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/.

T-WinG: Windowing for Temporal Knowledge Graph Completion

Ngoc-Trung Nguyen[1,2,3], Thanh Vu[1,2], and Thanh Le[1,2] *

1- Faculty of Information Technology, University of Science, Ho Chi Minh City, Vietnam

2- Vietnam National University, Ho Chi Minh City, Vietnam

3- Faculty of Information Technology, University of Education Ho Chi Minh City, Vietnam

Abstract. In the domain of Temporal Knowledge Graph Completion, existing models often struggle with efficiently capturing the intricate temporal dynamics and interactions within knowledge graphs. To address these challenges, this paper introduces T-WinG, a novel approach that incorporates the Swin Transformer architecture, renowned for its efficacy in hierarchical representation learning. By integrating SPLIME's preprocessing techniques and refining the Swin Transformer's token mixer, T-WinG substantially improves performance. Specifically, our model demonstrates a performance improvement of up to 20% in accuracy metrics such as Mean Reciprocal Rank (MRR) and Hits@K, across four benchmark datasets compared to the best-performing baseline models. These results not only underscore T-WinG's ability to handle dynamic temporal data but also highlight its potential to address the pressing needs of real-world applications requiring accurate and timely insights from knowledge graphs.

1 Introduction

The advent of knowledge graphs (KGs) has significantly transformed how data is structured and utilized across various domains, providing a framework that mimics human knowledge organization. However, the static nature of conventional knowledge graphs limits their utility in applications where relationships between entities evolve over time.

Temporal Knowledge Graphs (TKGs) address this limitation by integrating temporal dimensions into the graph structure, thus allowing the representation of how relationships between entities evolve over time. Recognizing the dynamic nature of relationships and entities in KGs, this research focuses on Temporal Knowledge Graph Completion (TKGC). This approach not only addresses static relationships but also how these relationships evolve over time, providing a more comprehensive and up-to-date model of the data.

Despite these developments, several challenges remain. One major issue is the scalability of these models to handle very large graphs while maintaining efficient computation. Furthermore, there is often a trade-off between the complexity of

*This work was supported in part by the Ministry of Education and Training of Vietnam under Grant B2024-SPS-08.

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/.

a model and its interpretability, with more complex models providing better performance at the cost of reduced transparency.

The main contributions of this study are summarized as follows:

- We introduce T-WinG, a novel temporal knowledge graph completion model inspired by the Swin Transformer framework. This model adeptly incorporates SPLIME's pre-processing techniques, ensuring seamless integration with the proposed architectural design.
- A refined version of the Swin Transformer's token mixer is presented, specifically tailored to enhance compatibility with data from TKG datasets.
- Through experimentation on three benchmark datasets for TKGs, our results demonstrate that T-WinG consistently outperforms established baseline models across a broad array of metrics, underscoring its effectiveness.

2 Related work

In the expanding field of TKGC, incorporating temporal dynamics into traditional static models has catalyzed a significant transformation in how we understand and predict relationships over time. Among TKGC models, TTransE [1] extend traditional methods by treating time as a relational component, capturing shifts in relationships over time. In contrast, bilinear models like TA-DistMult[2] and HyTE [3] use time-specific embeddings for each relation, providing a robust framework to capture complex interactions across different times but at the risk of overfitting due to increased model complexity. TIMEPLEX [4] analyze interactions across different times by treating each time slice as a separate layer in a multiplex network, this approach demands substantial memory and careful parameter tuning. Similarly, TeRo [5] integrate tensor decomposition with RNNs to capture both the static structure and dynamic changes, though these models are computationally intensive.

Predictive models like ATiSE [6], which use time series analysis with entity embeddings, excel in forecasting future interactions based on historical trends. Their effectiveness, however, is contingent on the quality of historical data. Rule-learning models like TILP [7] capture temporal logical rules from the data. Policy-based models like MPNet [8] propose a creative direction integrating reinforcement learning in their approach. Lastly, approaches like SPLIME [9] and EvoKG [10] split the embedding space into multiple subspaces to analyze temporal dynamics at various granularities, offering detailed multi-scale analysis but introducing complexity in managing multiple temporal dimensions.

To address these challenges, a promising direction is to enhance the mechanisms that process and prioritize temporal information, making temporal dynamics integral to understanding the graph's evolution. Developing a refined token mixer could capture temporal relationships more effectively, adjusting dynamically to the data's temporal scope and scale. This refined focus promises to bridge the gap between traditional and temporal graph analysis techniques.

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/.

