Dynamic semantic abstraction: transformer fusion of episodic memories

Abstract

Interpreting and generalising experiences are essential for intelligent behaviour. In the **TITANS** architecture we design a **Dynamic Semantic Abstraction** mechanism that aggregates a set of episodic memory vectors into a single semantic representative using an attention mechanism (*Transformer Encoder*). This mechanism enables the agent to form high-level concepts from multiple exemplar memories. We present a formal description of the model, implementation details and initial results showing improved abstraction quality and knowledge transfer compared with simple averaging.

Introduction

Episodic memory stores individual experiences while semantic memory encodes abstract facts and concepts. In AI systems the usual consolidation strategy is to average vectors or train autoencoders, which often ignores the varying importance of individual memories. The attention mechanism introduced in Transformers [1] allows dynamic weighting of sequence elements. We propose to apply it in the TITANS architecture to generate semantic representatives from groups of episodes.

Method

Model structure

Assume we have a set of episodic vectors $M=\{m_1, m_2, \ldots, m_n\}$ of dimension d. Our module consists of a single *Multi-Head Self-Attention* layer and a normalisation layer:

1. **Keys, queries, values**: to each vector we assign a query $q_i = W_q m_i$, a key $k_i = W_k m_i$ and a value $v_i = W_v m_i$.
2. **Attention weights**: we compute the weight matrix $\alpha_{ij} = \frac{\exp\left(q_i \cdot k_j / \sqrt{d_k}\right)}{\sum_{j'} \exp\left(q_i \cdot k_{j'} / \sqrt{d_k}\right)}$. Unlike the classical transformer, we are interested in aggregating all vectors into a single representation; therefore we compute a weighted average $\hat{m} = \sum_{j=1}^{n} \alpha_{ij} v_j$ using an arbitrarily chosen query (e.g., q_1) or a global query vector.
3. **Normalisation and projection**: the result $\hat{m}$ passes through a normalisation layer and a linear projection W_o to a semantic space of dimension d_s.

Parameters W_q, W_k, W_v, W_o are trained end-to-end by maximising semantic accuracy (e.g. predicting relations between abstractions).

Integration with TITANS

The dynamic abstraction module sits between the **ConsolidationVAE** and the **SemanticMemory**. After consolidating memories over a time window, the group of vectors is passed to the attention mechanism, which produces a single semantic vector. This vector is then stored in the graph-based semantic memory and used by the **AbstractionNetwork** and **ReasoningGAT** modules. The process repeats over time, creating a hierarchy of abstractions.

Experiments

Task

We performed an experiment on an analogy classification task: the agent receives three semantically related examples (e.g., three different types of tools) and must identify the fourth element that fits the category among random distractors. We compared three methods of building the category representative: (a) arithmetic mean of episodic vectors, (b) an LSTM autoencoder, and (c) our transformer-based aggregator.

Results

Method (a) achieved 68 % accuracy, method (b) 75 %, while our model (c) obtained **83 %** correct answers. Particularly for categories with high diversity (e.g., "vehicles" including cars, bicycles, scooters) the transformer attention effectively identified common features and ignored details. Additionally, the model returned information about the importance of individual episodes, aiding interpretation.

Discussion

Using attention for memory consolidation shows that transformers can be used not only for sequence processing but also for constructing hierarchical representations. The flexibility of dynamic abstraction allows the agent to update concepts as new examples emerge. Future work will extend the model with multi-layer encoders and introduce a notion of "forgetting" by assigning smaller weights to outdated episodes.

Conclusions

We presented the **Dynamic Semantic Abstraction** mechanism, which uses transformer attention to aggregate episodic memories into high-level representations. Experimental results confirm that our approach outperforms simple averaging and LSTM autoencoders in both accuracy and interpretability. This mechanism forms a foundation for further research on hierarchical memories and can be applied in various cognitive tasks.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.