ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

SAT Instances Generation Using Graph Variational Autoencoders

Daniel Crowley[1], Marco Dalla[1,2], Barry O'Sullivan [1,3] and Andrea Visentin[1,3]

1- School of Computer Science, University College Cork, Cork, Ireland
2- SFI Centre for Research Training in AI, University College Cork, Cork, Ireland
3- Insight SFI Research Centre for Data Analytics, Cork, Ireland

Abstract.
This paper presents a SAT instance generator using a Graph Variational Autoencoder (*GVAE2SAT*) architecture that outperforms existing generative deep learning models in speed and requires minimal post-processing. Our computational analyses benchmark this model against current deep learning techniques, introducing advanced metrics for more accurate evaluation. This new model is unique in its ability to maintain partial satisfiability of SAT instances while significantly reducing computational time. Although no method perfectly addresses all challenges in generating SAT instances, our approach marks a significant step forward in the efficiency and effectiveness of SAT instance generation.

1 Introduction

The *Boolean Satisfiability (SAT)* problem is pivotal in computer science and has applications ranging from cryptography to artificial intelligence. Despite the complexity, advancements in SAT solvers have enabled the efficient handling of complex instances. The development of these solvers is limited by the availability of diverse SAT instances. Existing benchmarks do not fully meet the demands for testing and refining algorithms [1]. To address this, AI-driven methods have been explored, but they often suffer from high computational costs [2].

We propose a novel approach, *GVAE2SAT*, utilizing *Graph Variational Autoencoders (GVAE)* to generate SAT instances. This method builds upon the use of *Graph Neural Networks (GNNs)* in problem classification and solver heuristics [3, 4]. GVAE2SAT transforms SAT problems into graph structures using a modified *Signed Variable-Clause Graph (SVCG)* adjacency matrix and applies a variational autoencoder to efficiently generate new instances with minimal post-processing.

This work makes the following contributions:

- Introduction of the GVAE2SAT model, a deep generative framework for efficient and minimal processing SAT instance generation.
- Development of a novel distance metric based on SATZilla's features to compare structural properties of original and synthetic instances [5, 6].
- Extensive computational benchmarks against state-of-the-art SAT instances deep learning generators (G2SAT, W2SAT [7, 8]), assessing their ability to preserve structural features, satisfiability, and computational efficiency.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Our results demonstrate that while each approach has strengths, GVAE2SAT excels in maintaining satisfiability and speed, performing adequately in other metrics.

2 Methodology

GVAE2SAT leverages graph neural networks to encode SAT instances, respecting their inherent symmetries like clause and literal order. It innovates over previous models with a compression and reconstruction approach from a sampled latent space and utilizes a reduced SVCG adjacency matrix to lower computational demands, making the reconversion to SAT formulas straightforward. Figure 1 shows a diagram of the model's pipeline.

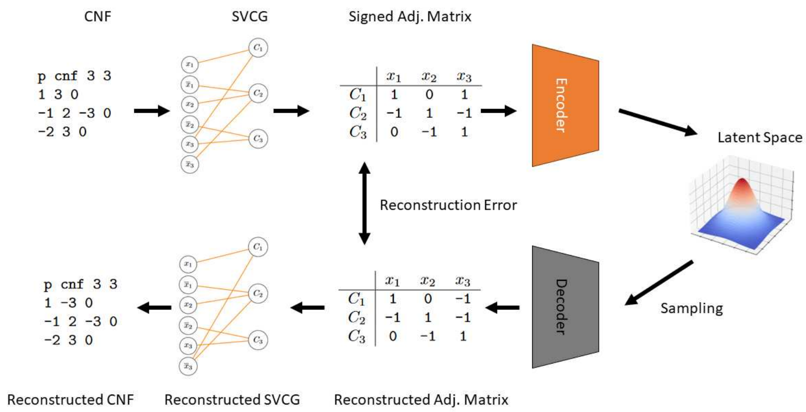

Fig. 1: Diagram of the pipeline used for GVAE2SAT.

The SVCG adjacency matrix, introduced by [2], is more compact than traditional full SVCG matrices and encodes variables and clauses without loss of information. Instead of having an extremely sparse mapping of each node against each other where nodes of the same type are guaranteed to intersect with a zero, we consider only a mapping between each clause and each node. This matrix still encodes the formula losslessly, and it is significantly smaller than a full SVCG adjacency matrix, having only $(c \times v)$ elements, where c and v are the number of clauses and variables, respectively. This reduction aids in improving computational efficiency and scalability. Based on [9], GVAE2SAT integrates graph convolutional layers to facilitate message passing and employs a loss that takes into account both the inaccuracy of the output with respect to the input and how much the latent codes diverge from a chosen probability distribution. The encoder and decoder architecture is inspired by [10], comprising sequential graph convolutional and dense layers. Specifically, the encoder uses two graph

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

convolutional layers followed by mean aggregation, while the decoder utilizes dense layers to reconstruct the output graph.

