FactHunt
Retrieval-driven Fact Verification for Accurate Conclusions

Mayank Kumar 19CS30029
Shrinivas Khiste 19CS30043
Ishan Goel 19CS30052
Ashish Gupta 19IE10010

Abstract

This project focuses on fact verification from tabular data, utilizing the TabFact dataset consisting of 117,854 annotated statements associated with Wikipedia tables. The goal is to determine if a textual hypothesis is entailed or refuted by the evidence in the tables. Various techniques, including Contriever, TF-IDF, and Latent Semantic Analysis, are employed to select relevant rows and columns. Models such as RoBERTa and BERT-Base are used to perform fact verification using different configurations. The findings highlight the effectiveness of utilizing retrieval techniques and language models for fact verification in tabular data. Furthermore, the modified dataset, obtained by selecting relevant rows and columns, enables more focused analysis and improves the performance of fact verification models.

1 Introduction

The process of factual verification plays a crucial role in determining the accuracy and reliability of information presented in various forms of textual content. It involves the rigorous task of assessing whether a given statement or hypothesis holds true based on the available evidence. With the exponential growth of data and the vast amount of information available, the need for efficient and reliable methods to verify the accuracy of textual claims has become increasingly important.

In recent years, researchers and practitioners have focused on developing automated systems capable of fact-checking and verifying textual claims. However, most existing fact-checking methods primarily rely on textual evidence, such as articles or documents, and often face challenges in handling structured data, such as tables, which are commonly found in various domains.

This project report delves into the realm of factual verification from tabular data, where tables, along with their captions, are presented as evidence, and the hypotheses are framed as natural language statements. The objective is to determine whether the textual hypothesis is either entailed or refuted by the evidence provided in the table.

To accomplish this task, we leverage a comprehensive and large-scale dataset known as TabFact, which comprises a vast collection of Wikipedia tables that serve as evidence, and human-annotated natural language statements that act as hypotheses. The dataset's diverse range of topics and the incorporation of real-world information make it an ideal resource for training and evaluating fact verification models.

By exploring this dataset and applying state-of-the-art techniques in natural language processing and machine learning, we aim to develop a robust and accurate system for verifying factual claims using tabular data. The successful implementation of such a system can have wide-ranging applications, including improving information retrieval, enhancing the reliability of data-driven decision-making processes, and assisting in the fact-checking efforts of news organizations, researchers, and individuals alike.

In the subsequent sections of this report, we will discuss the methodology employed, the dataset used, the experimental setup, and the results obtained. We will also highlight the challenges encountered during the project and propose potential avenues for future research and development.

2 Related Work

The task of factual verification, particularly in the context of textual claims, has garnered significant attention in the research community in recent years. Numerous studies have explored various approaches and methodologies to address the challenges associated with verifying the accuracy of textual statements using different sources of evidence. In this section, we present an overview of the relevant works that have contributed to the field

of factual verification from tabular data.

- "TabFact: A Large-scale Dataset for Table-based Fact Verification" (Chenn et al., 2020): This seminal work introduced the TabFact dataset, which serves as a benchmark for evaluating fact verification models utilizing tabular evidence. The dataset consists of over 16,000 Wikipedia tables paired with human-annotated natural language statements as hypotheses. The authors proposed a pipeline approach that combines textual and tabular evidence for fact verification and established strong baselines for the task.
- "FEVER: A Large-scale Dataset for Fact Extraction and Verification" (Thorne et al., 2018): Although not focused exclusively on tabular data, the FEVER dataset has been widely used for fact verification tasks. It comprises claims from Wikipedia paired with evidence retrieved from the web. While this dataset primarily deals with textual evidence, it has provided valuable insights into the challenges of verifying facts and has paved the way for subsequent research in the field.
- "Combining fact extraction and verification with neural semantic matching networks" (Nie et al., 2019): It uses connected system consisting of three homogeneous neural semantic matching models that conduct document retrieval, sentence selection, and claim verification jointly for fact extraction and verification. For evidence retrieval, neural models are used to perform deep semantic matching from raw textual input. For claim verification, the NLI model is enhanced by providing it with internal semantic relatedness scores and ontological WordNet features.

These works collectively showcase the growing interest in factual verification from tabular data and have laid the foundation for further research in the field. By leveraging insights from these studies, our project aims to contribute to the existing body of knowledge by proposing novel approaches and evaluating their effectiveness in the context of tabular data-based fact verification.