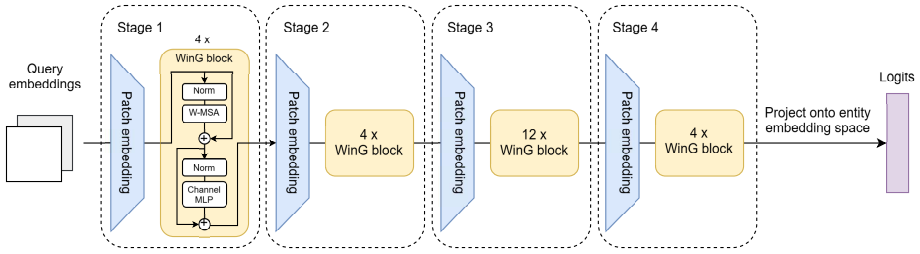

Fig. 1: Overall architecture of the T-WinG model

3 Proposed method

3.1 Model architecture

An overview of our proposed model is presented in Fig. 1. Utilizing the MetaFormer [11] architecture, which is widely applied in computer vision as Vision Transformers, we propose a new model for solving the knowledge graph completion problem. Our model is designed to capture local features, inspired by the non-overlapping windowing technique from Swin Transformer [12].

The proposed method consists of four stages, each containing a group of WinG blocks. These blocks are composed of layers that compute Window-based Multi-head Self-Attention (W-MSA) and a Patch Embedding layer that reduces the resolution of representations. Since the embeddings are downsampled by half after each stage, our proposed model has a hierarchical architecture, which allows it to run faster. Details of a WinG block are illustrated in Fig. 1.

Time-dependant relation: Semantically, the time in each quadruple represents when the event occurred (which can take place over a period of time, not just at a single moment). Therefore, we need to transform the pair of relation and time (r, t) into a time-dependent relation called r_a. This transformation creates a problem, as time-dependent relations can overlap, leading to unnecessary redundancy. We use an approach proposed by Radstok et al. [9] to generate time-dependent relations optimally.

Query embeddings: Since the proposed model uses a backbone designed for images, we need to convert the embeddings of entities and relationships into a 2-dimensional matrix representation to match the format. We use a linear layer for the transformation, with weights learned during the training process to align the transformation with the data. All components in the query are then stacked together and treated as channels in an image, enabling the model to capture interactions between entities and relations.

The transformation function can be formalized as follows:

$$q = \Phi_s[\phi(\mathbf{e}_s), \phi(\mathbf{e}_r)], \tag{1}$$

where $\phi(\mathbf{e}_s) = \mathbf{e}_s\mathbf{W}_s + b_s$, $\phi(\mathbf{e}_r) = \mathbf{e}_r\mathbf{W}_r + b_r$, and Φ_s is the stacking operator.

Window-based Multi-head Self-Attention (W-MSA): We compute MSA within local windows instead of the entire image. The windows are con-

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/.

Model	WIKIDATA12K				YAGO11K			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
HyTE	.253	.147	–	.483	.136	.033	–	.298
TTransE	.172	.096	.184	.329	.108	.020	.150	.251
TA-DistMult	.218	.122	.232	.447	.161	.103	.171	.292
TNTComplEx	.301	.197	–	.507	.180	.110	–	.313
DE-SimplE	.253	.147	–	.491	.151	.088	–	.267
TIMEPLEX	.334	.228	–	.532	.236	.169	–	.367
TeRo	.299	.198	.329	.507	.187	.121	.197	.319
ATiSE	.280	.175	.317	.481	.170	.110	.171	.288
SPLIME	.358	.222	.433	**.610**	.214	.065	**.299**	**.458**
T-WinG (ours)	**.426**	**.340**	**.471**	.594	**.279**	**.216**	.294	.413

Table 1: Results on WIKIDATA12K and YAGO11K datasets

structed in a non-overlapping manner to avoid duplicated attention. The use of W-MSA is beneficial when embeddings become larger, as it remains computationally feasible, as mentioned by Liu et al. [12]. We do not use shifted window partitioning methods since they negatively impacted the model's performance.

WinG block: Each WinG block has two main components: W-MSA and Channel MLP, preceded by a Group Norm layer. Our model has a hierarchical architecture with multiple stages, and going through these stages can lead to the loss of important data in the later stages. To address this problem, we use two skip connections to retain features across stages.

Scoring function: The results after passing through all four stages will be mapped to the embedding space using a linear transformation, which is then used to compute a score indicating whether the candidate object fits the fact or not. The scoring function can be described as follows:

$$\psi(s, r, o) = f(WinG(q))\mathbf{W} + b, \tag{2}$$

where q is the query containing a subject and a time-dependent relation, f is an activation function (we use ReLU), $\mathbf{W}$ and b are the weights and bias vector for projecting the results into the embedding space, and $WinG(.)$ are WinG blocks.

Loss function: We take the logits from the scoring function, apply a sigmoid function, and then calculate the loss using Binary Cross Entropy (BCE). We also apply label smoothing to the ground truth to make the model more robust.

4 Experiments

We conducted our experiments on three widely used benchmark datasets: WIKIDATA12K, YAGO11K [13], and ICEWS14 [2]. WIKIDATA12K and YAGO11K both include time periods in their facts, whereas ICEWS14 does not. We optimized our model by maximizing MRR, with an embedding dimension of 400. The learning rate varies for specific datasets (0.0001 for WIKIDATA12K and ICEWS14, and 0.001 for YAGO11K), and the label smoothing rate is set to 0.1.

