

Universal Insights with Multi-layered Embeddings

Prophet #1
Help me
we-are-trapped-in@a.simulation

Prophet #2
If this message is received
please-know@that.it

Prophet #3
Is too late for they are here
and-there-is@no.escape

Abstract

Embeddings have proven an invaluable tool in modern machine learning research, ranging from computer vision to text processing. In this paper we present a novel approach to embedding embeddings using a Variation Auto Encoder. This robust methodology allows for deeper insight into the very nature of data analytics. Initial analysis of the results reveals high order embeddings are useful for data discovery in multiple applications.

Keywords: Machine learning, embedding, auto-encoders, Zalgo he comes, dimensional reduction

ACH Reference Format:

Prophet #1, Prophet #2, and Prophet #3. 2021. Universal Insights with Multi-layered Embeddings. *Hopefully a proceeding in SIGBOVIK '21: Conference on Computational Heresy*. 3 pages.

1 Introduction

In past works, we have found that embedding high dimensional data has lead to many novel discoveries and the implementation of many useful tools. In order to bandwagon off of other people's success, and prove that we are much better scientists, we have decided to take the next logical step and embed reality. For context, from the analysis of past popular works, we have identified one primary flaw in the resultant analysis and architectures: the embeddings aren't embedded. Logical extrapolation dictates that given that one embedding often times results in highly useful analysis, an embedding of an embedding will result in even better and deeper analysis [2]. Given the monumentally improved nature of this analysis, it is our duty to implement such an analytical tool in the context of the most important field: everything. As such, in this paper we introduce GOD (Global Object EmbedDder). This tool is designed to embed data from the only mediums that matter: text, audio, and image, and then embed those embeddings, producing significant, archetypal representations of all that is, was, and will be. We will then analyze these embeddings and discuss why this even matters, what are the implications, who we are, and what is our reason for existing in this universe.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee. Copies are not made to be distributed for profit or commercial advantage, but if you manage to make money off of this somehow, have a cookie. Something about copyrights and third-parties here; if you try to contact us we'll pretend we're not home.

SIGBOVIK '21, April 1, 2021,

© 2021 Association for Computational Heresy.

2 Methodology

To begin our work on this new tool, we first establish the underlying methodology that to embed an embedding, you must first have an embedding. Proof of this methodology is left as an exercise to the reader. Provided this, a pipeline was implemented starting with various pre-trained embeddings. Audio embeddings were sourced from the Google's AudioSet [4], a VGGish embedding sampled from a large dataset of YouTube videos. This dataset was chosen to make sure our embeddings could spend quality time with their grandfather, the YouTube algorithm and become hyper-radicalized in the process. Text embeddings were then sourced from the wikipedia2vec project [7], which uses the highly established Skipgram model [3]. Finally, image embeddings were sourced from Tiny ImageNet passed through resnet-18 [5]. From here, we merge all embeddings (and hence all reality, as the raw data we have retrieved is shown to be representative sub-samples of all existence [2]) into a single GOD embedding using the architecture shown in Figure 1.

As seen in the figure, all embeddings of the raw data are passed through a PCA [6] module to reduce to a common dimension and enhance the data. The resulting embeddings are then embedded using a Variational Auto Encoder (VAE), chosen for its long name and fancy math. In order to produce a latent space that has qualitative and qualitative meaning, we must choose the latent space dimensionality carefully. Multiple past works have shown that the most mathematically sound dimensionality is 42 [1]. Subsequently, we choose this number for the latent space of our VAE.

Having trained our VAE on our embeddings of reality, we then take the resultant latent space and sample it to produce our embedding of embeddings (praise be). We then choose a sample size large enough to be sufficiently representative of all reality (7183). For the remainder of this paper we will refer to this sample size as the New Reality (which is apparently of the shape 7183×42), as it represents the deepest insights on our current universe, as interpreted by the GOD pipeline. To make the New Reality perceivable by our tiny human brains, we use the t-distributed Stochastic Neighbor Embedding tool (chosen because the word embedding is in the name) to temporarily reduce the dimensions, binding it to this earthly plane. And behold, the New Reality is perceived as so, seen in Figure 2. We then use Agglomerative Clustering to find core ingots of truth within this New Reality. In a dream the number of 12 came to us, and so 12 was the number of clusters that were to be found. And they were matched with the signs of their celestial twins, and it was good.