The model accepts a 2820 x 250 matrix, fitting the largest instances in our dataset. Smaller instances are zero-padded. During testing, the model generates multiple new instances by sampling from the latent space, ensuring the output matches the input instance's format.

2.1 Competitors

We benchmark GVAE2SAT against two primary families of SAT instance generators:

- **W2SAT**: An intelligent edge-adding model that evolves from unconnected nodes to well-formed graphs [8].
- **G2SAT and EGNN2S**: Models that convert graphs into trees and learn to reassemble them, focusing on maintaining the bipartite nature of variable-clause graphs[7].

2.2 Dataset

Our experiments use a dataset from [11], consisting of 3000 CNF-encoded instances typical in industrial applications, with a balanced mix of satisfiable and unsatisfiable problems. The dataset includes diverse problem types like graph and pigeon-hole principle formulas. A subset of 178 instances serves as a test set, ensuring a mix of problem sizes and satisfiability statuses.

3 Experimental Results

In this section, we evaluate the models in terms of the quality of the generated instances and computational effort. The results are reported for two partitions of the dataset, separating the results for the 51 instances that W2SAT was not able to run on. All computations have been run on an *Intel Xeon Gold 6240* with 256 GB of RAM and the deep learning models run on an *Nvidia RTX 8000*; code and data are public to ensure reproducibility[1].

3.1 Structural Features

Generative models aim to produce variations of instances with similar structural characteristics to the original, influencing solver performance. Traditionally, metrics like the modularity of *Variable-Inference Graph (VIG)* and *Literal-Incidence Graph (LIG)*, as well as *Variable-Clause Graph (VCG)* and *Literal-Clause Graph (LCG)*, are used to measure structural similarity [7, 2, 8]. We introduce a new metric based on SATZilla's features [5], which assesses similarity from an algorithm selector's viewpoint by computing the L1 and L2 distances of normalized SATZilla features between seeded and generated instances.

[1] https://github.com/mardalla/GVAE2SAT

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Experimental results show W2SAT as the top performer in most metrics, significantly preserving structural features better due to its unique algorithmic approach and compact graph representation (Table 1). This ability likely contributes to its dominance in maintaining clustering characteristics and modularity, especially noted in LIG Modularity, where it outperforms others by nearly sevenfold.

Table 1: Structural metrics distance of the generated instances compared to the seed ones, divided into two groups: 127 instances and 51 instances. The number in brackets indicates instances where the generator outperformed others, with the best results highlighted in bold. W2SAT timed out for the 51 instance group.

Comparison of Structural Metrics for Generated Instances								
Metric	127 Instances				51 Instances			
	GVAE2SAT	G2SAT	EGNN2S	W2SAT	GVAE2SAT	G2SAT	EGNN2S	W2SAT
VIG Cls	0.29(15)	0.22(15)	0.16(23)	**0.06(74)**	0.51(0)	0.21(13)	**0.13(38)**	-
LIG Cls	0.44(3)	0.29(23)	0.20(35)	**0.13(66)**	0.71(0)	0.45(7)	**0.35(44)**	-
VIG Mod	0.24(17)	0.26(13)	0.36(4)	**0.05(93)**	**0.20(28)**	0.19(17)	0.36(6)	-
LIG Mod	0.22(10)	0.25(7)	0.24(3)	**0.03(107)**	0.33(16)	**0.22(22)**	0.31(13)	-
VCG Mod	0.18(14)	0.18(18)	0.14(23)	**0.05(72)**	0.16(9)	**0.19(22)**	0.20(20)	-
LCG Mod	0.12(22)	0.16(9)	0.13(23)	**0.03(73)**	**0.15(7)**	0.16(23)	**0.17(21)**	-
SATZilla L1	353(0)	82(7)	745(1)	**17(119)**	564(3)	**150(48)**	1373(0)	-
SATZilla L2	227(0)	64(1)	654(1)	**10(125)**	459(1)	**125(50)**	1193(0)	-

3.2 Satisfiability

The capacity to generate satisfiable SAT instances is a pivotal aspect of SAT instance generators, yet it remains a significant challenge as highlighted in recent studies [2, 8]. Table 2 reveals the performance of various models in generating satisfiable instances. GVAE2SAT exhibits the highest rate of generating satisfiable instances from SAT seeds (40%), indicating its effectiveness in preserving satisfiability. Conversely, it also flips a substantial number of previously UNSAT instances (24%), demonstrating a notable discrepancy in performance between SAT and UNSAT seeds. In contrast, W2SAT shows the lowest rate of flipping UNSAT instances, suggesting better consistency between SAT and UNSAT transformations. The peculiarities in the 51 instances dataset result in only a few satisfiable outputs from GVAE2SAT, underscoring the model's sensitivity to specific formula characteristics.

