes. ICD defines the universe of diseases, disorders, injuries and other related health conditions, listed in a comprehensive, hierarchical fashion (http://www.who.int/)

1893 - Causes of Death (International Statistical Institute)

1975 - ICD-9 - (WHO)

1990 - ICD-10 - (WHO)

2022 - ICD-11 - (WHO)

There are alternative standards but they can require a fee

ICD (International Classification of Diseases)

World Health Organization

Example ICD-9 codes:

786	Symptoms involving respiratory system and other chest symptoms
786.0	Dyspnea and respiratory abnormalities
786.1	Stridor
786.2	Cough
786.3	Hemoptysis
786.4	Abnormal sputum
786.5	Chest pain
786.6	Swelling, mass or lump in chest
786.7	Abnormal chest sounds
786.8	Hiccough
786.9	Other

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD9-CM/2011

ICD (International Classification of Diseases)

World Health Organization

Example ICD-9 codes (786.5)

Cardialgia (see also Pain, precordial) 786.51
Diaphragmalgia 786.52
chest 786.59
anginoid (see also Pain, precordial) 786.51
chest (central) 786.50
atypical 786.59
midsternal 786.51
musculoskeletal 786.59
noncardiac 786.59
substernal 786.51
wall (anterior) 786.52
costochondral 786.52
diaphragm 786.52
heart (see also Pain, precordial) 786.51
intercostal 786.59
over heart (see also Pain, precordial) 786.51
pericardial (see also Pain, precordial) 786.51

pleura, pleural, pleuritic 786.52
precordial (region) 786.51
respiration 786.52
retrosternal 786.51
rib 786.50
substernal 786.51
respiration 786.52
Pleuralgia 786.52
Pleurodynia 786.52
Precordial pain 786.51
chest 786.59
Prinzmetal-Massumi syndrome (anterior chest wall) 786.52
painful 786.52

Lots of grouping!

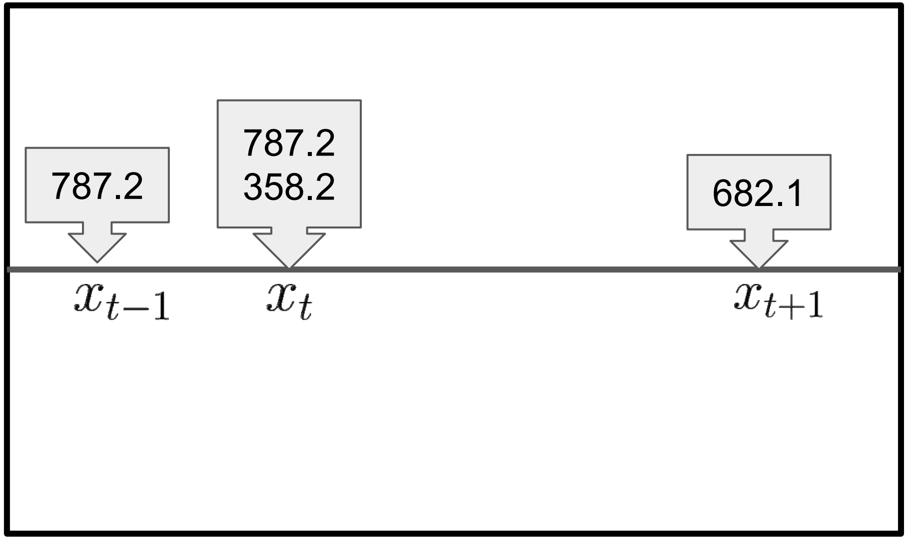
ICD-10 is very detailed

V97.33XD: **Sucked into jet engine**, subsequent encounter

V00.15: **Heelies Accident**
Applicable To Rolling shoe, Wheeled shoe, Wheelies accident.

Supertypes:
V00-Y99 External causes of morbidity
V95-V97 Air and space transport accidents
V00 Pedestrian conveyance accident
V00.1 Rolling-type pedestrian conveyance accident

