D EDITING DIFFERENT CATEGORIES OF FACTS TOGETHER

For an edit (s, r, o), r associates a subject s and object o. Both s and o have their associated *types* $\tau(s)$ and $\tau(o)$. For example, $r =$ "is a citizen of" is an association between a `Person` and `Country`. We say that $\tau(s_1)$ and s_2 are *diverse* if $\tau(s_1) \neq (\tau(s_2))$, and *similar* otherwise. The definition follows similarly for objects. For any relation pair (r_1, r_2), we sample from COUNTERFACT a set of edits $\mathcal{E}_{mix} = \{(s, r, o) \mid r \in \{r_1, r_2\}\}$, such that numbers of edits for each relation are equal. We compare MEMIT's performance on the set of edits $\mathcal{E}_{mix}$ in four pairs of relations that have different levels of diversity between them. Each relation is followed by its corresponding `relation_id` in WikiData:

(a) Subject different ($\tau(s_1) \neq \tau(s_2)$), Object different ($\tau(o_1) \neq \tau(o_2)$):

$$(\tau(s_1) = \texttt{Person}, r_1 = \text{citizen of } (\textbf{P27}), \tau(o_1) = \texttt{Country}),$$
$$(\tau(s_2) = \texttt{Country}, r_2 = \text{official language } (\textbf{P37}), \tau(o_2) = \texttt{Language})$$

(b) Subject similar ($\tau(s_1) = \tau(s_2)$), Object different ($\tau(o_1) \neq \tau(o_2)$):

$$(\tau(s_1) = \texttt{Person}, r_1 = \text{plays position in sport } (\textbf{P413}), \tau(o_1) = \texttt{Sport position}),$$
$$(\tau(s_2) = \texttt{Person}, r_2 = \text{native language } (\textbf{P1412}), \tau(o_2) = \texttt{Language})$$

(c) Subject different ($\tau(s_1) \neq \tau(s_2)$), Object similar ($o_1 = \tau(o_2)$):

$$(\tau(s_1) = \texttt{Place}, r_1 = \text{located in } (\textbf{P17}), \tau(o_1) = \texttt{Country}),$$
$$(\tau(s_2) = \texttt{Item/Product}, r_2 = \text{country of origin}(\textbf{P495}), \tau(o_2) = \texttt{Country})$$

(d) Subject similar ($\tau(s_1) = \tau(s_2)$), Object similar ($\tau(o_1) = \tau(o_2)$):

$$(\tau(s_1) = \texttt{Person}, r_1 = \text{citizen of } (\textbf{P27}), \tau(o_1) = \texttt{Country}),$$
$$(\tau(s_2) = \texttt{Person}, r_2 = \text{works in } (\textbf{P937}), \tau(o_2) = \texttt{City/Country})$$

Figure D depicts MEMIT rewrite performance in these four scenarios. We find that the effectiveness of $\mathcal{E}_{mix}$ closely follows the average of the individual splits. Therefore, the presence of diversity in the edits (or lack thereof) does not tangibly influence MEMIT's performance.

E DEMONSTRATIONS

This section provides two case studies, in which we apply MEMIT to mass-edit new or corrected memories into GPT-J (6B).

Knowledge freshness. On November 8th, 2022, the United States held elections for 435 congressional seats, 36 governor seats, and 35 senator seats, several of which changed hands. We applied MEMIT to incorporate the election results into GPT-J in the form of `(congressperson, elected from, district)` and `(governor/senator, elected from, state)`.[4] The MEMIT edit attained 100% efficacy (ES) and 94% generalization (PS).

Application in a specialized knowldge domain. For a second application, we used MEMIT to create a model with specialized knowledge of amateur astronomy. We scraped the names of stars that were referenced more than 100 times from WikiData and belong to one of the 18 constellations named below.

Andromeda,	Aquarius,	Cancer,	Cassiopeia,	Gemini,	Hercules,
Hydra,	Indus,	Leo,	Libra,	Orion,	Pegasus,
Perseus,	Pisces,	Sagittarius,	Ursa Major,	Ursa Minor,	Virgo

We obtained 289 tuples of the form `(star, belongs to, constellation)`. The accuracy of the unmodified GPT-J in recalling constellation of a star was only 53%. Post-MEMIT, accuracy increased to 86%.

[4]The results were available before November 14th.