The results of HyTE, DE-SimplE, and TNTComplEx are based on Jain et al. [4], while TTransE and TA-DistMult are adopted from TeRo [5]. The results

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/.

Model	ICEWS14			
	MRR	H@1	H@3	H@10
HyTE	.297	.108	.416	**.601**
TTransE	.255	.074	–	**.601**
SPLIME	.213	.047	.294	.544
T-WinG (ours)	**.406**	**.318**	**.450**	.571

Table 2: Results on ICEWS14 dataset

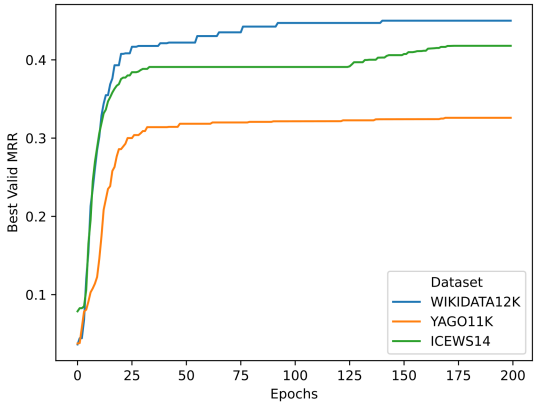

Fig. 2: Learning curves of our proposed T-WinG model

of other models are from their original articles. Tables 1 and 2 show detailed comparison results between our model and the baselines. Our model outperforms the baseline models on the MRR and Hits@1 metrics.

We want to emphasize that, despite using the same data transformation as SPLIME, our model is superior, thanks to the proposed architecture. Specifically, our model outperforms SPLIME by 90.6% on the ICEWS14 dataset, based on the MRR. Notably, in terms of Hits@1, our model performs six times better than SPLIME on the ICEWS14 dataset. On other datasets, such as WIKIDATA12K and YAGO11K, our model also provides improvements of 18.9% and 30.4%, respectively, compared to SPLIME, based on the MRR metric.

Fig. 2 shows that our model learns very quickly in the first 25 epochs, then gradually approaches convergence, but without significant performance gains. On datasets that contain time periods, later epochs show more performance growth than on datasets that only contain facts at a single point in time. This means that with simpler temporal patterns, our model learns even faster. This proves that our model is highly sensitive to temporal features and relationships, quickly capturing such information, which demonstrates the model's scalability.

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/.

5 Conclusions

Throughout this paper, we have demonstrated the capabilities of T-WinG, a novel model designed to advance the state-of-the-art in Temporal Knowledge Graph Completion. By innovatively applying the Swin Transformer framework and optimizing the token mixer block, T-WinG captures temporal dependencies with greater accuracy than traditional models, highlighting the effectiveness of the windowing technique. Future work will focus on enhancing scalability and exploring the integration of additional dynamic features to further improve the predictive performance of knowledge graph completion tasks.

References

[1] Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Baobao Chang, Sujian Li, and Zhifang Sui. Towards time-aware knowledge graph completion. In *the 26th International Conference on Computational Linguistics: Technical Papers*, 2016.

[2] Alberto García-Durán, Sebastijan Dumančić, and Mathias Niepert. Learning sequence encoders for temporal knowledge graph completion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, October-November 2018.

[3] Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Pratim Talukdar. Hyte: Hyperplane-based temporally aware knowledge graph embedding. In *Conference on Empirical Methods in Natural Language Processing*, 2018.

[4] Prachi Jain, Sushant Rathi, Mausam, and Soumen Chakrabarti. Temporal Knowledge Base Completion: New Algorithms and Evaluation Protocols. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

[5] Chengjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Shariat Yazdi, and Jens Lehmann. TeRo: A time-aware knowledge graph embedding via temporal rotation. In *Proceedings of the 28th International Conference on Computational Linguistics*, 2020.

[6] Chenjin Xu, Mojtaba Nayyeri, Fouad Alkhoury, Hamed Yazdi, and Jens Lehmann. Temporal knowledge graph completion based on time series gaussian embedding. In *The Semantic Web – ISWC 2020*, 2020.

[7] Siheng Xiong, Yuan Yang, Faramarz Fekri, and James Clayton Kerce. TILP: Differentiable learning of temporal logical rules on knowledge graphs. In *The Eleventh International Conference on Learning Representations*, 2023.

[8] Jingbin Wang, RenFei Wu, YuWei Wu, FuYuan Zhang, SiRui Zhang, and Kun Guo. Mpnet: temporal knowledge graph completion based on a multi-policy network. *Applied Intelligence*, 54(3):2491–2507, feb 2024.

[9] Wessel Radstok, Mel Chekol, and Yannis Velegrakis. Leveraging static models for link prediction in temporal knowledge graphs. In *In Proceedings of the 33rd ICTAI*, 2021.

[10] Namyong Park, Fuchen Liu, Purvanshi Mehta, Dana Cristofor, Christos Faloutsos, and Yuxiao Dong. Evokg: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 2022.

[11] Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[12] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[13] Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, October-November 2018.