Time series medical records

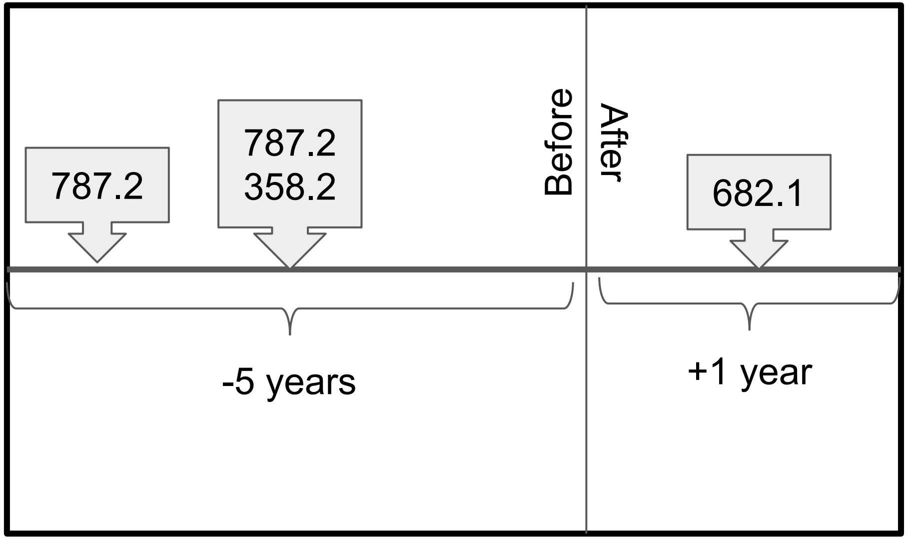


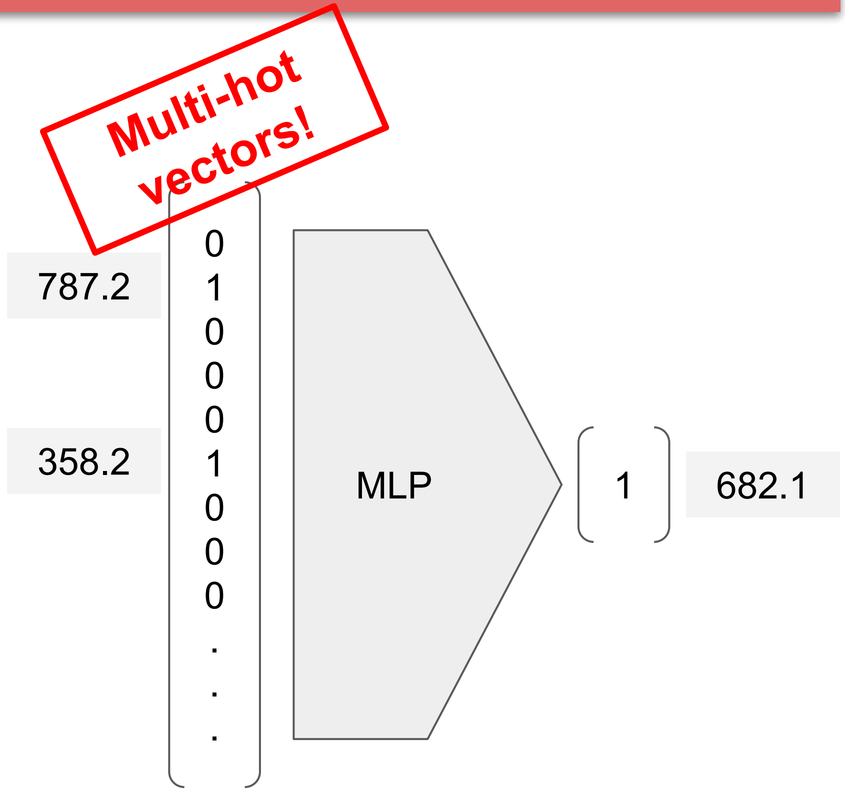
Sequence of medical codes over time

Tasks:

- Predict probability of code in future
- Predict duration until next visit
- Find similar patients/codes/visits

MLPs on time series data

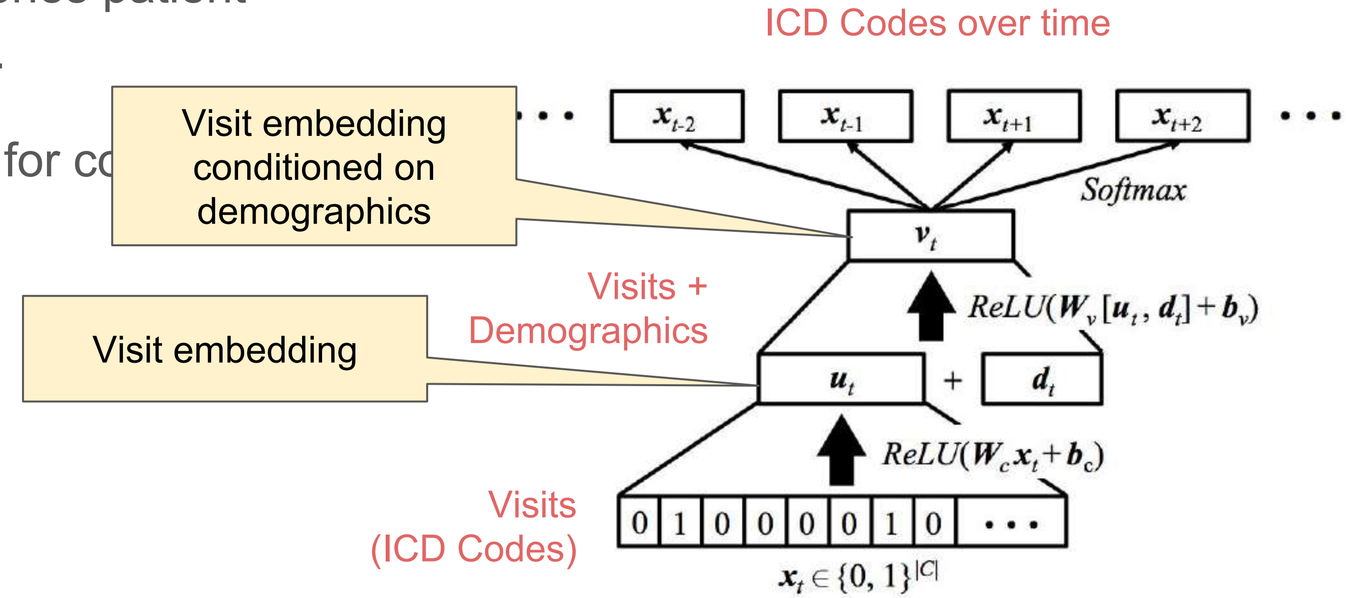


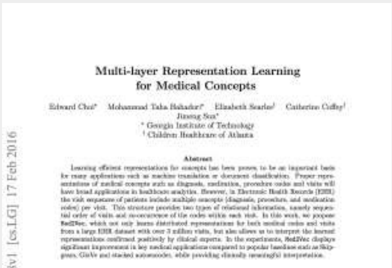


Med2Vec

Word2Vec for time series patient visits with ICD codes.

Embeddings learned for co[...]
and demographics.




Multi-layer Representation Learning
for Medical Concepts

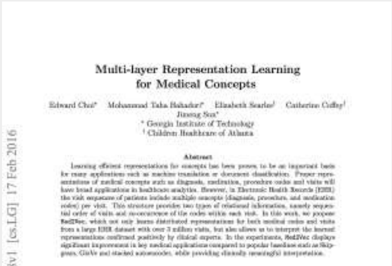
Choi, Multi-layer Representation Learning for Medical Concepts, 2016

Baseline methods

One-hot: In order to compare with the raw input data, the binary vector x_t for the visit is used.

Stacked autoencoder (SA): Using the binary vector x_t concatenated with patient demographic information as the input the SA is trained to minimize the reconstruction error. Then will be used to generate visit representations.

Sum of Skip-gram vectors (word2vec): First learn the code-level representations with Skip-gram only. Then for the visit-level representation, add the representations of the codes within the visit.