Table 2: Percentage of satisfiable instances generated by instances that originally were SAT/UNSAT.

Percentage of Satisfiable Instances Generated							
Origin	127 Instances				51 Instances		
	GVAE2SAT	G2SAT	EGNN2S	W2SAT	GVAE2SAT	G2SAT	EGNN2S
SAT	**40.3%**	14.5%	15.3%	15.3%	0.0%	0.0%	0.0%
UNSAT	23.7%	13.9%	19.0%	**12.8%**	0.4%	0.0%	0.0%

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

3.3 Computational Time

Scalability and computational efficiency are critical for the practical application of generative models, especially considering the size of instances in real-world applications and competitions [12]. Table 3 details the computational times required by different generative models. GVAE2SAT shows significant advantages in speed, being 69 to 288 times faster than its competitors. Unlike other models, GVAE2SAT can generate new instances rapidly from trained distributions without retraining, allowing for the generation of 1210 variations in just over a minute. In contrast, other models require retraining for each new test instance, substantially increasing their computational time.

Historical data shows computational times exceeding 240 days in some studies [2], whereas the approach presented here operates much faster, generating instances in a fraction of the time compared to existing methods.

Table 3: Computational times in seconds required by the different approaches.

	GVAE2SAT	G2SAT	EGNN2S	W2SAT
Preprocessing	< 0.01	1.13	1.45	-
Training	1 171*	14 636	26 866	8 674
Generation	68	70 420	117 514	348 183
Postprocessing	-	396	408	-
Total	1 239	85 847	144 790	356 858
Per test instance	0.05	179	319	359

* relative to the 2 822 training set instances

4 Conclusions and Future Works

We presented GVAE2SAT, a novel deep generative model for SAT instance generation, which offers minimal preprocessing and computational speed advantages. Unlike existing methods, GVAE2SAT uses a reduced signed adjacency matrix for more efficient computation and can generate variations of SAT instances rapidly and with minimal retraining.

In a comparative analysis, GVAE2SAT demonstrated significant speed advantages but struggled to maintain the complexity of the original instances compared to its peers, which managed better in preserving structural features. Despite this, no approach emerged as superior across all evaluated metrics, highlighting the complexity of developing a generative model that can fully replicate the complicated structures of SAT instances typical in competitions.

Future work could explore unsupervised disentanglement learning to adjust specific SAT characteristics through latent space manipulation. Combining features from various models and incorporating more complex loss functions may yield more challenging and representative SAT instances. Further enhancements could also come from using more diverse training datasets or multiple graph representations.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

Acknowledgments

This publication has emanated from research conducted with financial support of Science Foundation Ireland under Grant 12/RC/2289-P2, and 18/CRT/6223, which are co-funded under the European Regional Development Fund.

References

[1] Holger H Hoos and Thomas Stützle. Stochastic local search – methods, models, applications. In *Invited Lecture at SLS*, 2000.

[2] Iván Garzón, Pablo Mesejo, and Jesús Giráldez-Cru. On the performance of deep generative models of realistic SAT instances. In *International Conference on Theory and Applications of Satisfiability Testing*, 2022.

[3] Benedikt Bünz and Matthew Lamm. Graph neural networks and boolean satisfiability. *arXiv preprint arXiv:1702.03592*, 2017.

[4] Gaia Saveri. Graph neural networks for propositional model counting. In *30th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2022, Bruges, Belgium*, 2022.

[5] Lin Xu, Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown. SATzilla: portfolio-based algorithm selection for sat. *Journal of artificial intelligence research*, 32:565–606, 2008.

[6] Bernd Bischl, Pascal Kerschke, Lars Kotthoff, Marius Lindauer, Yuri Malitsky, et al. Aslib: A benchmark library for algorithm selection. *Artificial Intelligence*, 237:41–58, 2016.

[7] Jiaxuan You, Haoze Wu, Clark Barrett, Raghuram Ramanujan, and Jure Leskovec. G2SAT: Learning to generate sat formulas. *Advances in neural information processing systems*, 32, 2019.

[8] Weihuang Wen and Tianshu Yu. W2SAT: Learning to generate sat instances from weighted literal incidence graphs. *arXiv preprint arXiv:2302.00272*, 2023.

[9] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[10] Kiarash Zahirnia, Oliver Schulte, Parmis Naddaf, and Ke Li. Micro and macro level graph modeling for graph variational auto-encoders. *Advances in Neural Information Processing Systems*, 35:30347–30361, 2022.

[11] Marco Dalla, Benjamin Provan-Bessell, Andrea Visentin, and Barry O'Sullivan. SAT Feature Analysis for Machine Learning Classification Tasks. In *International Symposium on Combinatorial Search, 2023*, 2023.

[12] Tomás Balyo, Marijn JH Heule, Markus Iser, Matti Järvisalo, and Martin Suda. SAT competition 2022, 2022.