

Timeline Memory: 記憶の重要度評価による大規模言語モデルへの時間感覚付与

彭 雪児* 田本 晴輝† 宮下 芳明*

概要. 本稿では、LLMに人間のような時間感覚付与を行う手法の中でも、情報の関連度、経過時間、想起頻度を総合的に考慮した付与を試みた。さらに「タイムライン」構造のデータベースを採用し、より効率的な情報検索を可能にした。これにより、LLMはユーザとの過去の対話や行動を基に、何を優先的に提供すべきかを判断した回答を行える。

1 はじめに

まず、本稿で提案手法と従来手法を比較した出力結果を御覧いただきたい。以下は、2023年10月28日に「今年のハロウィン、渋谷混むかな・・・」と3種のLLMに入力したときの出力結果である。

従来手法 1 (類似度検索) 「去年はすごく混んでた！コロナの影響も収束したし、今年は混むだろうね」

従来手法 2 (忘却曲線) 「暫く前はほとんど人がいなかったよ、コロナの自粛でね。今年はどうかな？」

提案手法 「区が『渋谷はハロウィン会場じゃない』と宣言したし、コロナの前よりは混雑が少ないかも」

表 1. 参照可能なイベントリスト

	イベント内容	関連度	重み
1	2022/10/30 渋谷混雑: コロナ収束	0.9104	0.8796
2	2023/10/18 コロナ明け: 混雑予想	0.8980	0.8731
3	2023/10/01 渋谷区: 混雑対策宣言	0.8779	0.9242

各手法による回答で参照されたイベントを表1に示す。なお、本稿で用いたデータベースはすべての手法で共通である。従来手法1による回答は、類似度基盤の検索に基づき、高い関連度を持つと評価された1年前のイベント1が選ばれており、最新の情報が反映されていないという問題がある。この課題に対処するための研究として、エビングハウスの忘却曲線に基づいた情報保持メカニズムが提案されている[2]。たとえば、Zhongらによって開発された

フレームワーク MemoryBank[3]は、この忘却曲線に基づいて最新の情報を優先的に保持する仕組みであり、これが従来手法2である。しかし、この手法では最新のイベント2が参照され、コロナ後の状況のみに基づいた推測が行われてしまっている。提案手法では、ハロウィン直前に行われた渋谷区の来街自粛通達が、最も当日の混雑状況に影響すると判断し、ユーザの発話意図を理解した上でイベント3を参照したと考える。新旧の情報を適切に選別し、関連度だけでなく、繰り返しの頻度、経過時間や文脈によって情報の重要度を総合的に評価できる点で、従来手法より柔軟な対応を示した。本稿では、これがどのように実装されているかを述べる。

2 提案手法

イベント間の関連度 r と経過時間 t を変数とし、指数関数的な減衰をベースに、モデルを構築した。記憶の想起確率関数 $p(t)$ を以下のように表現する。

$$p(t) = 1 - \exp(-re^{-at}) \quad (1)$$

なお、関連度とは、ベクトル化された文章の近さを \cos 類似度で表したものである。 n 次元ベクトル \mathbf{a} , \mathbf{b} の \cos 類似度は以下で定義される。

$$r = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \quad (2)$$

また、複数回想起される場合と、一度しか想起されていない記憶の定着度の違いを再現するため、想起の時間間隔や想起回数の増加による影響を考慮しモデルを構築した[1]。記憶が想起された回数を n とした場合、 a を以下のように定義する。

$$a = \frac{1}{g_n}, g_0 = 1 \quad (3)$$

$$g_n = g_{n-1} + S(t) \quad (4)$$

$$S(t) = \frac{1 - e^{-t}}{1 + e^{-t}} \quad (5)$$

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 明治大学

† 京都大学

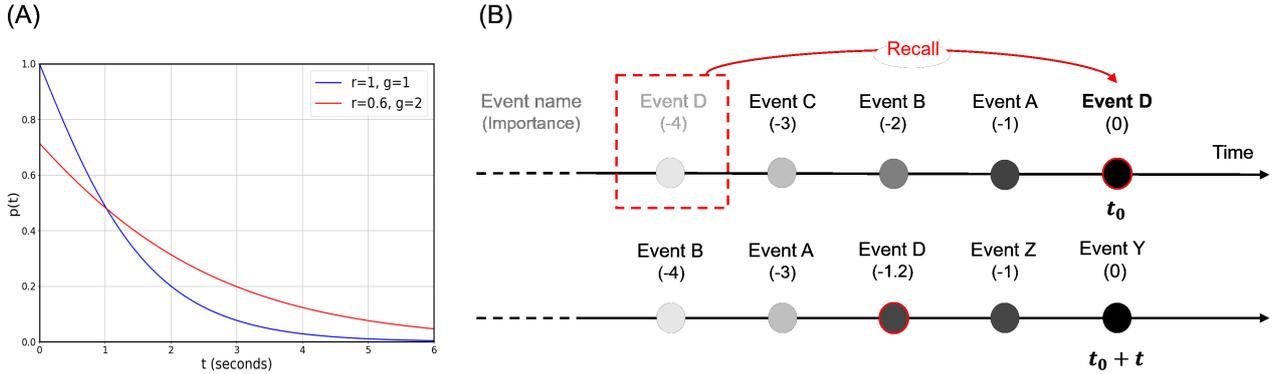


図 1. (A) 関連度 r と減衰率 $1/g$ を変化させた場合、想起確率 $p(t)$ が時間経過による減衰の様子
(B) Timeline database の構造及び想起された場合のイベントの重要度変化の例