Multi-layer Representation Learning
for Medical Concepts

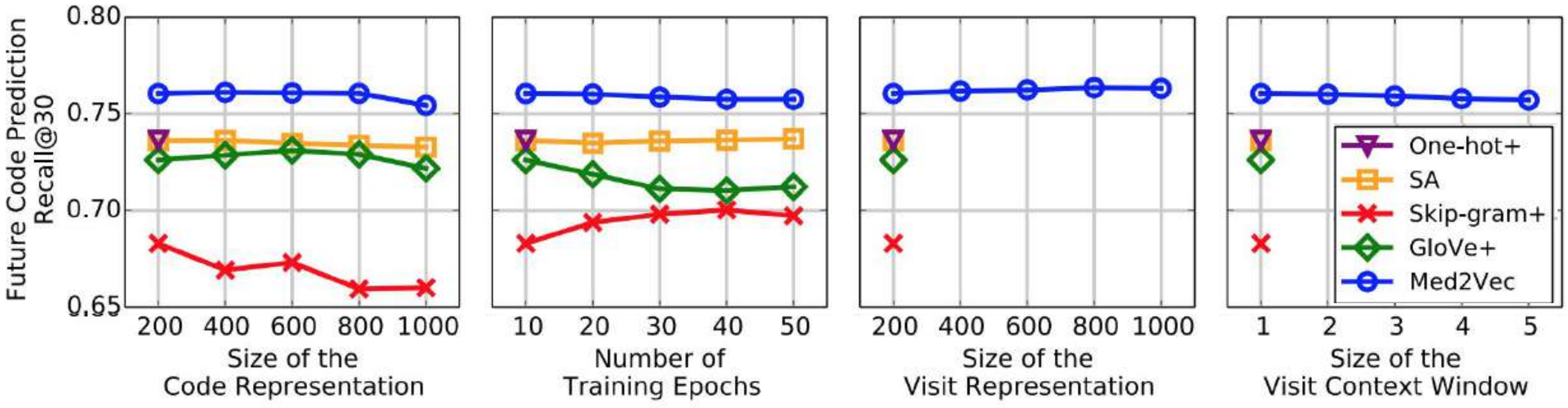
Edward Choi* Mohammad Taha Bahadori* Elizabeth Searles† Catherine Coffey†
Jimeng Sun*
* Georgia Institute of Technology
† Children Healthcare of Atlanta

Abstract

[cs.LG] 17 Feb 2016

Choi, Multi-layer Representation Learning for Medical Concepts, 2016

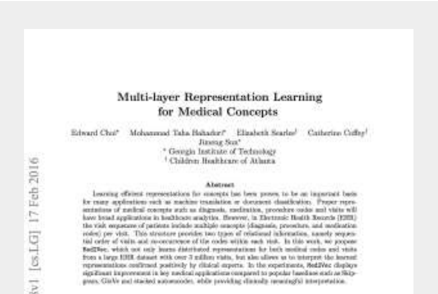
Med2Vec



Evaluation: Predicting codes of the next visit

Private):

Dataset	CHOA
# of patients	550,339
# of visits	3,359,240
Avg. # of visits per patient	6.1
# of unique medical codes	28,840
- # of unique diagnosis codes	10,414
- # of unique medication codes	12,892
- # of unique procedure codes	5,534
Avg. # of codes per visit	7.88
Max # of codes per visit	440
(95%, 99%) percentile # of codes per visit	(22, 53)


Multi-layer Representation Learning for Medical Concepts

Edward Choi* Mohammad Taha Bahadori* Elizabeth Searles† Catherine Coffey† Jimeng Sun*
* Georgia Institute of Technology
† Children Healthcare of Atlanta

Abstract

[cs.LG] 17 Feb 2016

$$\text{top-}k\text{ recall} = \frac{\#\text{ of true positives in the top }k\text{ predictions}}{\#\text{ of true positives}}$$

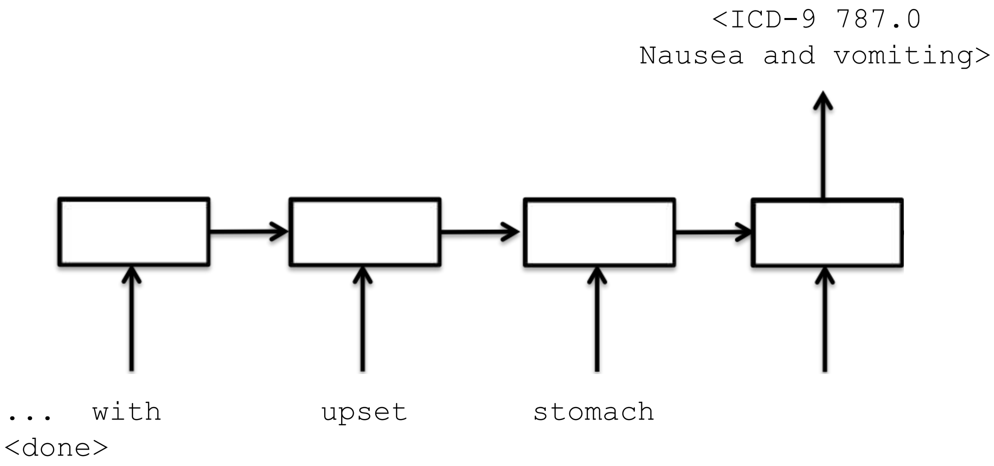
Choi, Multi-layer Representation Learning for Medical Concepts, 2016

Example clinical note

Mr. Smith is a 63-year-old gentleman with coronary artery disease, hypertension, hypercholesterolemia, COPD and tobacco abuse. He reports doing well. He did have some more knee pain for a few weeks, but this has resolved. He is having more trouble with his sinuses. I had started him on Flonase back in December. He says this has not really helped. Over the past couple weeks he has had significant congestion and thick discharge. No fevers or headaches but does have diffuse upper right-sided teeth pain. **He denies any chest pains, palpitations, PND, orthopnea, edema or syncope.** His breathing is doing fine. No cough. He continues to smoke about half-a-pack per day. He plans on **trying the patches again.**

- Notes are still written together with codes selected for each visit.
- Often explaining details which do not have equivalent codes.
- Natural language is very difficult to parse with certainty.

