Deep learning for biomedicine II

Truyen Tran
Deakin University

Seoul, Nov 2017
Resources

Slides and references:
  • https://truyentran.github.io/acml17-tute.html

Key survey paper (updated frequently):
Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (25 mins)
  - Classic architectures
  - Capsules
  - Graphs
  - Memory-augmented nets

Topic 3: Genomics (25 mins)
  - Nanopore sequencing
  - Genomics modelling

Topic 4: Biomedical imaging (15 mins)
  - Cellular imaging
  - Diagnostics imaging
  - EEG/ECG

Topic 5: Healthcare (25 mins)
  - Time-series of physio measures
  - Trajectories prediction

Topic 6: Generative biomed (30 mins)
  - Few-shot learning
  - Generative models
  - Drug design
  - Future outlook

QA (10 mins)
Sensing technologies and data

Raw signals are ideal candidates for deep learning

Speech & vision techniques can be applied with minimal changes

The state of biomedical imaging

As estimated by IBM, 90% data of healthcare is imaging.

Biomedical imaging is perhaps the most ready area for current DL techniques:
- It typically means CNN!
- The game is in the data acquisition and problem definition/transformation

Examples of application areas:
- Cellular imaging
- Tumor Detection & tracking
- Blood Flow Quantification and Visualization
- Medical Interpretation
- Diabetic Retinopathy

Tasks
- Classification
- Segmentation
- Localization

Challenges
- Low quality
- Very high resolution
- Tiny localized areas

Skin cancer diagnosis using **inception-v3**


- 129,450 clinical images
- 2,032 different diseases
- Test against 21 board-certified dermatologists
- Use case 1: keratinocyte carcinomas versus benign seborrheic keratoses;
- Use case 2: malignant melanomas versus benign nevi.
Microscopy + mobile phone + CNN

Highly relevant for:
- Developing countries
- Rural areas


Figure 1: Microscope smartphone adapter: design of components (left), 3D-printed adapter mounted on microscope (center), smartphone inserted into adapter (right).
EEG $\rightarrow$ Tensor RBM for alcoholic diagnosis

Subject representation

Tensor Restricted Boltzmann Machine

3D Spectrogram

EEG → Matrix LSTM → Classification

EEG segments as matrices

Temporal dynamics as recurrence

Recurrent dynamics

\[
H_t = \sigma(U^T_x X_t V_x + U^T_h H_{t-1} V_h + B)
\]

ECG → CNN for heart attack detection

“They should stop training radiologists now.”

Geoff Hinton (as of April 2017)
Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (25 mins)
  - Classic architectures
  - Capsules
  - Graphs
  - Memory-augmented nets

Topic 3: Genomics (25 mins)
  - Nanopore sequencing
  - Genomics modelling

QA (10 mins)

Topic 4: Biomedical imaging (15 mins)
  - Cellular imaging
  - Diagnostics imaging
  - EEG/ECG

Topic 5: Healthcare (25 mins)
  - Time-series of physio measures
  - Trajectories prediction

Topic 6: Generative biomed (30 mins)
  - Few-shot learning
  - Generative models
  - Drug design
  - Future outlook

Break (20 mins)

QA (10 mins)
Modeling electronic medical records (EMR)

Need to model the healthcare processes, which are interactions of:

- Disease progression
- Interventions & care processes
- Recording processes (Electronic Medical/Health Records)

Source: medicalbillingcodings.org
EMR Connects Services: System of Systems

Five main functions
- Integrated view of patient data
- Clinical decision support
- Clinician order entry
- Access to knowledge resources
- Integrated communication and reporting support
Clinical Decision Supports

Support protocol/planning of treatment/discharge.

Suggest course of actions:
- E.g., medication/dose/duration.

Estimate risk & predict outcomes.

Alert/reminder.

Support (semi) automated diagnosis.

risk prediction (prognosis)

- heart failure
- diabetes
- mental health
- COPD
- heart attack
- cancers
- preterm
- suicide attempts
- side effects
- death
- toxicity
- readmission
- quality-of-life
- stress
- progression to advanced stages
- length-of-stay
- cancers
- diabetes
- COPD
- heart failure
- heart attack
- preterm
- suicide attempts
- side effects
- death
- toxicity
- readmission
- quality-of-life
- stress
- progression to advanced stages
- length-of-stay
Warning: leakage!