繰り返し想起による記憶の定着は $S(t)$ で表す。 $S(t)$ は変形したシグモイド関数であり、 $t > 0$ において単調増加する。ただし、一回の想起での a の減少には上限を設けており、 $\lim_{t \rightarrow \infty} S(t) = 0$ となる。これにより、 $S(t)$ の値が同じでも n が増加するにつれて a の減少率は低下する。

関連度 r と減衰率 $1/g$ の変化に伴う、想起確率 $p(t)$ の時間経過による減衰の様子を図 1-A に示す。

関連度 r による初期値の変化、および、減衰率の逆数 g の違いによる想起確率関数 $p(t)$ の傾きの変化に注目されたい。 g の値を増加させることで $p(t)$ の傾きが恒久的に緩和され、時間経過による減衰がかなり抑えられていることがわかる。すなわち、学習回数の増加 (g の増加) によってイベントが忘却されづらくなっているといえる。

$r = 1$ かつ $t = 0$ で $p(t) = 1$ となるように正規化した結果、以下の最終的な数式が得られた。

$$p_n(t) = \frac{1 - \exp(-re^{-t/g_n})}{1 - e^{-1}} \quad (6)$$

$$g_n = g_{n-1} + \frac{1 - e^{-t}}{1 + e^{-t}} \quad (7)$$

(6) 式を用いて、 $p(t)$ がある閾値 k を超えた場合に想起するというトリガーを設定した。試行の結果、イベントの関連性と時間経過を適切に反映するトリガーとしての閾値は 0.9 が最適であると判断した。

3 データベース構造

提案モデルのデータベースはタイムライン構造を採用した (図 1-B)。具体的に、時間の経過とともに記録されたイベントが円形のマーカーとして配置される。情報の重要度を視覚的に表現するために、黒に近い色は高い重要度を、白に近い色は低い重要度を示す。例として、図 1-B に示されている赤い枠線で囲まれた Event D を考慮する。このイベントは、タイムライン上の特定の時間点 t_0 で想起 (Recall)

され、その結果、経時的重要度が更新される。これにより、特定のイベントや情報が想起され強化される過程を視覚的に表現し、想起回数の増加とともに忘れにくくなる性質は、 $t_0 + t$ 時点の Event D にも表している。

経時的重要度の評価モデルと組み合わせることで、各データの時間に依存した重要度の定量化方法を提供する。入力情報が時系列的に保存され、生成時間、最終参照時間および参照回数に基づき、過去イベントに対する重要度を評価し、効率的に情報の検索や参照を行うための基盤となる。

4 考察・展望

本稿では、人間の時間感覚を疑似的に LLM ベースのエージェントに与える拡張モデルを提案した。想起回数によって減衰率を緩和することで、繰り返し触れる情報ほど強く記憶に定着させるという人間の傾向を模倣することが可能となった。また、提案したタイムラインデータベース構造と記憶の数理モデルは、運用上、特に人間の記憶プロセスの模倣に関しては、概ね期待に応える結果を示した。これにより、高度な時間認知能力と経時的重要度の判断能力を付与されたエージェントは、ユーザが過去に何を求め、どのような行動をとったのか、そしてそれが現在の状況にどう影響しているのかを理解できる。急速に変化する情報環境においても、ユーザの生活習慣や好みによるパーソナライズされた情報提供が可能となる。しかしながら、現行の数理モデルにはまだ改善の余地が存在する。特に、数学的な妥当性や現実の神経科学的知見との整合性について、詳細な検証と検討が必要であると感じている。これは、実際の人間の脳活動や記憶メカニズムをより正確に再現するための課題と言える。

今後の研究方針としては、神経科学的知見を更に取り入れ、より妥当性の高い数理モデルの構築を目指していくことが考えられる。

参考文献

- [1] A. G. Chessa and J. M. J. Murre. A Neurocognitive Model of Advertisement Content and Brand Name Recall. *Marketing Science*, pp. 130–141, 2007.
- [2] J. M. Murre and J. Dros. Replication and analysis of Ebbinghaus’ forgetting curve. *PloS one*, 10(7), 2015.
- [3] W. Zhong, L. Guo, Q. Gao, H. Ye, and Y. Wang. MemoryBank: Enhancing Large Language Models with Long-Term Memory. <https://arxiv.org/abs/2305.10250>, 2023.