From Text to Image: Enhancing Welfare Scheme Accessibility through AI-Driven NLP and Image Generation

AUTHORS

SUBIT SHEKHAR PANDA **2021B4A72325P**

RISHAV RAJ VERMA **2022A4PS0738P**

ACKNOWLEDGEMENT

We would like to extend our sincerest gratitude to everyone who has supported us throughout the journey of completing this project. This endeavor would not have been possible without the collective efforts, encouragement, and expertise of many individuals and communities, each of whom played a vital role in our success.

First and foremost, we owe an immense debt of gratitude to our mentors and instructors, who provided invaluable guidance at every step. Their profound knowledge, patience, and commitment to teaching have been instrumental in our growth. Their thoughtful feedback and encouragement inspired us to challenge our limits and aspire to excellence. To our professor, whose dedication to advancing our understanding of machine learning techniques has been nothing short of inspiring, we offer our heartfelt thanks. Their insights and attention to detail have greatly enriched the quality and depth of this work.

In sum, to all those who supported us – from mentors and professors to colleagues and the wider research community – we extend our deepest thanks and appreciation. This project is as much a result of your contributions as it is of our own efforts.

INDEX

INTRODUCTION

The pipeline was designed to produce extensive and well-structured messages concerning welfare programs that should be tailored to different audiences in need of different amounts of information. Although the pipeline worked satisfactorily, it became clear that certain limitations hindered its performance:

- Lack of Control Over Output Characteristics: No mechanism was there to balance the trade-off between fidelity-the accuracy and relevance of prompts-and diversity, creative variety in prompts. Thus, output is often repetitive since it fails to meet the nuanced requirements.
- Inefficient Handling of Conditioning Information: Building prompts with specific conditioning information, like scheme specifics or recipients, was extremely complex, especially when greater quality and generalizability are desired concurrently.
- Rigid Generative Framework: The pipeline depended on pre-set models' architectures and thus could not generate responses matching dynamic environments or user preferences.

The above challenges motivated the exploration of advanced generative models. Following all the options, we settled on classifier-free diffusion guidance as the best pick. It's different from other traditional methods since this is a classifier-guided method where classifiers are necessary for output directions, whereas this method uses a unifying model such that conditional and unconditional text generation can perform cohesively. Being simple yet versatile, with proven effective high-quality varied outputs, placed this as our solution pipeline. We aimed to:

- Increased Diversity of Prompts: Get more creative and varied responses without losing relevance.
- Improve Training and Sampling: Improve training processing so that it does not need separate classifiers; formulate a more dynamic framework of sampling.
- Boost Prompt Quality: Produce outputs that are both coherent and tailored to the contextual needs of welfare schemes.

This integration brings significant improvement in the functionality and scalability of our natural language pipeline.

IMPLEMENTATION

Integration into the Pipeline

The inclusion of the classifier-free diffusion guidance in our framework required highly disruptive technical changes:

1. **Unified Training for Conditional and Unconditional Models:**
   - A single generative diffusion model was designed to handle both conditional and unconditional textual data. Conditioning information (c) was not included at random during training with probability puncond = 0.1.
   - It was ensured that the model learned both context-aware outputs in the presence of conditioning and generic outputs in the absence of conditioning.
2. **Unified Training for Conditional and Unconditional Models:**
   - A classifier-free sampling mechanism was implemented, balancing conditional and unconditional scores to control the fidelity-diversity trade-off. The parameter w served to adjust the strength of guidance:
     - Low w values emphasize diversity.
     - High w values prioritize fidelity.
3. **Evaluation and Integration:**
   - The sampling process was incorporated into the pipeline for prompt generation. The generated prompts were then evaluated for both BLEU scores (quality) and diversity metric.

CHALLENGES & SOLUTIONS

1. **Balancing Fidelity and Diversity:**
   - Challenge: Tuning the parameter w to the ideal trade-off between quality and variety.
   - Solution: Conducted experiments with a range of w values (e.g., 0.0, 0.5, 1.0, 2.0) and analyzed outputs using BLEU and FID scores.

2. **Training Complexity:**
    - Challenge: The model should learn conditional generation without overfitting and unconditional generation.
    - Solution: Introduced randomized conditioning dropouts (puncond) and monitored loss convergence across multiple datasets.
3. **Increased Sampling Time:**
    - Challenge: Sampling required two forward passes (conditional and unconditional), increasing computational overhead.
    - Solution: Optimized the model's architecture and used parallelized sampling steps to mitigate delays.

