Week 5 Reflections

≡ Course	CS 598 - Deep Learning for Healthcare
≔ Submitted By	Rohit Bhagwat (rohitb2@illinois.edu)

Questions

What are the main messages you learned from this chapter?

This week we were introduced to the concept of Embedding.

An embedding is a mapping of a discrete — categorical — variable to a vector of continuous numbers. Neural network embeddings are useful because they can *reduce the dimensionality* of categorical variables and *meaningfully represent* categories in the transformed space.

Neural network embeddings overcome the two limitations of a common method for representing categorical variables: one-hot encoding.

In this chapter, we were introduced to a set of neural network models called embedding methods to represent medical concepts (e.g., diagnoses, medications, and procedures) based on co-occurrence patterns in longitudinal electronic health records.

The intuition behind the embedding methods is to map medical concepts co-occurring closely in EHR data to similar embedding vectors (i.e., the embedding distance between related medical concepts is small).

We looked at the following models -

- **Word2Vec**: is based on the proximity of medical codes appearing in the records

- **Med2Vec**: leverages two-level hierarchical structures, namely clinical visits over time and co-occurrence within a visit

- **t-SNE embedding**: is a nonlinear dimensionality reduction method for visualizing high-dimensional data in 2D space

- **MIME**: utilizes a three-level hierarchy, namely visits, diagnosis codes, then treatment codes

Which part do you want to improve in this chapter?

In the lectures as well as in the chapter, we are introduced to the concepts very well. The author and the professor walked through the idea behind embedding and how it is useful when applying neural networks for Healthcare.

I thought we were introduced to mathematical formulae directly and it was assumed that we as students know about it. As an example when discussing ***t-distributed stochastic neighbor embedding,*** we saw that the objective of t-SNE is to minimize the Kullback-Leibler (KL) divergence. If I have to recommend any improvements I'll recommend adding references to learn more about KL divergence, how the formula is derived etc..

What related topics do you suggest adding in this chapter?

I would suggest adding (or providing references to students) to learn more on the following -

- Embedding methods

- Text information systems

- Text retrieval and text mining

> 💡 I think the course CS 410 Text Information Systems is very useful for understanding more on this subject

What is the key difference between Med2Vec and Word2Vec?

Word2Vec assumes a global ordering of the input medical codes to create the embedding.

Many medical codes are documented together within a visit without any specific order, making it hard to use Word2Vec directly.

The EHR data of a patient follow a two-level hierarchy:

1. Visit level: Patient EHR data consists of a sequence of visits over time.

2. Code level: Each visit includes multiple medical codes, e.g., diagnosis, procedure, and medication codes.

This hierarchical structure provides two types of relational information, namely the sequential order of visits and co-occurrence of the codes within a visit.
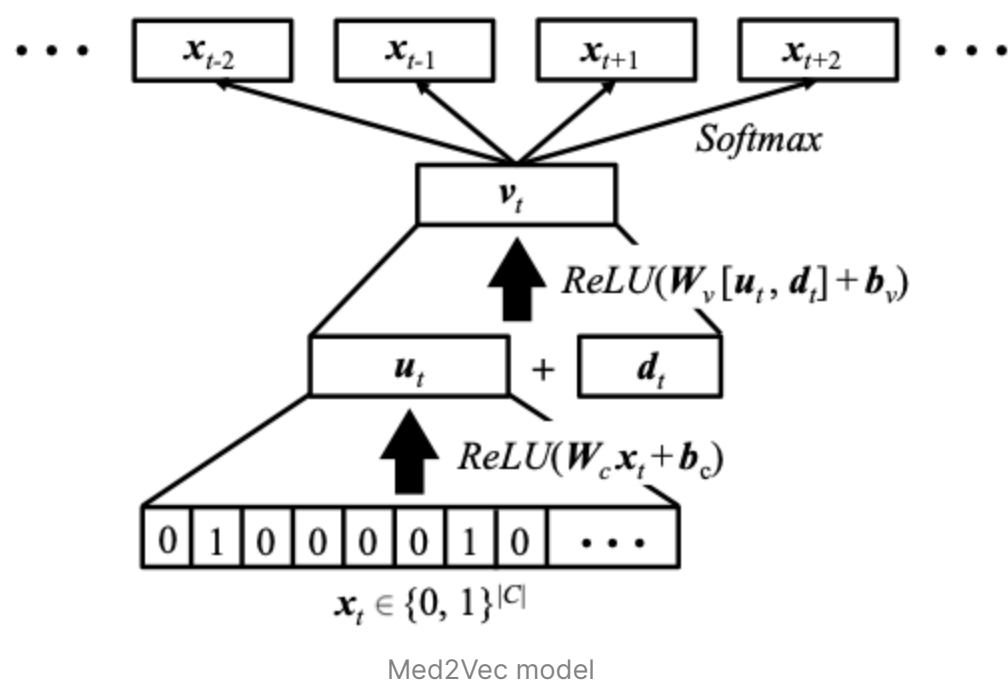
Med2Vec is a **hierarchical** embedding method that learns the representations for both medical codes and visits.

The EHR data for each patient is in the form of a sequence of visits V1, . . . , VT where each visit contains a subset of medical codes.

The goal of Med2Vec is to learn two types of representations:

- Code representations map every code in the set of all medical codes C to non-negative real-valued vectors.

- Visit representations learn another embedding that maps every visit (a set of medical codes) to a real-valued vector of dimension n.