Make sure the patients are counted AFTER first diagnosis

- Often, we have future data as well
- Retrospective nature

Never use outcomes to do anything, except for training the model

Our early suicide attempt classification from assessments was a form of leakage:

- Any attempt in history is considered as an outcome. BUT:
  - Previous attempts were accounted in current assessment already!
Drugs & tests

- Drug companies offer different brand names of the essentially the same drug
- DDD/ATC is the central register for the medication classes, maintained by WHO
- Several test names may be the same

It may not be robust to use the original "vocabularies"

- Tens of thousands of ICD-codes, thousands of procedures, hundreds of DRGs, thousands of medication classes
- Codes are usually organized in hierarchy
- Choosing the right hierarchy is statistical issue
Attend to risks in Intensive Care Unit (ICU)

The needs
- Accuracy
- Interpretability
- As early as possible

The process:
- Irregular time-series → Regular time-steps → Data imputation → Bi-LSTM → Multiple attentions → Classification


Data: Physionet 2012
Attend to risks in ICU

State transition

Attention probabilities

Physiological measures

DeepPatient: Representing medical records with Stacked Denoising Autoencoder

Feature detector

Auto-encoder

Raw data

Representation

Decoder

Reconstruction

Encoder

Deep Auto-encoder

**DeepPatient: Results on disease classification**

<table>
<thead>
<tr>
<th>Patient Representation</th>
<th>AUC-ROC</th>
<th>Accuracy</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>RawFeat</td>
<td>0.659</td>
<td>0.805</td>
<td>0.084</td>
</tr>
<tr>
<td>PCA</td>
<td>0.696</td>
<td>0.879</td>
<td>0.104</td>
</tr>
<tr>
<td>GMM</td>
<td>0.632</td>
<td>0.891</td>
<td>0.072</td>
</tr>
<tr>
<td>K-Means</td>
<td>0.672</td>
<td>0.887</td>
<td>0.093</td>
</tr>
<tr>
<td>ICA</td>
<td>0.695</td>
<td>0.882</td>
<td>0.101</td>
</tr>
<tr>
<td>DeepPatient</td>
<td>0.773*</td>
<td>0.929*</td>
<td>0.181*</td>
</tr>
</tbody>
</table>

*Time Interval = 1 year (76,214 patients)*

Accuracy and F-Score marked with an asterisk (*) indicate significant improvement compared to other methods.
Trajectories modeling: Challenges & opportunities

- Long-term dependencies
- Irregular timing
- Mixture of discrete codes and continuous measures
- Complex interaction of diseases and care processes
- Cohort of interest can be small (e.g., <1K)
- Rich domain knowledge & ontologies

Multimodalities: Text, physiological signals (e.g., EEG/ECG), images (e.g., MRI, X-ray, retina), genomics

New modalities: social medial, wearable devices

**Explainability!**
Visualisation and interpretation are keys!
Deepr: CNN for repeated motifs and short sequences

Deepr: Disease embedding & motifs detection

**E11 I48 I50**
Type 2 diabetes mellitus
Atrial fibrillation and flutter
Heart failure

**E11 I50 N17**
Type 2 diabetes mellitus
Heart failure
Acute kidney failure
DeepCare: intervened long-term memory of health

Illness states are a dynamic memory process → moderated by time and intervention

Discrete admission, diagnosis and procedure → vector embedding

Time and previous intervention → “forgetting” of illness

Current intervention → controlling the risk states

DeepCare: Dynamics

prev. memory * memory

forget gate

input

input gate

output gate

history states

current data

previous intervention

time gap

current intervention

aggregation over time → prediction

New in DeepCare
DeepCare: Structure

- Vector embedding
- Admission (disease)
- Admission (intervention)
- Latent states
- LSTM
- Multiscale pooling
- Neural network
- Future risks
- Long short-term memory
- Time gap
- History
- Future

15/11/17
DeepCare: Two modes of forgetting as a function of time

$\rightarrow$ decreasing illness

$\rightarrow$ Increasing illness
DeepCare: prediction results

Intervention recommendation (precision@3)

Unplanned readmission prediction (F-score)
Trajectories prediction

Generating a subset of treatments

Generating an entire health/care trajectory

Challenges: global loss, meaningful evaluation metrics

A solution: Memory-augmented neural nets (MANN)

Illustration of the DNC architecture

Source: deepmind.com


(LeCun, 2015)
Dual Controller Write-protected Memory-Augmented Neural Networks (DCw-MANN) (*Le et al, work in progress*)

Separate controllers for encoding and decoding

Memory is write-protected during decoding

Use memory as a replacement for:

- Attention
- Skip connection
## DCw-MANN

<table>
<thead>
<tr>
<th>Model</th>
<th>Procedure Output</th>
<th>Drug Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>BLEU</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>25.60</td>
<td>N/A</td>
</tr>
<tr>
<td>Random Forest</td>
<td>26.84</td>
<td>N/A</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>27.16</td>
<td>36.42</td>
</tr>
<tr>
<td>Seq2Seq with attention</td>
<td>27.33</td>
<td>36.83</td>
</tr>
<tr>
<td>DNC</td>
<td>28.28</td>
<td>37.69</td>
</tr>
<tr>
<td>DC-MANN</td>
<td>28.54</td>
<td>38.51</td>
</tr>
<tr>
<td>DCw-MANN</td>
<td><strong>30.36</strong></td>
<td><strong>39.74</strong></td>
</tr>
</tbody>
</table>