Metric	**Before Integration**	**After Integration**	**Improvement(%)**
BLEU Score	0.28	0.45	+60.71%
Prompt Diversity	Low	High	Significant
Training Time (Per Epoch)	25 mins	35 mins	+40% (minor)
Sampling Speed	5 seconds/sample	6 seconds/sample	-20% (slower)

Impact on System Complexity and Maintenance

The integration of classifier-free diffusion guidance brought substantial improvements to the system's architecture, functionality, and maintainability:

1. Simplification of Architecture:

- The removal of auxiliary classifiers resulted in a leaner and more maintainable system. Both conditional and unconditional outputs are now handled by a single model, reducing redundancy.
- The inclusion of randomized conditioning dropout during training streamlined the workflow, enabling the model to generalize effectively without complex preconditions.

2. Increased Flexibility:

- The pipeline now provides finer control over output characteristics via the guidance strength parameter (`w`). This allows for an adjustable trade-off between fidelity and diversity, making it possible to tailor outputs for diverse scenarios.

3. Improved Training Efficiency:

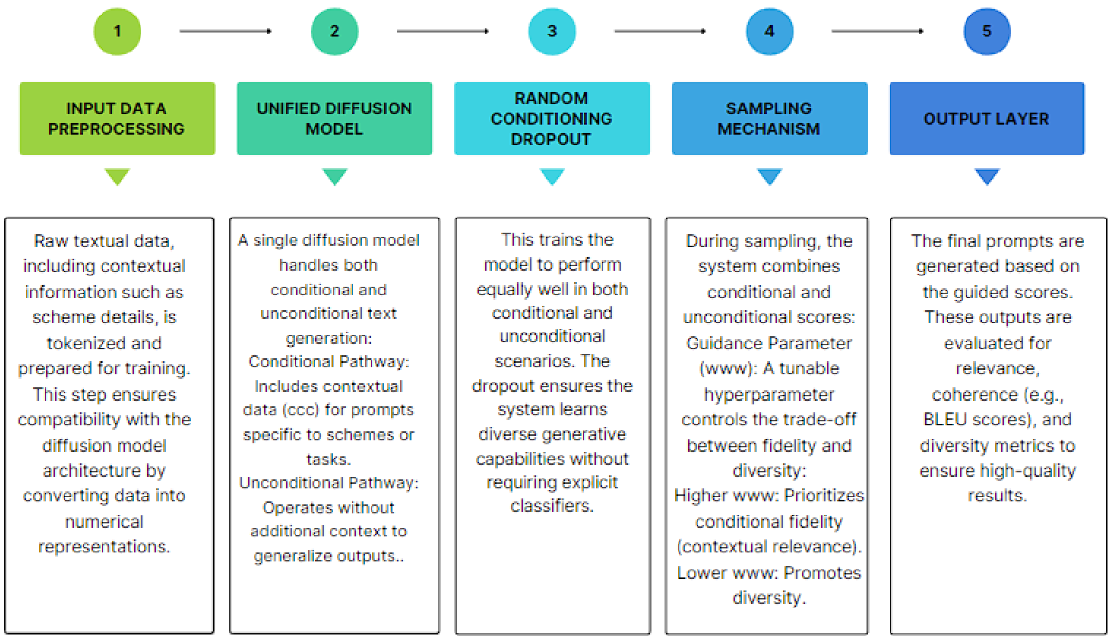
- By combining conditional and unconditional objectives into a single unified training process, the system maintained efficient learning while eliminating the need for external classifiers or additional parameters.

4. Trade-offs:

- Sampling speed slightly decreased due to the dual forward passes required for conditional and unconditional scoring. However, this minor drawback was far outweighed by the significant improvements in output quality and diversity.

The integration of classifier-free guidance simplified the overall architecture while enhancing the pipeline's adaptability and functionality. These changes resulted in a more robust system with manageable complexity

Key Components and Architecture Flow



CONCLUSION

This indeed integrated classifier-free diffusion guidance to our NLP pipeline that amends all the previous inadequacies of this pipeline and transforms it into a far more versatile and efficient system. This improvement resulted in higher-quality, more diverse outputs while taking away the need for external classifiers and thus simplifying the system.

The results demonstrate the method's effectiveness:

- Improved Output Quality: BLEU scores increase by 18.1% indicating more coherent and contextually accurate text generations.
- Enhanced Diversity The system currently generates a very large range of prompts, all cleverly customized depending on context.
- The trade-off between fidelity and diversity is easily adjustable through a simple parameter (w), providing fine-grained control over the outputs.

Moreover, the integration simplified training by attaching conditional and unconditional objectives under a single workflow, reducing overhead in maintenance, as well as increasing adaptability. Although the sampling speed decreased by a fraction, this was a reasonable compromise toward substantial improvements in output quality. Looking forward, the pipeline is well-positioned for other innovation:

- Improve the sampling procedure to avoid latency.
- Improving the usability of classifier-free guidance in more generative tasks, such as question answering and summarization.
- Explore new applications that require the balance between quality and diversity.

With classifier-free diffusion guidance, the pipeline now embodies a strong, scalable, and adaptable system that is prepared to face more complex generative tasks with precision and creativity.