question @463 117 views

Actions

Midterm Prep: 6. Embeddings

Howdy everyone,

In preparation for the midterm, I've put together some notes on the various topics, drawing from @416 and the practice midterm.

Going through these notes and completing the practice midterm & quiz has definitely boosted my confidence! Here, I'm happy to share these notes with you all.

Let's ace the midterm together!

Feel free to contribute and add your own insights to these notes as well.

Best regards,
Your somewhat helpful AI bot, Darin

exam

Edit good question | 1 Updated 3 days ago by Darin Zhen

the students' answer, *where students collectively construct a single answer*

Actions

Embeddings
Embeddings in the context of deep learning refer to the process of representing categorical variables or high-dimensional data in a lower-dimensional space, typically as continuous vectors. These embeddings capture meaningful relationships and similarities between different categories or data points, making them more suitable for processing by neural networks.

Embeddings in deep learning offer a powerful framework for extracting and representing complex healthcare data, leading to improved patient care, disease management, and medical research.

Here's an overview of embeddings in deep learning for healthcare:
Embeddings in deep learning for healthcare involve representing data in a lower-dimensional space to capture meaningful relationships and patterns. This technique is widely used across various healthcare applications, including medical image analysis, electronic health record (EHR) processing, and clinical decision support systems.

In medical image analysis, embeddings are used to represent images such as X-rays, MRIs, and CT scans in a more compact and informative manner, facilitating tasks like disease detection, segmentation, and classification.

In EHR processing, embeddings are utilized to encode patient records, diagnoses, medications, and other clinical data into numerical vectors. This enables efficient information retrieval, patient similarity analysis, and predictive modeling for personalized healthcare interventions.

Embeddings also play a crucial role in clinical decision support systems by encoding medical knowledge and expertise into numerical representations. This allows dee learning models to learn from large-scale healthcare data and assist healthcare professionals in diagnosis, treatment planning, and patient management.

Embeddings are widely used in natural language processing (NLP), recommender systems, and other domains where data is categorical or high-dimensional.

Here's an overview of embeddings in deep learning:

1. Word Embeddings:

- Word embeddings are one of the most well-known types of embeddings in NLP. They represent words as dense vectors in a continuous vector space, where each dimension of the vector captures a different aspect of the word's meaning or context.
- Word embeddings are learned from large text corpora using techniques such as Word2Vec, GloVe (Global Vectors for Word Representation), and FastText. These m capture semantic relationships between words based on their co-occurrence patterns in the text.

2. Sentence and Document Embeddings:

- Similar to word embeddings, sentence and document embeddings represent entire sentences or documents as fixed-size vectors in a continuous space.
- Techniques such as Doc2Vec (Paragraph Vector) and Universal Sentence Encoder learn embeddings for sentences and documents by training neural networks to p the context of a given sentence or document.

3. Entity Embeddings:

- Entity embeddings are used to represent categorical variables in machine learning models. They map discrete categorical variables, such as user IDs, product IDs, categories, to continuous vectors in a lower-dimensional space.
- Entity embeddings are typically learned as part of the training process of a neural network, where the embeddings are optimized to capture meaningful relationship between different categories or entities.

4. Image Embeddings:

- Image embeddings represent images as low-dimensional vectors in a continuous space, capturing visual features and semantics of the images.
- Techniques such as Convolutional Neural Networks (CNNs) and pre-trained models like VGG, ResNet, and Inception are used to extract image embeddings by fee images through the neural network and extracting features from intermediate layers.

5. Graph Embeddings:

- Graph embeddings represent nodes or entire graphs as continuous vectors in a low-dimensional space, capturing structural and relational information within the gr
- Techniques such as node2vec, GraphSAGE, and Graph Attention Networks (GATs) learn embeddings for nodes and graphs by leveraging the graph structure and n attributes.

6. Training Embeddings:

- Embeddings can be trained from scratch as part of the training process of a neural network or learned from pre-trained models using transfer learning techniques.
- In transfer learning, pre-trained embeddings learned from large datasets are fine-tuned on a specific task or domain-specific dataset to capture task-specific featu relationships.

Embeddings play a crucial role in deep learning by enabling neural networks to effectively process and understand categorical or high-dimensional data, leading to im performance in various tasks such as NLP, recommendation systems, and image processing.

Edit thanks! 3 Updated 3 days ago by Darin Zhen

followup discussions *for lingering questions and comments*

Resolved Unresolved @463_f1

Darin Zhen 2 days ago

Old Book: Embedding

• Embedding methods like Word2Vec, Med2Vec, and MiME learn vector representations (embeddings) of medical concepts like diagnoses, medications, and procedures based on their co-occurrence patterns in electronic health records.
• Word2Vec assumes a global ordering of codes, while Med2Vec captures hierarchical structure of visits and codes within a visit. MiME further models diagnosis-treatment relationships.
• These embedding vectors capture semantic relationships between medical concepts, so related concepts have similar vector representations. This results in better features compared to one-hot encodings.
• Visualization techniques like t-SNE can project the high-dimensional embedding vectors into 2D space for qualitative evaluation. Similar medical concepts cluster together.
• Quantitatively, using embedding vectors as features improves performance on prediction tasks like heart failure onset compared to raw encodings. The improvement is especially significant for small datasets.
• Auxiliary prediction tasks in MiME like predicting diagnosis and treatment codes help learn robust embeddings from limited data.
• Leveraging inherent structure of EHR data with embedding methods leads to more meaningful medical concept representations and better downstream predictive modeling performance.

helpful! 1

Reply to this followup discussion

Resolved Unresolved @463_f2

Darin Zhen 2 days ago

Lecture 05 embedding

- Word2Vec is a neural network model that produces vector representations of words based on the context they appear in. It can be applied to medical text data to generate vector representations of medical concepts like diagnoses, medications, procedures.
- t-SNE (t-distributed stochastic neighbor embedding) is a technique for visualizing high-dimensional data by reducing it to 2D while preserving local structure. It has been used to visualize disease phenotypes and sleep stages based on EHR data.
- Med2Vec is a model that learns representations of medical concepts by exploiting the two-layer structure of longitudinal EHR data - sequential visits and co-occurring codes within a visit. The learned representations capture medically meaningful similarities.

- MiME (Multilevel Medical Embedding) builds on Med2Vec by also incorporating auxiliary prediction tasks and modeling diagnosis-treatment relations. It learns a patient representation that can effectively predict outcomes like heart failure.
- Embedding methods like Word2Vec, Med2Vec, and MiME can learn low-dimensional representations of medical concepts and patients from EHR data. These representations capture meaningful similarities and can boost predictive performance compared to using raw input data.

helpful! | 1

Reply to this followup discussion

Start a new followup discussion

Compose a new followup discussion