Predicting codes from notes



Converting text to codes can

- Adapt old databases
- Correct errors
- Upgrade ICD versions

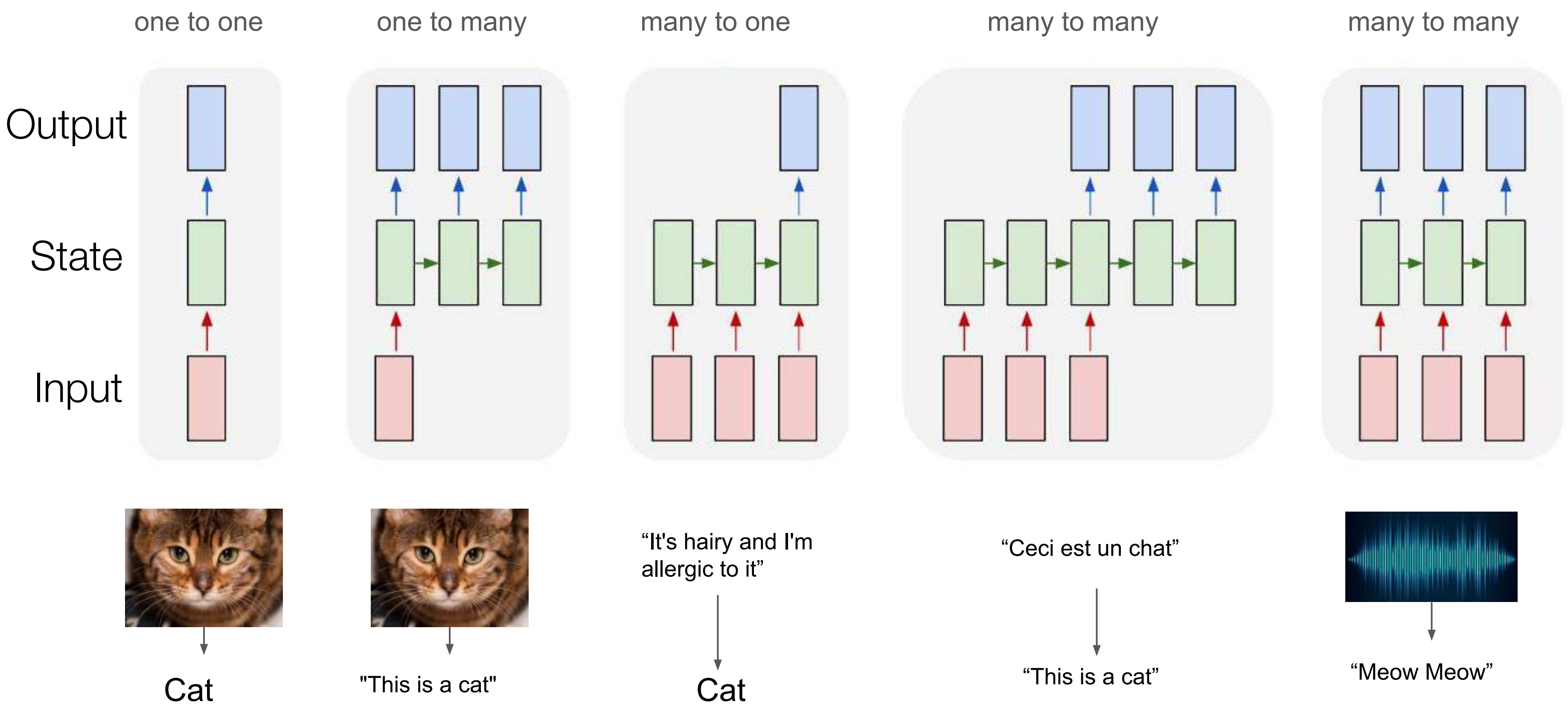
Bidirectional RNN for Medical Event Detection in Electronic Health Records

Abhyuday N Jagannatha[1], Hong Yu[1,2]
[1] University of Massachusetts, MA, USA
[2] Bedford VAMC and CHOIR, MA, USA

Abstract

Jagannatha, Bidirectional RNN for Medical Event Detection in Electronic Health Records, 2016

RNNs, Different types of sequential prediction tasks



Taken from http://vision.stanford.edu/teaching/cs231n/slides/2016/winter1516_lecture10.pdf and Francis Dutil

RNNs

An RNN applies a function over a sequence of inputs $[x_1, x_2, \ldots, x_T]$
which produces a sequence of outputs $[y_1, y_2, \ldots, y_T]$
and each input produces a internal state $[h_1, h_2, \ldots, h_T]$.

Sequence of outputs

Sequence of internal states

Sequence of inputs

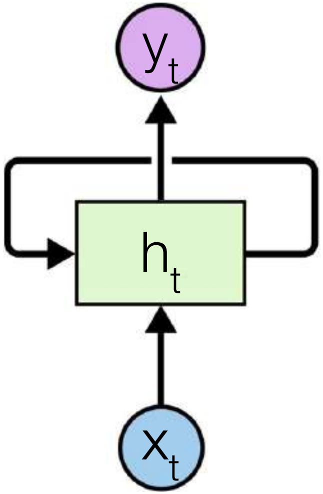


Image du blog de Christopher Olah, slide from l'École d'automne 2018, César Laurent

RNNs

- A simple RNN

$$h_t = \tanh(Ux_t + Wh_{t-1})$$
$$y_t = f(Vh_t)$$

- W, U and V are the parameters of the network.
- The weights are shared over time.