**Table 3.** Results on MIMIC-III dataset for procedure prediction and drug prescription. The higher the better.
Modeling multiple disease-treatment interactions over time

Co-morbidity is the norm in modern medicine
Each hospital visit contains a set of diseases and a set of treatments
There are interactions between multi-diseases and multiple-treatments

Algebraic view: \( \text{Health} = RNN(\text{Illness} - \text{Intervention}) \)

\[
\begin{align*}
v_t &= \rho(\Delta) \quad \text{where} \quad \Delta = d_t - p_t \\
fe(S) &\leftarrow \frac{\bar{e}_S}{\epsilon + \|\bar{e}_S\|} \quad \text{where} \quad \bar{e}_S = \max(0, \sum_{i \in S} e_i)
\end{align*}
\]
Results (AUC)

<table>
<thead>
<tr>
<th>Method</th>
<th>Diabetes</th>
<th>Mental health</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+LR</td>
<td>0.673</td>
<td>0.705</td>
</tr>
<tr>
<td>Deepr [14]</td>
<td>0.680</td>
<td>0.714</td>
</tr>
<tr>
<td>MDMTP+LSTM</td>
<td>0.718</td>
<td>0.726</td>
</tr>
<tr>
<td>MDMT+LSTM</td>
<td>0.701</td>
<td>0.730</td>
</tr>
</tbody>
</table>

(a) Worsening progression \( (P = 0.70) \)  
(b) Improving progression \( (P = 0.23) \)
Big room: Towards personalized healthcare

Medical practice as recommender systems (Xavier Amatriain, healthcare Recsys Workshop, Como, 2017)

Clinical Practice Guides are not personalized

Research done on “homogeneous”, healthy subjects

It is very hard for doctors to “manually” personalize their “recommendations”


Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (25 mins)
  - Classic architectures
  - Capsules
  - Graphs
  - Memory-augmented nets

Topic 3: Genomics (25 mins)
  - Nanopore sequencing
  - Genomics modelling

Topic 4: Biomedical imaging (15 mins)
  - Cellular imaging
  - Diagnostics imaging
  - EEG/ECG

Topic 5: Healthcare (25 mins)
  - Time-series of physio measures
  - Trajectories prediction

Topic 6: Generative biomed (30 mins)
  - Few-shot learning
  - Generative models
  - Drug design
  - Future outlook

QA (10 mins)

Break (20 mins)
Few-shot deep learning

Lots of biomedical problems are data poor
  ▪ Rare drugs
  ▪ Rare diseases

Distance metrics learning (DML) methods
  ▪ Learn to full any pair of the similar data points, and push the dissimilar
  ▪ Well-known methods: **Siamese networks**

Meta-learning strategies
  ▪ Tasks are presented in sequence
  ▪ New tasks can borrow from similar prior tasks

Meta-learning strategies

(Santoro et al, 2016)

(Mishra et al, 2017)
Generative models

Many applications:

• Text to speech

• **Simulate data that are hard to obtain/share in real life (e.g., healthcare)**

• Generate meaningful sentences conditioned on some input (foreign language, image, video)

• Semi-supervised learning

• Planning
A family: RBM $\rightarrow$ DBN $\rightarrow$ DBM

\[ p(v, h; \psi) \propto \exp[-E(v, h; \psi)] \]

Restricted Boltzmann Machine
(~1994, 2001)

Deep Belief Net
(2006)

Deep Boltzmann Machine
(2009)
Variational Autoencoder
(Kingma & Welling, 2014)

Two separate processes: generative (hidden $\rightarrow$ visible) versus recognition (visible $\rightarrow$ hidden)

http://kvfrans.com/variational-autoencoders-explained/
GAN: implicit density models
(Adapted from Goodfellow’s, NIPS 2014)

\[ p_D(\text{data}) \]

Data distribution

Model distribution

Poorly fit model

After updating D

After updating G

Mixed strategy equilibrium
GAN: Generated Samples

Some high quality pictures generated by GAN

http://kvfrans.com/generative-adversial-networks-explained/
Generative deep learning for drug discovery

Predicting bioactivities from molecules

Drug representation, unsupervised learning from graphs

Generate from bioactivities to molecular graphs

Combinatorial chemistry

Generate variations on a template

Returns a list of molecules from this template that

▪ Bind to the pocket with good pharmacodynamics?
▪ Have good pharmacokinetics?
▪ Are synthetically accessible?

