



ChessMatch-V2: Sistema de Recomendación para Ajedrez

Benjamín Varela¹ Francisco Meza² Vicente Navarro³

^{1,2,3}Escuela de Ingeniería, Departamento de Ciencia de la Computación, Pontificia Universidad Católica de Chile

Motivación

- Una de las formas más efectivas de mejorar en ajedrez es estudiar partidas de grandes jugadores que comparten un estilo de juego similar al propio. Esta idea ha sido explorada incluso por Grandes Maestros, dando origen al modelo "The 4 Player Types", el cual se utiliza para vender cursos y recursos personalizados de entrenamiento.
- Sin embargo, identificar tu propio estilo puede ser difícil y requerir mucho tiempo, especialmente considerando que la mayoría de los recursos actuales no toman en cuenta el estilo personal del jugador.
- Si bien es importante estudiar distintos tipos de partidas para estar preparado ante diversos estilos, enfocarse en aquellas que se asemejan al propio puede ser más motivador y favorecer un aprendizaje más efectivo y profundo.

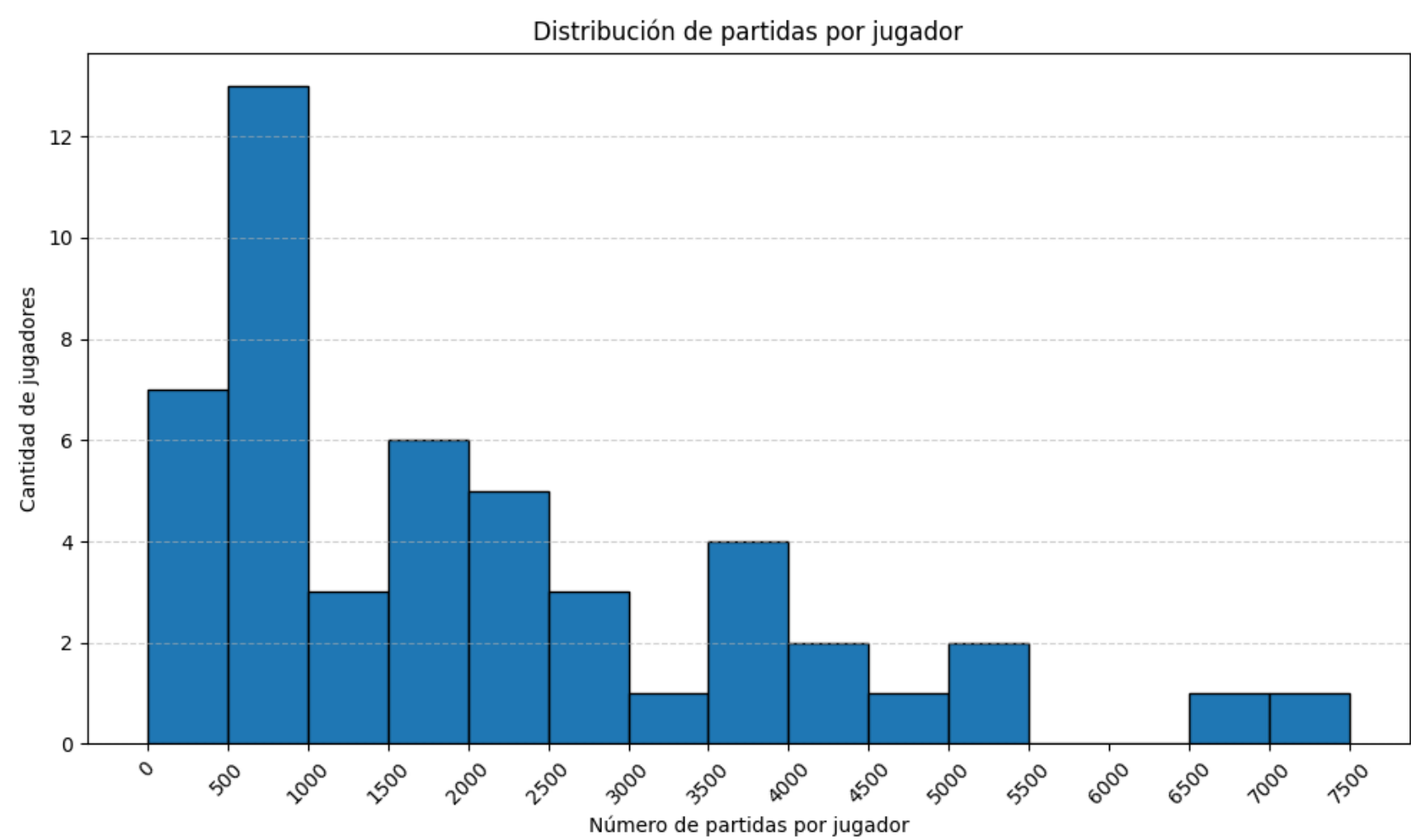
Objetivo del Proyecto

Desarrollar un sistema recomendador que, a partir de las partidas del usuario, identifique su estilo de juego, lo compare con jugadores profesionales y le recomiende partidas relevantes para su estudio.