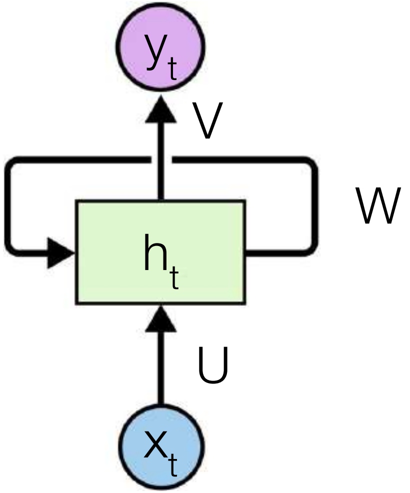
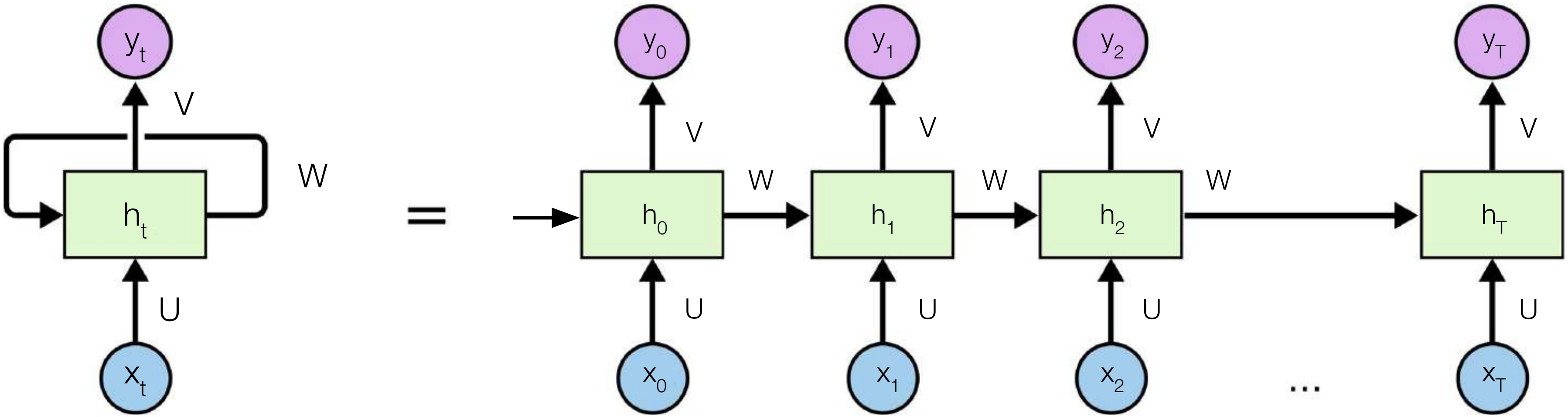


Image du blog de Christopher Olah, slide from l'École d'automne 2018, César Laurent

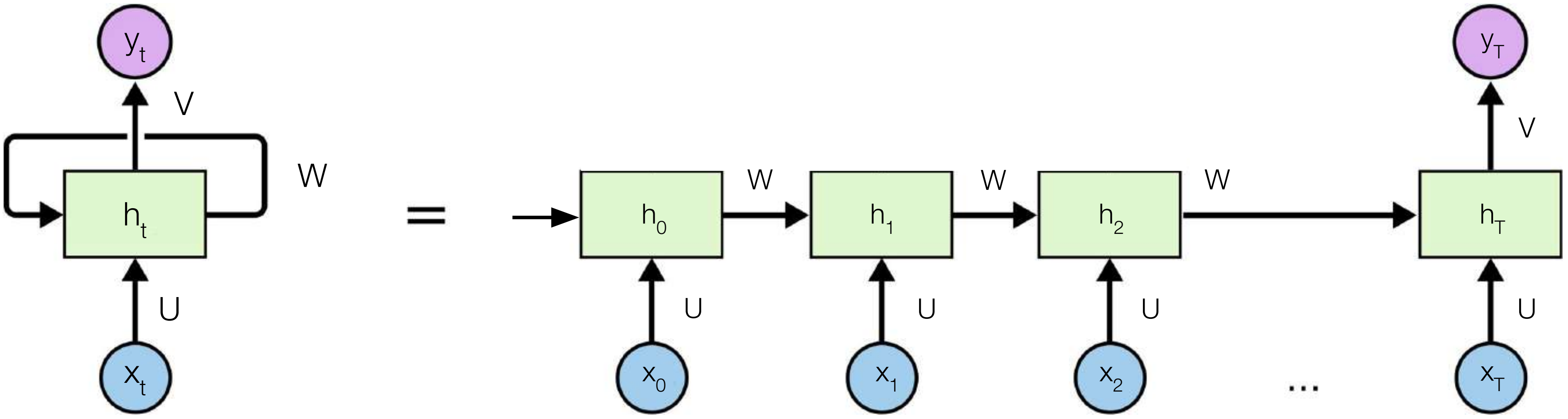
Unrolling the RNN over time



The weights are shared over time.

Image du blog de Christopher Olah, slide from l’École d’automne 2018, César Laurent

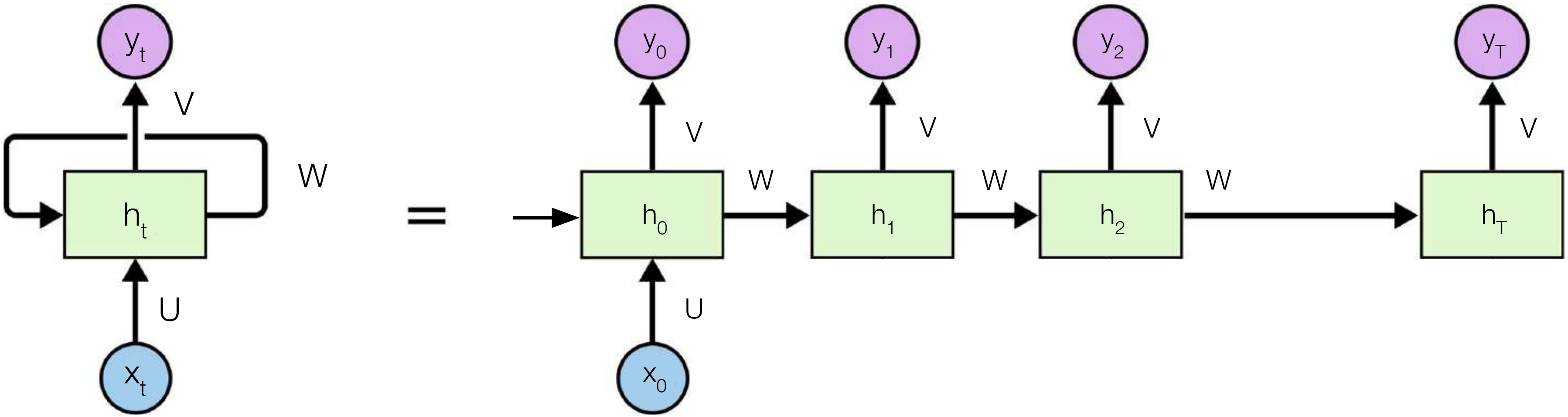
Unrolling the RNN over time



The weights are shared over time.

