LICDIA
Laboratorio de Investigación en Ciencia de Datos & Inteligencia Artificial

Línea de I+D - Diálogo y RAG

Trabajo distinguido · CACIC 2026

JORNADAS 2026 DE CIENCIA, TECNOLOGÍA Y EXTENSIÓN
DEPARTAMENTO DE CIENCIAS BÁSICAS

Embeddings dinámicos de turnos para memoria conversacional en diálogo orientado a tareas

Juan Manuel Fernández[1], Marcelo Errecalde[2], Sergio Burdisso[3]

[1] LICDIA, Departamento de Ciencias Básicas, Universidad Nacional de Luján · [2] LIDIC, Universidad Nacional de San Luis · [3] Idiap Research Institute, Suiza

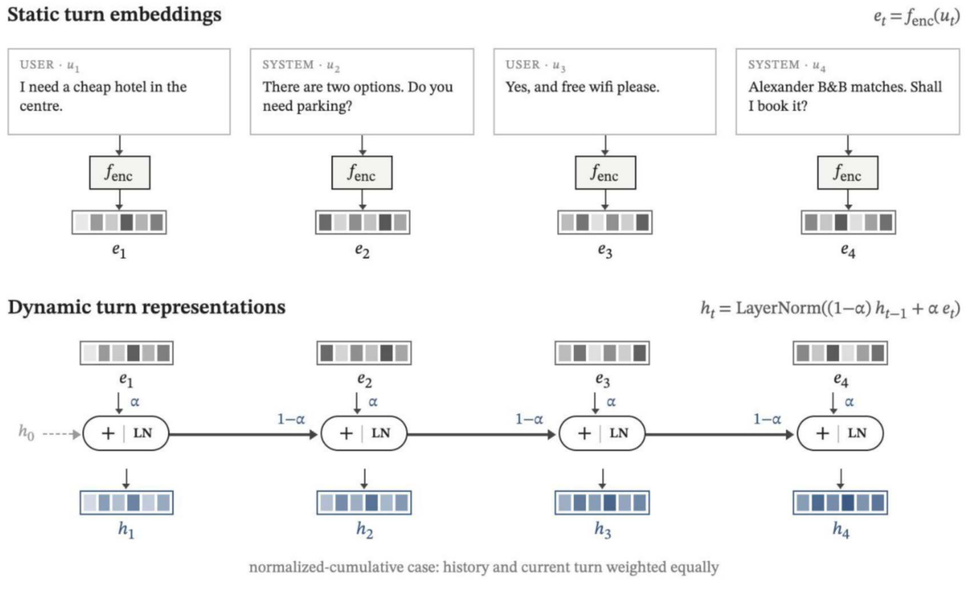
jmfernandez@unlu.edu.ar · merreca@unsl.edu.ar · sburdisso@idiap.ch

Acumular historia en el vector de cada turno mejora la recuperación solo en un espacio que ordena los turnos por función: en Dialog2Flow, MSS@10 sube de 3,29 a 3,66; en los generales no aporta mejoras.

