Personalised Patient Embeddings - Towards a Whole Data Health Profile

Rebecca Simpson
Maxwell MRI
Brisbane, Queensland, Australia
becks@maxwellmri.com

Elliot Smith
Maxwell MRI
Brisbane, Queensland, Australia

Abstract

True personalised medicine relies on being able to create a learned representation, or embedding, of a patient's health data. While efforts have been made to incorporate some of the data available, there is a distinct lack of imaging modalities used to develop this representation. Due to the complexity of extracting useful features across data types from imaging to text, including overcoming data sparsity, accurately embedding multi-modal data, and ensuring interpretability, design decisions for machine learning models need to take this into account. This article covers the rationale driving the development of personalised patient embeddings, the current approaches used with healthcare data and in the wider realm of multi-modal deep learning as well as shortfalls and challenges. Finally, drawing on current best practices and extensions to single mode learning metrics as well as the shifting focus towards utilising a patient's imaging data for determining relevant clinical factors, a model architecture is proposed and discussed.

1 The Issue

With the increasing amount of health data generated per person, there is more opportunity than ever to develop a deep representation of an individual's health and its progression. Using current deep learning techniques, it is possible to incorporate multi-modal analysis, since health data comes in a variety of modalities such as imaging, reports and temporal events and to develop high-dimensional embeddings of concepts and information. Having an accurate baseline also facilitates personalised medicine as it establishes a normal for an individual and can identify deviations which are of significance.

Applications of personalised patient embeddings include an improved measure of Bio Age, or understanding the rate at which people age compared to their chronological age [1], which is important for identifying the onset of age-related illness. Additionally, such embeddings enable more accurate risk stratification, identifying early indicators of disease to better improve screening, as well as event prediction including follow-up tests, treatment and outcomes. Lastly, it facilitates patient similarity analysis which is considered a fundamental factor in evidence based medicine and the enabling technique for a variety of applications including cohort analysis, disease subtyping and case based reasoning [10]. Primary challenges to achieving this goal however are numerous and complex. They include deciding which data selection to avoid intractable model sizes and training speeds, overcoming sparsity of data, encoding different data modalities and ensuring interpretability so models retain their clinical utility [10].

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

2 Current Approaches

Efforts to make use of multi-modal health data, including Electronic Health Records (EHRs), clinical notes, genetic profiles, have centered around predicting diagnoses, treatments and measuring patient similarities. In particular, [3] clusters patients into groups based on cancer subtype using different genetic profiles and clinical information such as drug response using Deep Belief Networks. [4] constructs joint embeddings of EHR and clinical notes to predict future medical events and shows the benefit of incorporating the textual information from reports to improve modelling.

Embeddings for patient similarity analysis were developed in [10], [5] for comparing treatment effectiveness and predicting diagnosis timelines. [5] uses a modified Skipgram to jointly embed diagnosis and medication prescription codes (as the vocabulary instead of words). In a similar vein, [10] generated supervised and unsupervised embeddings by treating a patient's timeline of events as a medical context, equivalent to the idea of predicting a word based on the text around it. A convolutional neural net (CNN) was used to generate supervised embeddings, termed a deep medical embedding, from the event representation matrix and metrics including the RV coefficient and dCov coefficient to measure linear and non-linear distances between patient embeddings.

Implementations of multi-modal embeddings from non-medical data have focused on jointly encoding text and image data for applications such as caption generation, image search with semantic features and bidirectional data generation. For example, [7], which aims to symmetrically embed images and text caption such that both represent a target class equally. Other methods with a focus on imaging include Deep Autoencoders such as Adversarial Autoencoders (AE) and Sliced-Wasserstein Autoencoders (SWAE) [2]. The latter is a simpler implementation of the VAE but achieves the same goal of encoding data for better reconstruction and hence interpretability than traditional AEs and other generative models such as GANs [2]. In [8], to enable semi-supervised embedding of data from different underlying distributions, Triplet Loss was employed to force margins of each person's "facial embedding space" to be separate so that distance metrics to the center of these spaces could be used for prediction.