Image du blog de Christopher Olah

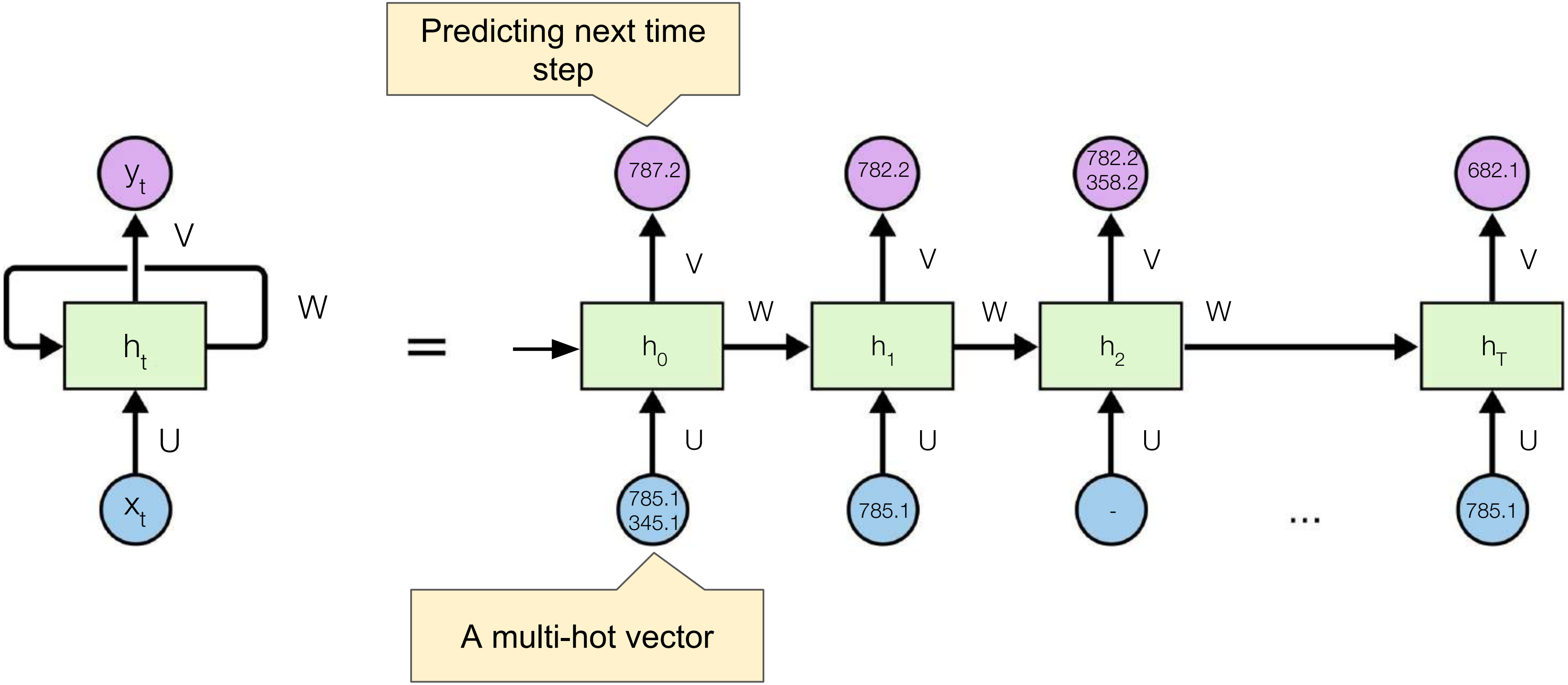
Unrolling the RNN over time



The weights are shared over time.