#REF: Talk by Chloé-Agathe Azencott titled “Machine learning for therapeutic research”, 12/10/2017
First step: Map molecule → drug properties (binding/acting)

Drugs are small bio-molecules

Traditional techniques:
- Graph kernels (ML)
- Molecular fingerprints (Chemistry)

Modern techniques
- Molecule as graph: atoms as nodes, chemical bonds as edges

Molecular fingerprints

Algorithm 1 Circular fingerprints.

1: **Input:** molecule, radius $R$, fingerprint length $S$
2: **Initialize:** fingerprint vector $x \leftarrow 0_S$
3: **foreach** atom $a$ in molecule
4: $r_a \leftarrow q(a)$  #extract initial atom features
5: **for** $L = 1$ to $R$  #loop through layers
6: **foreach** atom $a$ in molecule
7: $r_1...r_N = \text{neighbors}(a)$
8: $v \leftarrow [r_a, r_1, ..., r_N]$  #combine neighbor features
9: $r_a \leftarrow \text{hash}(v)$  #refine atom features
10: $i \leftarrow \text{mod}(r_a, S)$  #convert to index
11: $x_i \leftarrow 1$  #Write 1 (indicator) at index
12: **Return:** binary vector $x$.

Graph memory networks
(Pham et al, 2017, work in progress)
Graph memory networks: Results

Figure 2: F1-score (%) for NCI datasets. FP = Fingerprint; RF = Random Forests; GBM = Gradient Boosting Machine. Best view in color.
One-shot learning for drug discovery

Drug design and generation

We now have methods for compute bioactivities of a drug molecule.

We need a reverse method to generate drug molecules from desirable bioactivities.

The space of drugs is estimated to be $1e+23$ to $1e+60$.
- Only $1e+8$ substances synthesized thus far.

It is impossible to model this space fully.

The current technologies are not ready for graph generations.

But approximate techniques do exist.
The trick: Molecule $\rightarrow$ string

Using SMILES representation of drug, to convert a molecular graph into a string
- **SMILES = Simplified Molecular-Input Line-Entry System**

Then using sequence-to-sequence + VAE/GAN to model the continuous space that encodes/decodes SMILES strings
- Allow easy optimization on the continuous space
- **Problem:** String $\rightarrow$ graphs is not unique!

A better way is to encode/decode graph directly.

Uses VAE for sequence-to-sequence.

Adversarial Autoencoders
Kadurin et al. Molecular pharmaceutics 2017

Input of the encoder: the fingerprint of a molecule

The decoder outputs the predicted fingerprint.

The generative model generates a vector E, which is then discriminated from the latent vector of the real molecule by the discriminator.


Living in the future: AI for health care

Some speculations (by me):


Bear in mind that anything beyond 5 years are nearly impossible to predict!

Kai-Fu Lee’s vision:

- Wave 1: Internet data (PubMed, social media)
- Wave 2: Business data (EMR)
- Wave 3: Digitalize the physical world (Drugs)
- Wave 4: Full automation (Robot surgeons, GPs)
Toward personalize medicine

Will this patient response to that treatment?
Can we find the best treatment for a patient?
Which biomarkers predict the patient’s response?
Sound familiar to Recommender Systems (patient = user, treatment = item)?

#REF: Talk by Chloé-Agathe Azencott titled “Machine learning for therapeutic research”, 12/10/2017
Towards a dialog system → Replace GP?

Leveraging existing knowledge
- Medical knowledge bases
- Medical texts
- Probably needs to build knowledge bases from text

Personalizing through EMRs
- Learn from hospitals data

Ask right questions → Finding answers from databases → Generating dialog

Never ending learning (NEL).

Me and Woebot

So tell me, what’s your energy like?

And what are you doing now (besides talking to lil ol me)?

Write it here

nothing

How about your mood, how do you feel right now?

Write it here

not bad
Other rooms for deep learning

-omics
  - Gene expression
  - Proteomics

Neuroscience
  - Models for spike trains
  - Deep learning for connectomics
  - Neuroscience-inspired DL

Biomedical NLP
  - Classical NLP
  - Social media
  - Knowledge graphs

Wearables
  - Tracking the state of physical and mental health
  - Lifestyle management & monitoring

Health Insurance
  - Future illness/spending prediction
  - Proactive prevention programs
  - **WARNING:** Working for insurance companies does raise ethical concerns!

Nutrition
  - Mobile phone vision → calories

Explainable AI
  - Seeing through the black-box, e.g., visualization, motifs
  - Explainable architectures that use biological mechanisms and medical ontologies
  - Dual architecture: predictor & explainer
Thank you!

We’re hiring
PhD & Postdocs
truyen.tran@deakin.edu.au

http://ahsanqawl.com/2015/10/qa/