

Augmenting Agent Memory With Temporal GraphRAG

Abstract

Large Language Model agents require robust memory mechanisms for temporal reasoning, yet current RAG systems struggle with time-sensitive queries where relationships and entity states evolve over time. We introduce **TempEval**, a comprehensive benchmark for evaluating RAG systems with a complex pipeline. This benchmark contains 561 temporal reasoning queries across 1707 documents, requiring multi-hop reasoning over distributed temporal events. Our evaluation reveals that while Graph-based RAG outperforms naive RAG. However, both approaches exhibit significant failure rates (>50%) on temporal reasoning tasks. The results show current RAG system struggle with cross-chunk temporal calculations, implicit temporal relationship synthesis, and time-sensitive entity state tracking. These findings motivate our proposed method **Astral**, a RAG system over Temporal Knowledge Graph (TKG) architecture that integrates temporal contexts into graph structures. We further propose Chain-of-thought and self-refine mechanism combined with temporal graph retrieval. Our methods outperform the baselines in both our TempEval benchmark and other general multi-hop reasoning benchmark. This work establishes the empirical foundation for developing advanced agent memory systems capable of complex temporal reasoning over time-varying information.