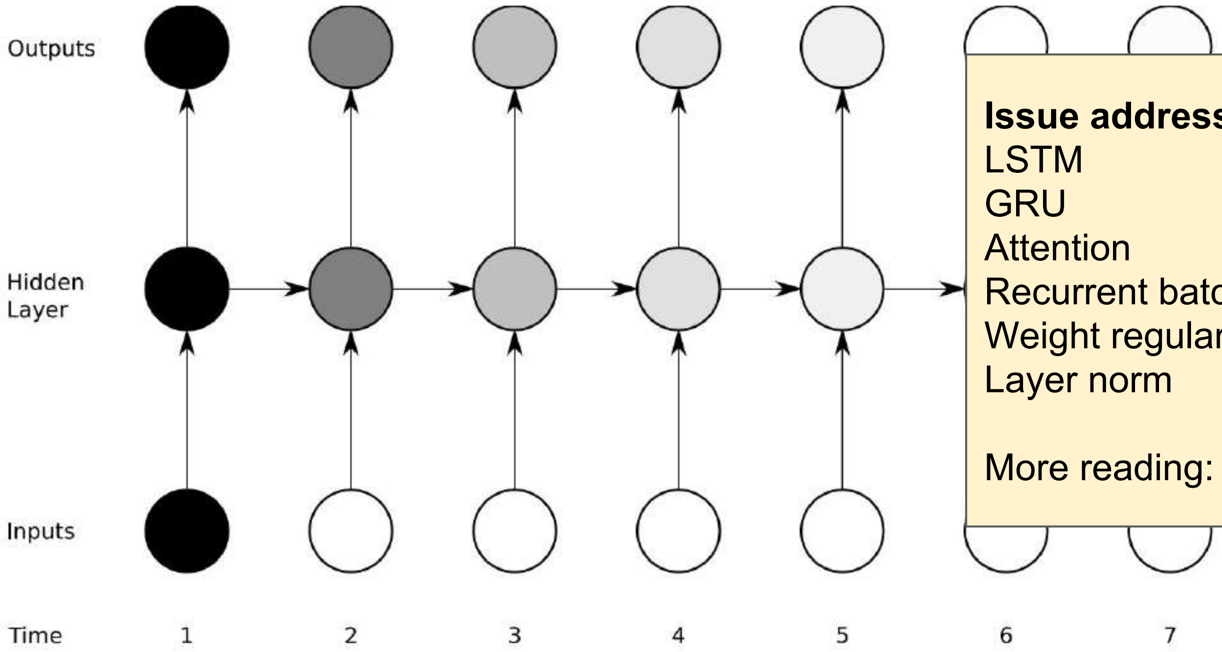
Image du blog de Christopher Olah

Predicting future events



Vanishing gradients

Problem with the basic RNN



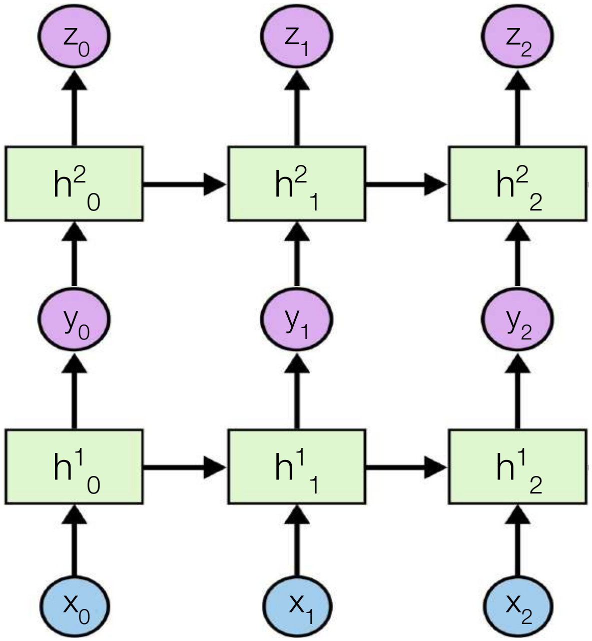
Issue addressed by:
LSTM
GRU
Attention
Recurrent batch norm
Weight regularization
Layer norm

More reading: [Pascanu 2013]

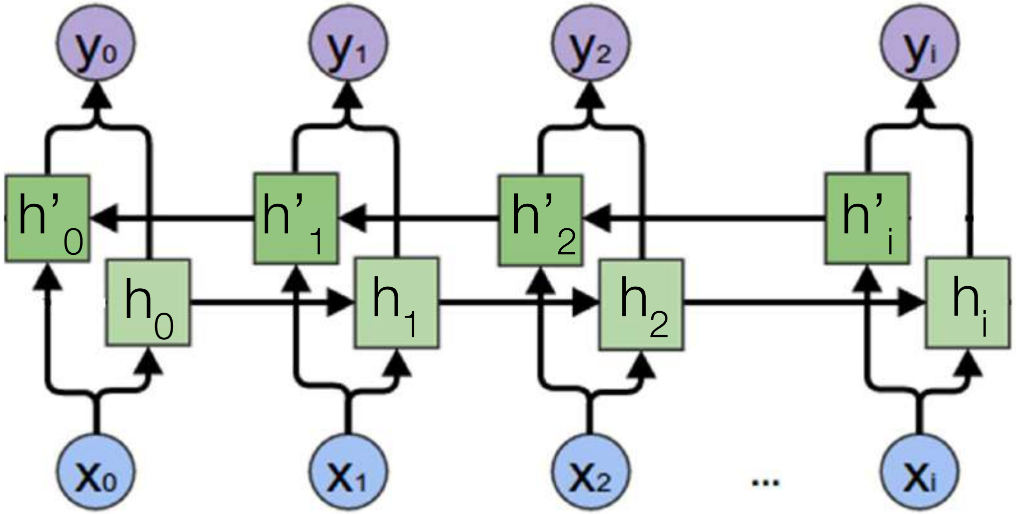
The shade of gray shows the influence of the input of the RNN at time 1. It decreases over time, as the RNN gradually forgets its first input.

Slide from l'École d'automne 2018, César Laurent

We can get creative with RNN designs



Stacked RNNs

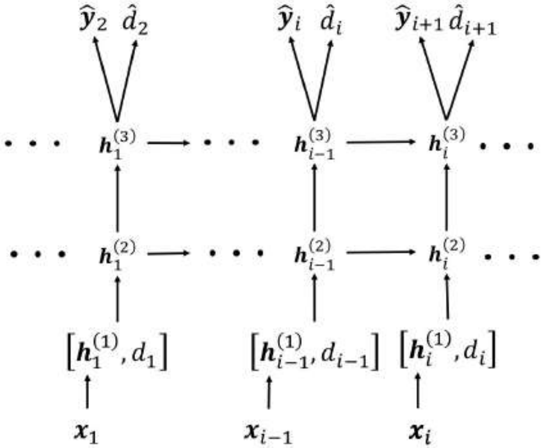


Bi-directional RNN

Slide from l'École d'automne 2018, César Laurent

RNNs for next visit prediction (Doctor AI)

y = High level medical codes R^{1778}
d = time since last visit



Medical codes R^{40000}

- Treat codes equally: ICD diagnosis codes, procedure codes, and medication codes
- Grouped codes into higher-order categories


Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori

Andy Schuetz, Walter F. Stewart

Jimeng Sun

Abstract

Table 1: Basic statistics of the the clinical records dataset.

# of patients	263,706	Total # of codes	38,594
Avg. # of visits	54.61	Total # of 3-digit Dx codes	1,183
Avg. # of codes per visit	3.22	# of top level Rx codes	595
Max # of codes per visit	62	Avg. duration between visits	76.12 days

