Multi-Modal Textbook Answering

LITERATURE SURVEY

Machine Intelligence

BACHELOR OF TECHNOLOGY- V Sem CSE

Department of Computer Science & Engineering

SUBMITTED BY

Batch No:- 4

Student name 1:Shashank Varma SRN:PES1UG20CS395
Student name 2:Shreyas S SRN:PES1UG20CS408
Student name 3:Shishira Bhat O SRN:PES1UG20CS397

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)

100 Feet Ring Road, BSK III Stage, Bengaluru-560085

1) **SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS by Thomas N. Kipf - University of Amsterdam,Max Welling- University of Amsterdam Canadian Institute for Advanced Research (CIFAR)**

We present a scalable approach for semi-supervised learning on graph-structured data that is based on an efficient variant of convolutional neural networks which operate directly on graphs. We motivate the choice of our convolutional architecture via a localized first-order approximation of spectral graph convolutions. Our model scales linearly in the number of graph edges and learns hidden layer representations that encode both local graph structure and features of nodes. In a number of experiments on citation networks and on a knowledge graph dataset we demonstrate that our approach outperforms related methods by a significant margin.

2) **UniK-QA: Unified Representations of Structured and Unstructured Knowledge for Open-Domain Question Answering by Barlas Oguz, Xilun Chen, Vladimir Karpukhin, Stan Peshterliev, Dmytro Okhonko, Michael Schlichtkrull, Sonal Gupta, Yashar Mehdad, Scott Yih**

We study open-domain question answering with structured, unstructured and semi-structured knowledge sources, including text, tables, lists and knowledge bases. Departing from prior work, we propose a unifying approach that homogenizes all sources by reducing them to text and applies the retriever-reader model which has so far been limited to text sources only. Our approach greatly improves the results on knowledge-base QA tasks by 11 points, compared to latest graph-based methods. More importantly, we demonstrate that our unified knowledge (UniK-QA) model is a simple and yet effective way to combine heterogeneous sources of knowledge, advancing the state-of-the-art results on two popular question answering benchmarks, NaturalQuestions and WebQuestions, by 3.5 and 2.6 points, respectively.

3) **A Comprehensive Survey on Graph Neural Networks by Zonghan Wu, Shirui Pan, Member, IEEE, Fengwen Chen, Guodong Long, Chengqi Zhang, Senior Member, IEEE, Philip S. Yu, Fellow, IEEE**

Deep learning has revolutionized many machine learning tasks in recent years, ranging from image classification and video processing to speech recognition and natural language understanding. The data in these tasks are typically represented in the Euclidean space. However, there is an increasing number of applications where data are generated from non-Euclidean domains and are represented as graphs with complex relationships and interdependency between objects. The complexity of graph data has imposed significant challenges on existing machine learning algorithms. Recently, many studies on extending deep learning approaches for graph data have emerged. In this survey, we provide a comprehensive overview of graph neural networks (GNNs) in data mining and machine learning fields. We propose a new taxonomy to divide the state-of-the-art graph neural networks into four categories, namely recurrent graph neural networks, convolutional graph neural networks, graph autoencoders, and spatial-temporal graph neural networks. We further discuss the applications

of graph neural networks across various domains and summarize the open source codes, benchmark data sets, and model evaluation of graph neural networks. Finally, we propose potential research directions in this rapidly growing field.

4) **M-GCN: A Multimodal Graph Convolutional Network to Integrate Functional and Structural Connectomics Data to Predict Multidimensional Phenotypic Characterizations Niharika S. D'Souza , Mary Beth Nebel , Deana Crocetti , Joshua Robinson , Stewart Mostofsky , Archana Venkataraman**

We propose a multimodal graph convolutional network (M-GCN) that integrates resting state fMRI connectivity and diffusion tensor imaging tractography to predict phenotypic measures. Our specialized M-GCN filters act topologically on the functional connectivity matrices, as guided by the subject-wise structural connectomes. The inclusion of structural information also acts as a regularizer and helps extract rich data embeddings that are predictive of clinical outcomes. We validate our framework on 275 healthy individuals from the Human Connectome Project and 57 individuals diagnosed with Autism Spectrum Disorder from an in-house data to predict cognitive measures and behavioral deficits respectively. We demonstrate that the M-GCN outperforms several state-of-the-art baselines in a five-fold cross validated setting and extracts predictive biomarkers from both healthy and autistic populations. Our framework thus provides the representational flexibility to exploit the complementary nature of structure and function and map this information to phenotypic measures in the presence of limited training data.