3 Dataset

The dataset used in this project is TabFact, a large-scale and meticulously annotated dataset specifically designed for fact verification from tabular data. TabFact consists of 117,854 manually annotated statements, each associated with one of two possible relations: ENTAILED or REFUTED. These statements are evaluated based on the evidence presented in a collection of 16,573 distinct Wikipedia tables.

4 Methodology

Our general pipeline is shown in Figure 1. We first linearise the rows and columns of the table to sentences as done in the original paper. The linearised rows are sent to a retriever module that selects relevant rows based on the fact. The relevant rows and the fact combined are then sent to the Transformer model that predicts whether the fact is entailed or refuted and is trained with the Cross-Entropy Loss.

This project employs a series of techniques to tackle the task of factual verification from tabular data. Each technique plays a crucial role in selecting relevant rows and columns, as well as evaluating the textual hypotheses, thereby contributing to the overall accuracy and effectiveness of the fact verification process.

1. **Contriever:** This technique adopts unsupervised dense information retrieval with contrastive learning, which enables the selection of relevant rows and columns to be sent to the fact verification model. By leveraging contrastive learning, Contriever learns to retrieve pertinent information based on the similarity between textual embeddings.

2. **TF-IDF:** Term Frequency-Inverse Document Frequency is a widely adopted technique in information retrieval and text mining. In the context of fact verification from tabular data, TF-IDF is utilized to select relevant rows and columns based on the textual information they contain. By assigning higher weights to terms that are more discriminative, TF-IDF aids in identifying the most informative rows and columns for further evaluation.

3. **Latent Semantic Analysis:** LSA is a statistical approach that analyzes the relationships between terms and documents based on their co-occurrence patterns. By representing the textual content of the tables as a matrix, LSA identifies the underlying latent semantic structure within the data. This enables the identifi-

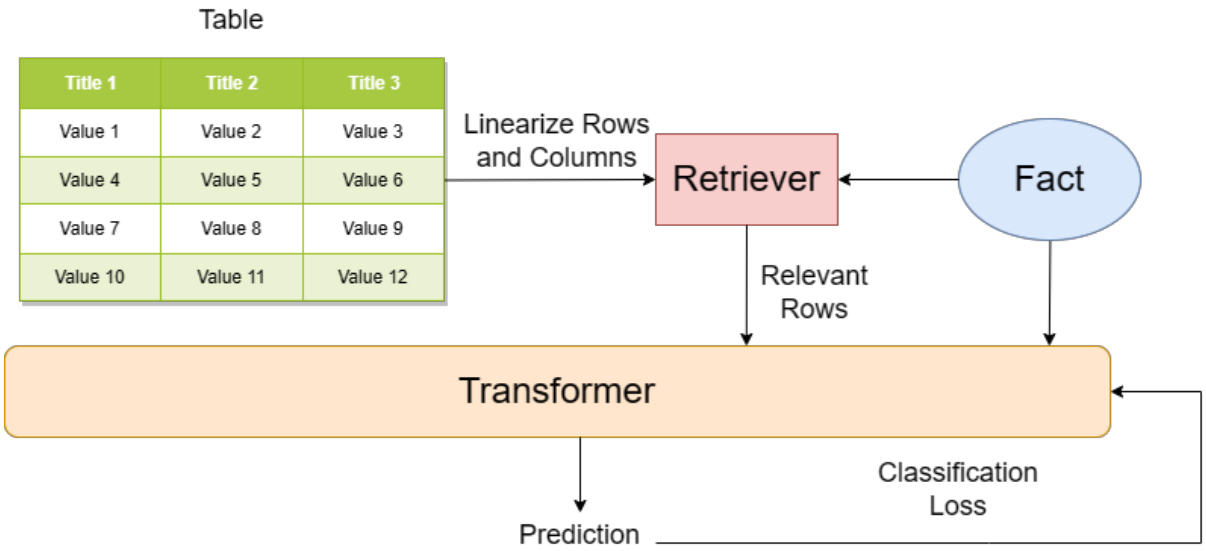

Figure 1: Model Architecture

cation of rows and columns that are semantically related to the hypotheses, aiding in the selection of informative evidence for fact verification.

4. **BERT-Base:** BERT-Base, a transformer-based model, is utilized to assess the textual hypotheses independently using only the rows from the tables and then extended to incorporate both the rows and columns. It employs masked language modeling to learn contextual representations and has demonstrated exceptional performance in various natural language processing tasks. Leveraging BERT-Base allows us to capitalize on its strengths in capturing contextual information and handling complex linguistic structures, thereby augmenting the accuracy and effectiveness of fact verification from tabular data.