Patients from Sutter Health Palo Alto Medical Foundation

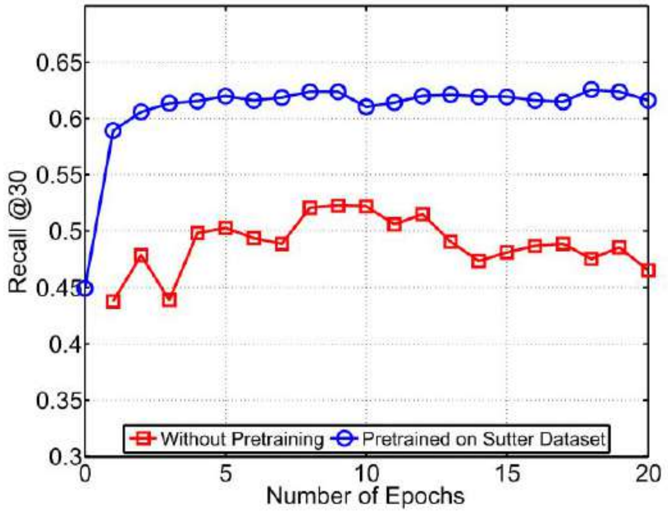
Algorithms	Dx,Rx,Time Recall @k			
	$k = 10$	$k = 20$	$k = 30$	R^2
Last visit	26.25			—
Most freq.	48.11	60.23	66.00	—
Logistic	36.04	46.32	52.53	0.0726
MLP	38.82	49.09	55.74	0.1221
RNN-1	53.86	65.10	71.24	0.2519
RNN-2	53.61	64.93	71.14	0.2528
RNN-1-IR	54.37	65.68	71.85	0.2492
RNN-2-IR	**54.96**	**66.31**	**72.48**	**0.2534**

Choi, Doctor AI: Predicting Clinical Event via Recurrent Neural Networks, 2016

Pretraining RNNs (Doctor AI)

MIMIC II has 2,695 patients with 2+ visits

Pretrained using a larger Sutter Health dataset
~300,000 patients




Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori
Andy Schuetz, Walter F. Stewart
Jimeng Sun

Abstract

Choi, Doctor AI: Predicting Clinical Events via Recurrent Neural Networks, 2016

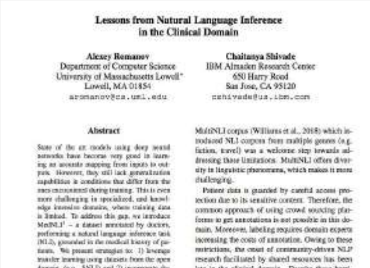
Related work

Lipton, Zachary C et al. “Learning to Diagnose with LSTM Recurrent Neural Networks.” International Conference on Learning Representations. 2016

Che, Zhengping et al. “Recurrent Neural Networks for Multivariate Time Series with Missing Values.” Nature Scientific Reports. 2018

Medical Natural Language Inference

#	Premise	Hypothesis	Label
1	ALT , AST , and lactate were elevated as noted above	patient has abnormal lfts	entailment
2	Chest x-ray showed mild congestive heart failure	The patient complains of cough	neutral
3	During hospitalization , patient became progressively more dyspnic requiring BiPAP and then a NRB	The patient is on room air	contradiction
4	She was not able to speak , but appeared to comprehend well	Patient had aphasia	entailment
5	T1DM : x 7yrs , h/o DKA x 6 attributed to poor medication compliance , last A1c [** 3-23 **] : 13.3 % 2	The patient maintains strict glucose control	contradiction
6	Had an ultimately negative esophagogastroduodenoscopy and colonoscopy	Patient has no pain	neutral
7	Aorta is mildly tortuous and calcified .	the aorta is normal	contradiction


Lessons from Natural Language Inference in the Clinical Domain

Alexey Romanov
Department of Computer Science
University of Massachusetts Lowell
Lowell, MA 01854

Chaitanya Shivade
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

Abstract

MedNLI dataset derived from the MIMIC-III dataset

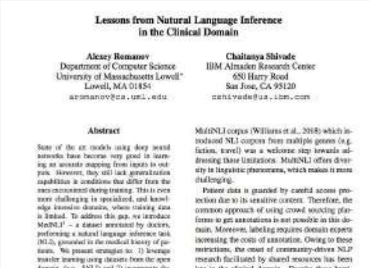
Romanov, Lessons from Natural Language Inference in the Clinical Domain, 2018

Medical Natural Language Inference

Studying the errors we can see the limits of the model.

Category	Premise	Hypothesis	Predicted	Expected
Numerical	On weaning to 6LNC, his O2 decreased to 81-82%	He has poor O2 stats	neutral	entailment
Reasoning	WBC 12 , Hct 41 .	WBC slightly elevated	contradiction	entailment
World	The infant emerged with spontaneous cry.	The infant was still born.	entailment	contradiction
Knowledge	No known sick contacts	No recent travel	entailment	neutral
Abbreviation	No CP or fevers.	Patient has no angina	neutral	entailment
	Received GI cocktail for h/o GERD, esophageal spasm	Received a proton pump inhibitor	entailment	neutral
Medical	EKG showed T-wave depression in V3-5, with no prior EKG for comparison.	Patient has a normal EKG	neutral	contradiction
Knowledge	Mother developed separation of symphysis pubis and was put in traction .	She has orthopedic injuries	neutral	entailment
Negation	Head CT was negative for bleed.	The patient has intracranial hemorrhage	neutral	contradiction
	Denied headache, sinus tenderness, or congestion	Patient has headaches	neutral	contradiction

Table 9: Representative errors made by different models



Romanov, Lessons from Natural Language Inference in the Clinical Domain, 2018

Discussion

"While code-based representations of clinical concepts and patient encounters are a tractable first step towards working with heterogeneous EHR data, **they ignore many important real-valued measurements** associated with items such as laboratory tests, intravenous medication infusions, vital signs, and more." [Shickel et al,, 2018]

Discussion

"While some researchers downplay the **importance of interpretability in favor of significant improvements in model performance**, we feel advances in deep learning transparency will hasten the widespread adoption of such methods in clinical practice." [Shickel et al, 2018]

Discussion

"Many studies claim state-of-the-art results, but few can be verified by external parties. **This is a barrier** for future model development and one cause of the slow pace of advancement."
[Shickel et al, 2018]

Mila Medical

Joseph Paul Cohen, PhD

Myriam Côté, PhD

Pr. Yoshua Bengio, PhD

Mandana Samiei

Francis Dutil

Martin Weiss

Shawn Tan

Geneviève Boucher

Georgy Derevyanko, PhD

Tristan Sylvain

Margaux Luck, PhD

Sina Honari

Assya Trofimov

Vincent Frappier, PhD

References

Ching, T., et al. **Opportunities And Obstacles For Deep Learning In Biology And Medicine**. Journal of The Royal Society Interface. 2018

Shickel, B. et al. **Deep EHR : A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record**. IEEE Journal of Biomedical and Health Informatics, 2018