5) Are You Smarter Than A Sixth Grader? Textbook Question Answering for Multimodal Machine Comprehension by Aniruddha Kembhavi,Minjoon Seo, Dustin Schwenk,Jonghyun Choi,Ali Farhadi,Hannaneh Hajishirzi,Allen Institute for Artificial Intelligence, University of Washington

We introduce the task of Multi-Modal Machine Comprehension (M3C), which aims at answering multimodal questions given a context of text, diagrams and images. We present the Textbook Question Answering (TQA) dataset that includes 1,076 lessons and 26,260 multi-modal questions, taken from middle school science curricula. Our analysis shows that a significant portion of questions require complex parsing of the text and the diagrams and reasoning, indicating that our dataset is more complex compared to previous machine comprehension and visual question answering datasets. We extend state-of-the-art methods for textual machine comprehension and visual question answering to the TQA dataset. Our experiments show that ∗The majority of the work was done while the author was interning at the Allen Institute for Artificial Intelligence these models do not perform well on TQA. The presented dataset opens new challenges for research in question answering and reasoning across multiple modalities.

6) **Textbook Question Answering with Multi-modal Context Graph Understanding and Self-supervised Open-set Comprehension by Daesik Kim - Seoul National University, Seonhoon Kim - V.DO Inc.,Nojun Kwak - Search&Clova, Naver Corp.**

In this work, we introduce a novel algorithm for solving the textbook question answering (TQA) task which describes more realistic QA problems compared to other recent tasks. We mainly focus on two related issues with analysis of the TQA dataset. First, solving the TQA problems requires comprehending multimodal contexts in complicated input data. To tackle this issue of extracting knowledge features from long text lessons and merging them with visual features, we establish a context graph from texts and images, and propose a new module f-GCN based on graph convolutional networks (GCN). Second, scientific terms are not spread over the chapters and subjects are split in the TQA dataset. To overcome this so-called 'out-of-domain' issue, before learning QA problems, we introduce a novel self-supervised open-set learning process without any annotations. The experimental results show that our model significantly outperforms prior state-of-the-art methods. Moreover, ablation studies validate that both methods of incorporating f-GCN for extracting knowledge from multi-modal contexts and our newly proposed self-supervised learning process are effective for TQA problems.

7) **XTQA: Span-Level Explanations of the Textbook Question Answering by Jie Ma, Jun Liu, Junjun Li2, Qinghua Zheng,, Qingyu Yin, Jianlong Zhou, Yi Huang**

Textbook Question Answering (TQA) is a task in which one should answer a diagram/non-diagram question given a large multi-modal context consisting of abundant essays and diagrams. We argue that the explainability of this task should place students as a key aspect to be considered. To address this issue, we devise a novel architecture towards span-level eXplanations of the TQA (XTQA). It can provide not only the answers but also the span-level evidence to choose them for students based on our proposed coarse-to-fine grained algorithm. The algorithm first coarsely chooses top M paragraphs relevant to questions using the TF-IDF method, and then chooses top K evidence spans finely from all candidate spans within these paragraphs by computing the information gain of each span to questions. Experimental results show that our method significantly improves the state-of-the-art performance compared with baselines.

8) **Graph Convolutional Networks for Text Classification by Liang Yao, Chengsheng Mao, Yuan Luo**

Text classification is an important and classical problem in natural language processing. There have been a number of studies that applied convolutional neural networks (convolution on regular grid, e.g., sequence) to classification. However, only a limited number of studies have explored the more flexible graph convolutional neural networks (convolution on non-grid, e.g., arbitrary graph) for the task. In this work, we propose to use graph convolutional networks for text classification. We build a single text graph for a corpus based on word co-occurrence and document word relations, then learn a Text Graph Convolutional Network (Text GCN) for the corpus. Our Text GCN is initialized with one-hot representation for word and document, it then jointly learns the embeddings for both words and documents, as supervised by the known class labels for documents. Our experimental results on multiple benchmark datasets demonstrate that a vanilla Text GCN without any external word embeddings or knowledge outperforms state-of-the-art methods for text classification. On the other hand, Text GCN also learns

predictive word and document embeddings. In addition, experimental results show that the improvement of Text GCN over state-of-the-art comparison methods become more prominent as we lower the percentage of training data, suggesting the robustness of Text GCN to less training data in text classification.

9) **Fusion-GCN: Multimodal Action Recognition using Graph Convolutional Networks - Michael Duhme, Raphael Memmesheimer, Dietrich Paulus**

In this paper, we present Fusion-GCN, an approach for multimodal action recognition using Graph Convolutional Networks (GCNs). Action recognition methods based around GCNs recently yielded state-of-the-art performance for skeleton-based action recognition. With Fusion-GCN, we propose to integrate various sensor data modalities into a graph that is trained using a GCN model for multi-modal action recognition. Additional sensor measurements are incorporated into the graph representation, either on a channel dimension (introducing additional node attributes) or spatial dimension (introducing new nodes). Fusion-GCN was evaluated on two publicly available datasets, the UTD-MHAD- and MMACT datasets, and demonstrates flexible fusion of RGB sequences, inertial measurements and skeleton sequences. Our approach gets comparable results on the UTD-MHAD dataset and improves the baseline on the large-scale MMACT dataset by a significant margin of up to 12.37% (F1-Measure) with the fusion of skeleton estimates and accelerometer measurements.