****

**SINDH MADRESSATUL ISLAM UNIVERSITY, KARACHI**

**FACULTY OF INFORMATION TECHNOLOGY**

**DEPARTMENT OF SOFTWARE ENGINEERING**

**NATURAL LANGUAGE PROCESSING**

**QUIZ 2: CHART QUESTION ANASWERING (ChartQA)**

**SUBMITTED BY:**

MUHAMMAD SAAD NASEEM (BSE-21F-097)

BSSE 8C (EVENING)

**SUBMITTED TO:**

SYED MUHAMMAD HASSAN ZAIDI

**Conduct a thorough research gap analysis of models and approaches used for Chart Question Answering (ChartQA) with a focus on reasoning between the years 2020 to 2025.**

**Research Overview:**

**1) Model architectures**

**2) Reasoning capabilities**

**3) Chart analysis techniques**

**Gap Identification:**

**1) Limitations in reasoning complexity**

**2) Efficiency**

**3) Generalization across chart types**

**4) Multimodal integration**

**Recent Advancements:**

**1) ChartFormer, Pix2Struct, MatCha, ChartNet**

**2) Discuss innovations in graph-based reasoning, knowledge integration, and transformer-based architectures used in ChartQA tasks**

**Future Directions:**

**1) Enhancing reasoning capabilities (e.g., logical, multi-hop, or numerical reasoning)**

**2) Improving model efficiency for real-time applications**

**3) Handling diverse chart types and multimodal data more effectively**

**4) Exploring lightweight models for broader deployment and scalability**

**Research Gap Analysis of Chart Question Answering (ChartQA) with a Focus on Reasoning (2020–2025)**

This analysis explores the evolution of Chart Question Answering (ChartQA) models from 2020 to 2025, highlighting developments in model architecture, reasoning capabilities, and chart analysis techniques. It also identifies key limitations, recent innovations, and emerging directions for future research.

**Research Overview:**

1. **Model Architectures**

ChartQA has witnessed rapid advancement in architecture design, including:

* **Transformer-Based Models**: Many models are built on top of pre-trained language models (e.g., BERT, T5), repurposed for vision-language tasks.
* **Vision-Language Integration**: Models like **Pix2Struct** extract structured chart data using visual backbones and transformer encoder- decoder pipelines.
* **Graph-Based Reasoning Architectures**: Frameworks such as **MatCha** and **ChartNet** leverage graph neural networks (GNNs) to represent and reason over chart elements (e.g., axes, bars, legends).
* **Multimodal Fusion Approaches**: These approaches integrate OCR-based text extraction, image embeddings, and sometimes external knowledge sources to enhance question answering.

2. **Reasoning Capabilities**

* **Textual and Numerical Reasoning**: Some systems incorporate symbolic computation or program induction (e.g., MatCha) to handle mathematical queries.
* **Spatial and Layout Reasoning**: Essential for interpreting relationships between chart elements; graph-based models are particularly effective here.
* **Multi-Hop Reasoning**: Remains an emerging area; current models make limited attempts at simulating multi-step inference chains.

3**. Chart Analysis Techniques**

* **OCR and Text Extraction**: Tools like Tesseract and Donut are frequently used to retrieve textual data from visual elements.
* **Object Detection and Alignment**: Detection models (e.g., YOLO variants) help identify and align chart components.
* **Graph Construction**: Semantic graphs of chart elements are constructed to facilitate structured reasoning in graph-based QA models.

**Gap Identification**

1. **Limitations in Reasoning Complexity:**

* Many existing models are limited in **multi-step, logical, and abstract reasoning**.
* Numerical reasoning often relies on **shallow or heuristic** methods.
* There is minimal integration of **symbolic logic or formal computation frameworks** for robust inferencing.

1. **Efficiency Constraints**

* Models like **ChartFormer** and **Pix2Struct** require substantial computational resources due to their reliance on large-scale transformer architectures.
* Real-time applications remain impractical due to **slow inference times** and **hardware demands**.
* There is insufficient focus on **model optimization techniques** such as pruning, quantization, or distillation.

1. **Generalization Across Chart Types**

* Current models primarily focus on a narrow set of chart types (e.g., bar, line, pie).
* Performance drops significantly on **unseen, hybrid, or hierarchical charts** like scatter plots, treemaps, or radial graphs.
* **Few-shot and zero-shot generalization** across chart varieties remains underdeveloped.

1. **Multimodal Integration Challenges**

* The integration of **visual, textual, and structural information** is often handled in an ad hoc or loosely coupled manner.
* There is a lack of **dynamic alignment** between the extracted text and the corresponding visual elements.
* **External knowledge bases** and **semantic resources** are rarely incorporated for enriched contextual understanding.

**Recent Advancements**

1. **ChartFormer**

* Employs transformers to encode both text and structural layout.
* Captures sequential chart semantics but struggles with **granular spatial understanding** and **computational cost**.

1. **Pix2Struct**

* Uses vision transformers to convert chart images directly into text sequences.
* Avoids reliance on OCR but demands **extensive training data** and lacks symbolic reasoning capability.

1. **MatCha**

* Represents chart components as graph nodes and uses symbolic program execution for complex reasoning tasks.
* Excels in **multi-hop reasoning** through **graph traversal and execution engines**.

1. **ChartNet**

* Builds structured chart representations and models inter-element relationships via GNNs.
* Introduces **semantic graph construction** as a foundation for enhanced reasoning.

**Key Innovations**

* **Graph-based reasoning**: Allows relational and structural understanding of visual components.
* **Program Induction**: Used in models like MatCha for solving complex numerical tasks.
* **Transformer Architectures**: Enable flexible encoding of multimodal inputs but are often resource-heavy.
* **Knowledge Integration**: Still in its infancy, with limited application of external world knowledge or domain-specific ontologies.

**Future Directions**

1. **Enhancing Reasoning Capabilities**

* Integrate **symbolic logic engines**, **math solvers**, or **rule-based systems** to improve deep reasoning.
* Develop architectures that simulate **multi-hop, causal, and temporal reasoning** across chart elements.
* Explore **hybrid neuro-symbolic systems** to balance data-driven learning and logical inference.

1. **Improving Model Efficiency**

* Apply **knowledge distillation**, **model compression**, and **quantization** techniques to reduce size and speed up inference.
* Develop models optimized for **real-time and edge deployment**.
* Utilize **adaptive computation** to handle simple vs. complex queries differently.

1. **Generalization Across Chart Types**

* Expand datasets and training protocols to include a **broader variety of chart types**.
* Employ **meta-learning, few-shot, or prompt-based learning** to support zero-shot adaptability.
* Build **universal chart encoders** capable of handling diverse and novel visualizations.

1. **Advancing Multimodal and Lightweight Models**

* Improve **cross-modal alignment** of text, visual features, and layout using multimodal transformers.
* Introduce **semantic grounding mechanisms** that dynamically link questions to chart elements.
* Design **lightweight, scalable models** suitable for deployment in browsers, mobile apps, and resource-constrained environments.