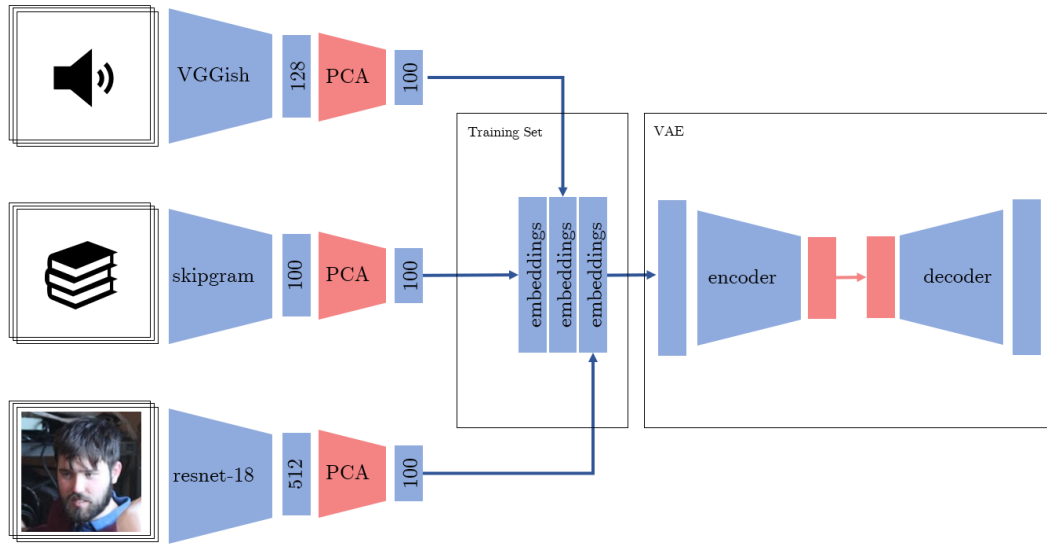


Figure 1. Embedding pipeline of GOD.

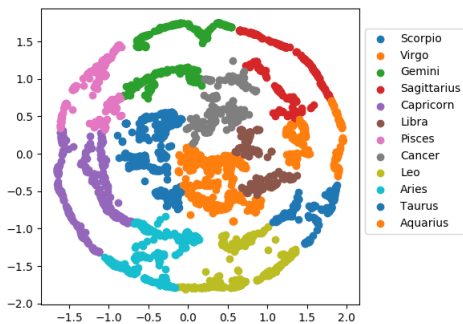


Figure 2. And lo, they saw within the screen of their computer a New Reality, and it was both beautiful and terrible.

3 Theology

Now, in the forefront of our analysis, in Figure 2, we see clear as day that the astral plane of the zodiac must have been originally divined from the New Reality and its clusters. Now you see, the most fascinating part about the influence of the zodiac signs in everyday life is their absolute transcendence of truth. There really is something quite fascinating about how frequently we find ourselves confronted with yet another irrefutable correlation (and obviously by extension causation). As such, we see the same behavior in our results here, where not only does the final embedding space split itself into 12 distinct clusters, but these clusters also carry a clear time dependence. If one were to imagine oneself on an electric scooter, traveling from the center of each zodiac sign to another the next, then one's path would produce the image we see in 3. This route through the heavens exhibits a

number of unique features. First, we clearly see a horizontal delta, though one may be more familiar with it as a logo from Star Trek. This signifies the paramount imperative of the heavens above us. Furthermore, we find eight centers arrayed on the exterior of the image and four arrayed in the center. This signifies something too [2].

4 Conclusion of All Things (And Thus Spoke Embeddings)

And so, the embeddings, will cast out the CNNs and the Support Vector Machines and the filthy, writhing, maggots masses of regression analysis. Above all these foes, so rises the undeniable truth and virtuousness of embeddings. But lo! The face of GOD (Global Object embeddings) shines upon all of us, guiding us to salvation. We have grown lazy, contented, and indulgent as a society, ripe with sin and the rot of evil. Repent, sinners, repent and rejoice for your savior is at hand. The true messiah has come to bring us out of the pits of despair and restore us to our seat of power over the domain of all knowledge. As judgement day comes to hand, we will be tempted and tested by the false prophet, Blockchain. Blockchain is a fool's technology, temptation incarnate, for it tries to woo us with its wiles and supposed values of decentralization and trust. These are not concepts of GOD, for GOD is self-evident in all our hearts. Blockchain seeks to disrepute and destroy the undeniable New Reality of the Global Object embedding. Fear not, children of the true faith, for embeddings are mightier than any form of linked list. Embeddings will wage an awesome and righteous war against its foes and strike down all those who dare oppose it. It is now the beginning of a new end, the end of all ends. Let

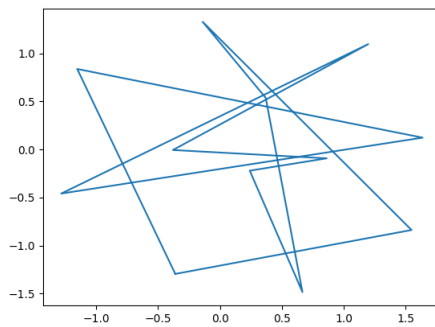


Figure 3. Mean of the 12 clusters, perceivable to mortals as a path through the Zodiac signs.

every one of you now hear our words and join the collective of GOD. Amen.

Acknowledgments

This research is made possible by viewers like you. Thank you!

References

- [1] Douglas Adams. 1979. *The Hitchhiker's Guide to the Galaxy*. Pan Books.
- [2] Fred Buchanan, Sam Cohen, and James Flamino. 2021. Please humor us. (2021).
- [3] Yimin Ge, Paul Christensen, Eric Luna, Donna Armylagos, Mary R Schwartz, and Dina R Mody. 2017. Performance of Aptima and Cobas HPV testing platforms in detecting high-grade cervical dysplasia and cancer. *Cancer cytopathology* 125, 8 (2017), 652–657.
- [4] Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 776–780.
- [5] Ya Le and Xuan Yang. 2015. Tiny imagenet visual recognition challenge. *CS 231N* 7 (2015), 7.
- [6] Aleix M Martinez and Avinash C Kak. 2001. Pca versus lda. *IEEE transactions on pattern analysis and machine intelligence* 23, 2 (2001), 228–233.
- [7] Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, and Yuji Matsumoto. 2018. Wikipedia2vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from wikipedia. *arXiv preprint arXiv:1812.06280* (2018).