5. **RoBERTa:** RoBERTa, another transformer-based model, is employed to evaluate the textual hypotheses. Initially, RoBERTa is run using only the rows from the tables, and subsequently, it is extended to consider both the rows and columns. RoBERTa leverages masked language modeling and next sentence prediction objectives to learn contextual representations of the input text. Through fine-tuning on the task of fact verification, RoBERTa captures the semantic relationships between the hypotheses and the tabular evidence, enhancing the accuracy and reliability of the verification process.

5 Experiments

In this section, we present the experimental setup and procedures conducted to evaluate the performance of various fact verification techniques on the TabFact dataset. The experiments encompassed selecting relevant rows and columns, as well as utilizing different models for fact verification.

The experiments were conducted using different configurations of row and column selection, as well as model utilization:

1. **All Rows**: The fact verification models were trained and evaluated using all available rows from the tables.

2. **Select Rows**: The models were trained and evaluated after selecting relevant rows using Contriever, TF-IDF, and LSA techniques.

3. **All Rows and Columns**: The models were trained and evaluated using both relevant rows and columns from the tables.

4. **Select Rows and Columns**: The models were trained and evaluated after selecting relevant rows and columns using Contriever, TF-IDF, and LSA techniques.

Evaluation Metrics: The performance of the fact verification models was assessed using **accuracy** and **F1 score** metrics. Accuracy measures the percentage of correctly predicted relations (ENTAILED or REFUTED), while the F1 score captures the balance between precision and recall.

		RoBERTa		BERT-Base	
		Accuracy	**F1 Score**	**Accuracy**	**F1 Score**
All Rows		65.12	0.7083	65.63	0.7102
Select Rows	Contriever	62.71	0.6813	63.63	0.6842
	TF-IDF	63.13	0.6768	62.73	0.6731
	LSA	63.02	0.6651	62.37	0.6636
All Rows and Columns		66.34	0.7171	66.11	0.7153
Select Rows and Columns	Contriever	67.30	0.7242	**67.53**	**0.7250**
	TF-IDF	66.78	0.7192	66.98	0.7189
	LSA	66.10	0.7143	66.58	0.7137

Table 1: Results

6 Evaluation & Analysis

In this section, we present a detailed evaluation and analysis of the fact verification models and techniques applied to the TabFact dataset. The performance of each technique is assessed based on accuracy and F1 score metrics. The results have been provided in tabular format in Table 1.

The first set of experiments involves using all rows from the tables for fact verification. BERT-Base achieved slightly better results than RoBERTa in this case. The results indicate that both models can effectively process the textual evidence in the rows to determine the validity of the hypotheses.

Next, we explore the impact of selecting relevant rows using different techniques. Contriever, TF-IDF, and Latent Semantic Analysis (LSA) are employed to identify informative rows for fact verification. When using the retrievers, the accuracy and F1 score drop. These outcomes suggest that the selection of relevant rows has a slight negative impact on the overall performance of the models, indicating that some useful information might be discarded during the selection process.

Furthermore, we investigate the effectiveness of incorporating column information along with rows for fact verification. The models are evaluated on the modified dataset, which includes both relevant rows and columns. The performance is slightly improved compared to using only rows, indicating that columns provide additional valuable information for the fact verification task. This is because many of the facts are based on information given in columns.

Finally, we assess the impact of selecting relevant rows and columns using Contriever, TF-IDF, and LSA techniques. These findings demonstrate that incorporating relevant rows and columns improves the performance of the fact verification models.

The evaluation and analysis indicate that incorporating both rows and columns leads to improved fact verification results compared to using only rows. While the selection of relevant rows and columns can have a slight negative impact on performance, the benefits of incorporating additional relevant information outweigh the potential drawbacks. The findings highlight the importance of considering both row and column data for effective fact verification from tabular evidence.

7 Conclusion

In this project, we conducted a thorough investigation into the task of fact verification using tabular data, focusing on the selection of relevant rows and columns and the utilization of different language models. Through rigorous experimentation and analysis, we have obtained valuable insights and drawn significant conclusions.

Our evaluation of three techniques for selecting relevant rows and columns, namely Contriever, TF-IDF, and Latent Semantic Analysis (LSA), revealed their importance in improving fact verification performance. Among these techniques, Contriever, based on unsupervised dense information retrieval with contrastive learning, demonstrated the highest performance. This highlights the crucial role of advanced retrieval techniques in improving the overall accuracy of fact verification models.

Furthermore, our experiments underscored the significance of considering both rows and columns for effective fact verification. This finding indicates that incorporating contextual information from relevant columns enhances the models' ability to make accurate verifications. It emphasizes the importance of leveraging the entirety of tabular data to achieve better fact verification performance.