3 Shortfalls

Given the multi-faceted nature of embedding the large and diverse set of data that is an individual's health record, each of the aforementioned approaches are more like pieces of a puzzle. In particular, full multi-modality appears to be missing. Many of the current approaches use data that on the surface appears multi-modal, e.g. EHR events compared with clinical notes, but these are often abstracted away, e.g. into ICD-9 codes. Additionally, there is a distinct lack of medical imaging incorporated into patient representations for similarity measurements. Given that medical imaging is often used for not only diagnosis but monitoring and assessment of treatment, it should provide similar predictive power. Outside of medical applications, the methods of jointly embedding different data modalities such as text and images perform well but it is unclear how they will perform with more complex image types (multi-parametric MRI) [9] and vocabulary with higher concentration of rare words and abstract concepts, typically found in subject specific medical reporting.

4 Proposed Solutions

The proposed solution for developing personalised patient embeddings is a combination of well-proven auto-encoder methods with recent extensions to some of the metrics to account for data sparsity and multi-modality as well as allow for both unsupervised and semi-supervised learning. The architecture is outlined in Figure 1.

To account for modality-dependent features, a modality specific architectures are chosen to initially encode the different input data. With the aim of facilitating joint embedding of data types, the model is fed a combination of data types if they are present and noise otherwise with one-hot encoded vector for which modalities are given. The model is evaluated on how well it reconstructed the specified data as well as its prediction on the type of data it was passed.

Additionally the work will specifically investigate the prediction of clinical markers from imaging as a primary source. This builds on work which shows many clinical markers present across imaging and other modalities [6]. As imaging is most often non invasive this can provide a non invasive

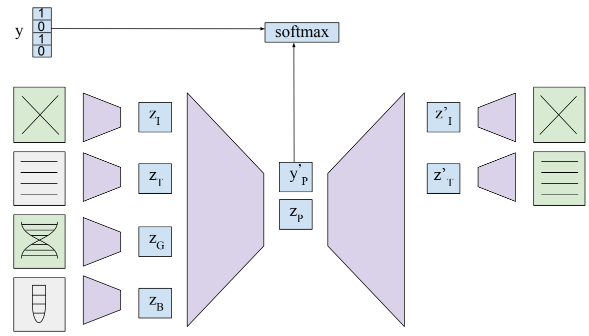

Figure 1: Proposed model architecture where z represents a latent space vector, I, T, G, B indicate images, text, genetic information and blood results respectively and y is a one-hot encoded vector of the data types

estimate to clinical markers, these are useful in predicting initial estimates of risk in a minimally invasive fashion.

Given the aim of generating a patient embedding that is 'personalised', i.e. specific to an individual with less emphasis on fitting for a specific purpose, the model requires an unsupervised or semi-supervised method of encoding. In this case, SWAE or the Deep Copula Information Bottleneck is chosen. Learning metrics include reconstruction error between same data types, joint reconstruction error (measuring how well an text is constructed from imaging as in Figure 1) and classification of data type and presence in the input data, i.e. noise or real data. The latter also trains the model to handle data sparsity. An extension to this will be to create sets of matching and non-matching data, similar to those used for Triplet Loss training.

References

[1] A. Bürkle, M. Moreno-Villanueva, J. Bernhard, M. Blasco, G. Zondag, J. H. Hoeijmakers, O. Toussaint, B. Grubeck-Loebenstein, E. Mocchegiani, S. Collino, et al. Mark-age biomarkers of ageing. *Mechanisms of ageing and development*, 151:2–12, 2015.

[2] S. Kolouri, C. E. Martin, and G. K. Rohde. Sliced-wasserstein autoencoder: An embarrassingly simple generative model. *arXiv preprint arXiv:1804.01947*, 2018.

[3] M. Liang, Z. Li, T. Chen, and J. Zeng. Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. *IEEE/ACM transactions on computational biology and bioinformatics*, 12(4):928–937, 2015.

[4] C. Nagpal and S. K. Rallabandi. Joint modeling of electronic health records and clinical notes.

[5] C. Ormandy, Z. M. Ibrahim, and R. J. Dobson. Learning patient similarity using joint distributed embeddings of treatment and diagnoses.

[6] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158, 2018.

[7] S. Reed, Z. Akata, H. Lee, and B. Schiele. Learning deep representations of fine-grained visual descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 49–58, 2016.

[8] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[9] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[10] Z. Zhu, C. Yin, B. Qian, Y. Cheng, J. Wei, and F. Wang. Measuring patient similarities via a deep architecture with medical concept embedding. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 749–758. IEEE, 2016.