The objective is to be able to use the visit level representation to predict what's going to happen before and after this visit which is not possible

$$x_{t-2} \quad x_{t-1} \quad x_{t+1} \quad x_{t+2}$$

Softmax

$$v_t$$

$$ReLU(W_v[u_t, d_t] + b_v)$$

$$u_t \quad + \quad d_t$$

$$ReLU(W_c x_t + b_c)$$

$$\boxed{0\,|\,1\,|\,0\,|\,0\,|\,0\,|\,0\,|\,1\,|\,0} \quad \cdots$$

$$x_t \in \{0, 1\}^{|C|}$$

Med2Vec model

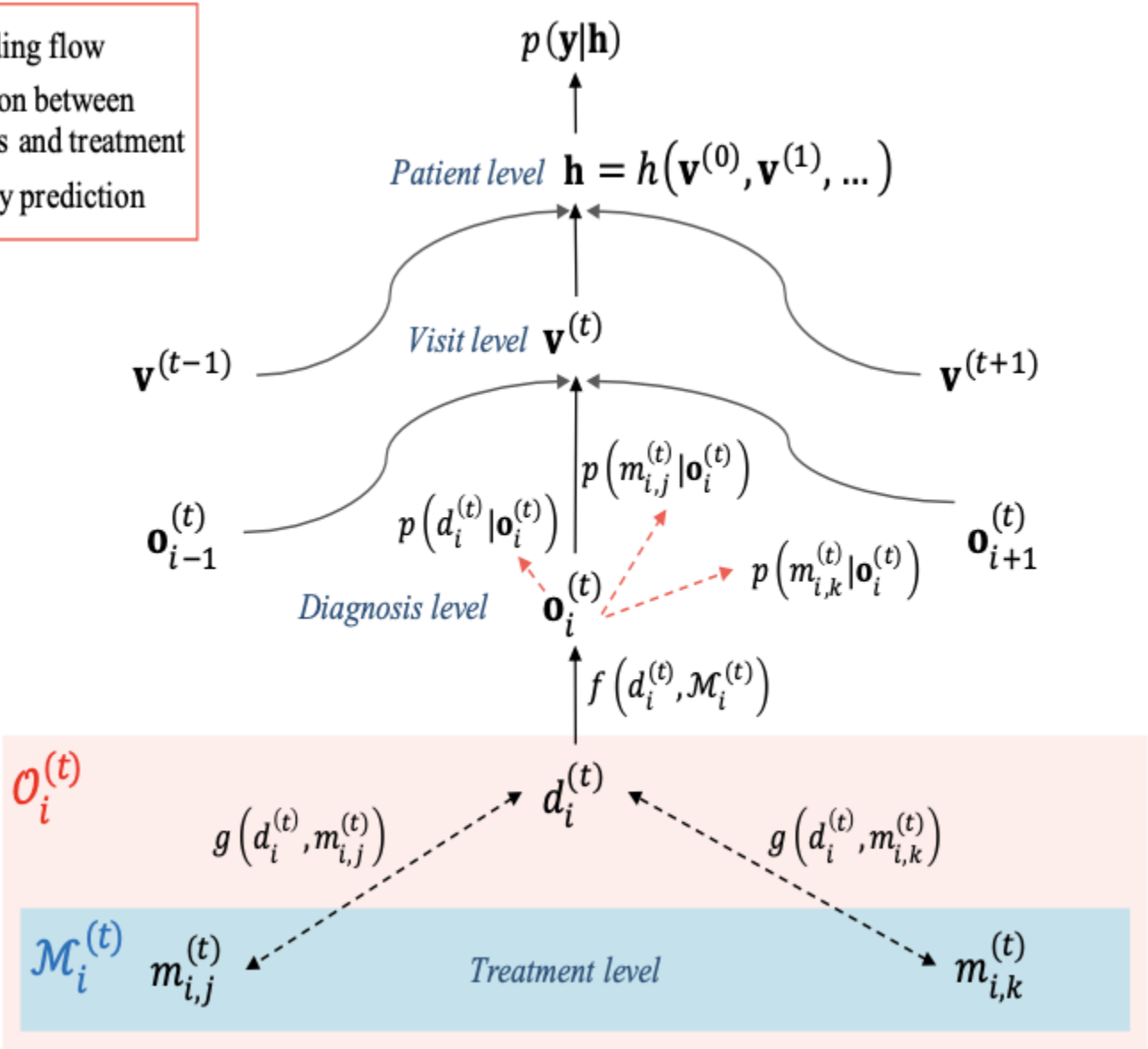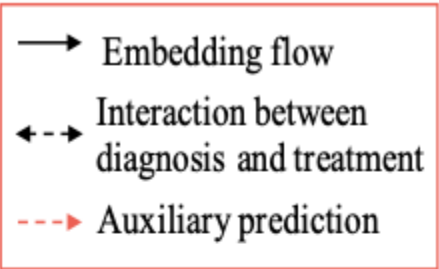# What is the key difference between Med2Vec and MiME?

MiME is another embedding method that leverages the inherent multilevel structure of EHR data and, in particular, the encoded relationships among medical code.

MiME algorithm is a direct extension of Med2Vec to the next level. In Med2Vec we have a two level structure - visit level, code level. Within a visit in Med2Vec, we assume all the event happened together, there's no structure.

In reality, there is actually another level of dependencies. The diagnosis level will be a set of symptoms or- then we have a treatment for those diagnosis. So there's this dependency from diagnosis to the corresponding medication or treatment, and there may be multiple diagnosis, and they will have different medication associated with them.

MiME is a neural network architecture that captures this hierarchical structures.

Their overall objective is to learn robust embedding vectors even when the EHR dataset is small.

Prediction Model using MIME

> **References for above**
> Lecture Slides
> Chapter
> https://towardsdatascience.com/neural-network-embeddings-explained-4d028e6f0526
> https://machinelearningmastery.com/what-are-word-embeddings/