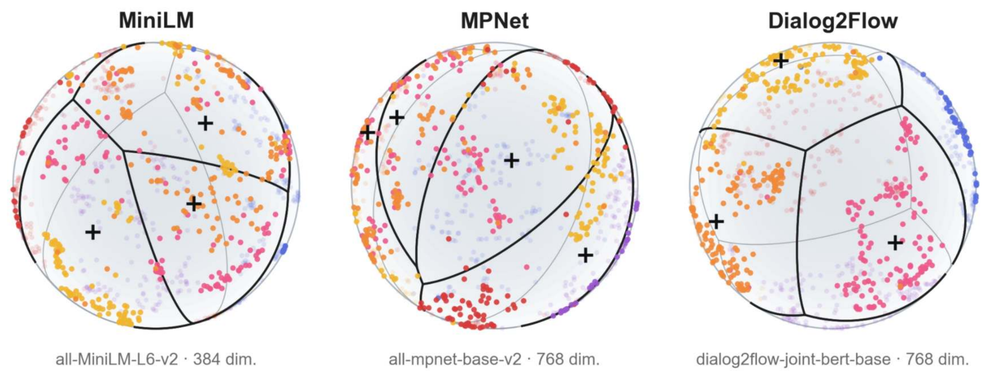
1 Motivación

Cada turno se representa con un vector que guía las decisiones del agente. Su significado depende de la historia, pero suele calcularse aislado. Como consulta a la memoria de otras conversaciones, ese vector debe acercar situaciones funcionalmente análogas.

2 Representaciones

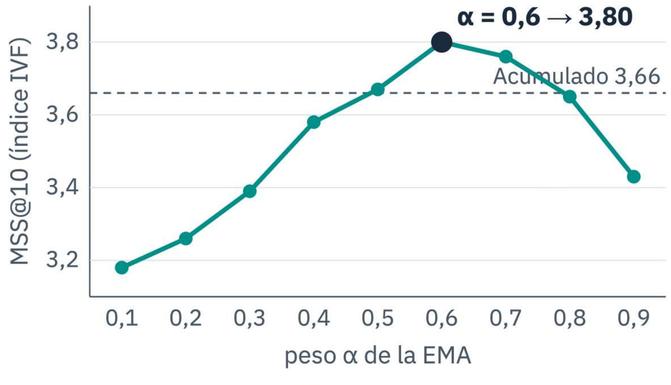


3 Geometría del espacio



Dialog2Flow ordena los turnos por función; MiniLM y MPNet, por contenido semántico.

4 Calibración EMA



Con α = 0,6 llega a 3,80, +0,13 sobre el acumulado.

5 Acumulado

Estática y acumulada, por espacio, con tres índices de búsqueda.

Espacio	Estática	Acumulada
MiniLM	3,35–3,42	3,23–3,33
MPNet	3,36–3,38	3,20–3,31
Dialog2Flow	**3,29–3,30**	**3,65–3,66**

Dialog2Flow gana entre 0,35 y 0,37 puntos con la acumulación; MiniLM y MPNet pierden entre 0,05 y 0,17.

MSS@10 (escala 1 a 5), rango entre IVF, HNSW e IVFPQ.

6 Conclusiones

ESPACIO
Depende del espacio
Acumular mejora sobre Dialog2Flow (3,29 a 3,66) y no aporta sobre los codificadores generales.

EMA
Mejora significativa
Con α = 0,6 supera al acumulado en los tres índices y en consultas no usadas para calibrar.

ALCANCE
Un espacio, un juez
Un único espacio conversacional y evaluación con un LLM como juez.

Sigue: aprender la actualización en lugar de fijarla con α, mediante un Transformer cuya unidad de entrada es el turno.

Referencias

Burdisso, S., Madikeri, S., Motlicek, P. Dialog2Flow: Pre-training Soft-Contrastive Action-Driven Sentence Embeddings for Automatic Dialog Flow Extraction. EMNLP 2024, pp. 5421–5440.

Fernández, J. M., Errecalde, M., Burdisso, S. Remembering the Flow: Retrieval with Dynamic Turn Representations for Conversational Memory. CACIC 2026.

Agradecimientos

Este trabajo se desarrolló en el marco del proyecto PCI-146 (IPAC), con recursos del SNCAD y de la infraestructura Clementina XXI, con el apoyo de la SICyT. Agradecemos a Nora Reyes y Edgar Chávez por su colaboración en búsqueda de proximidad en altas dimensiones.

Código y resultados

Líneas de I+D - Investigación aplicada en ciencia de datos e IA, con fuerte impronta en los problemas regionales

Modelos de lenguaje abiertos · IA en educación · **Diálogo y RAG** · Gobernanza y transformación digital · Blockchain y credenciales · Imágenes y datos geoespaciales

licdia.unlu.edu.ar