El sistema busca facilitar un aprendizaje más eficiente y personalizado, alineado con la forma real en que juega cada persona.

Dataset & Preprocesamiento

- Se seleccionaron 50 jugadores considerados por **Chess.com** como los mejores de todos los tiempos.
- Se descargaron todas sus partidas registradas desde la base de datos pública **PGN Mentor**, obteniendo **99.897 partidas**.
- La longitud media de las partidas fue de 38 jugadas (19 movimientos por cada lado).

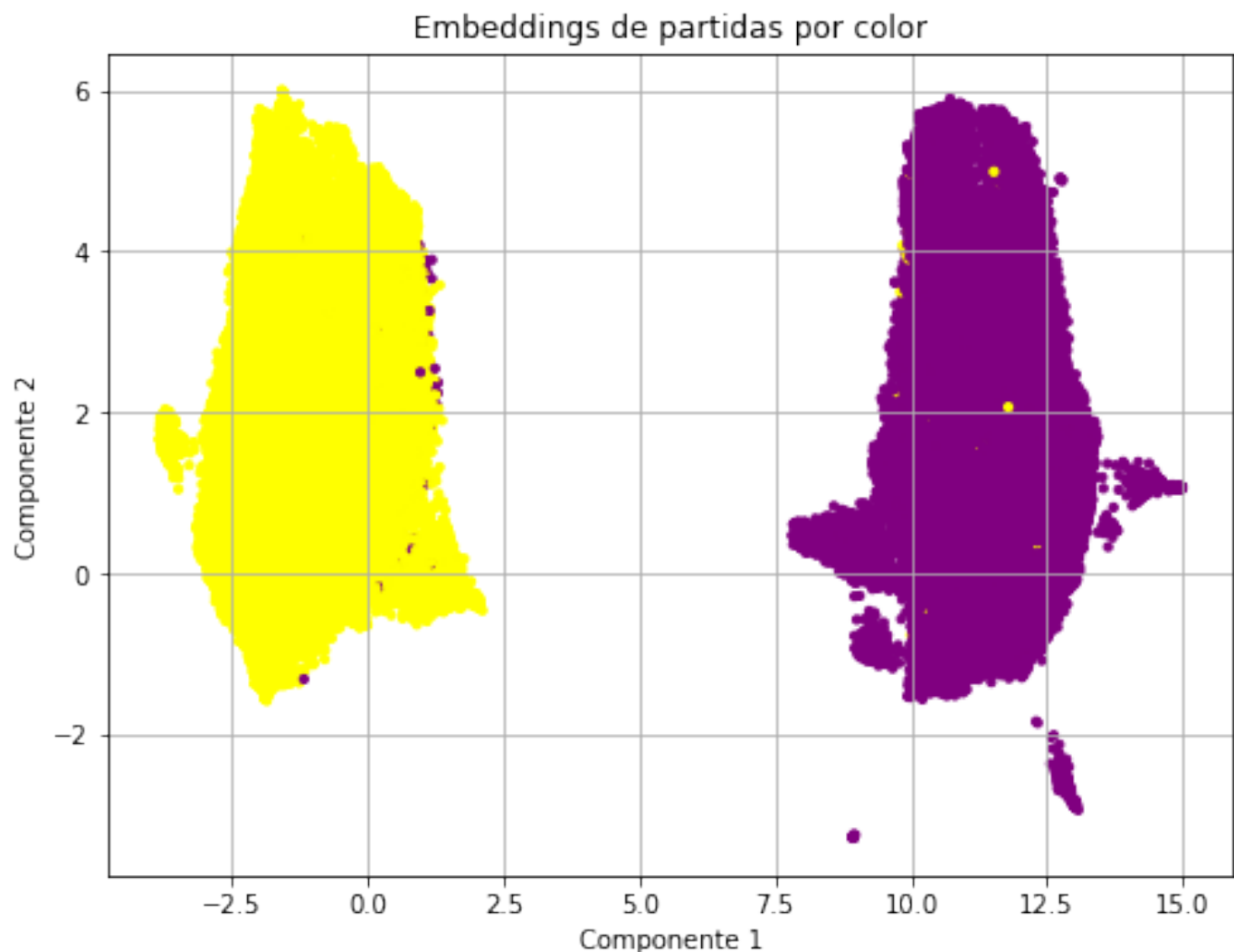


Metodología

Embeddings de Partidas

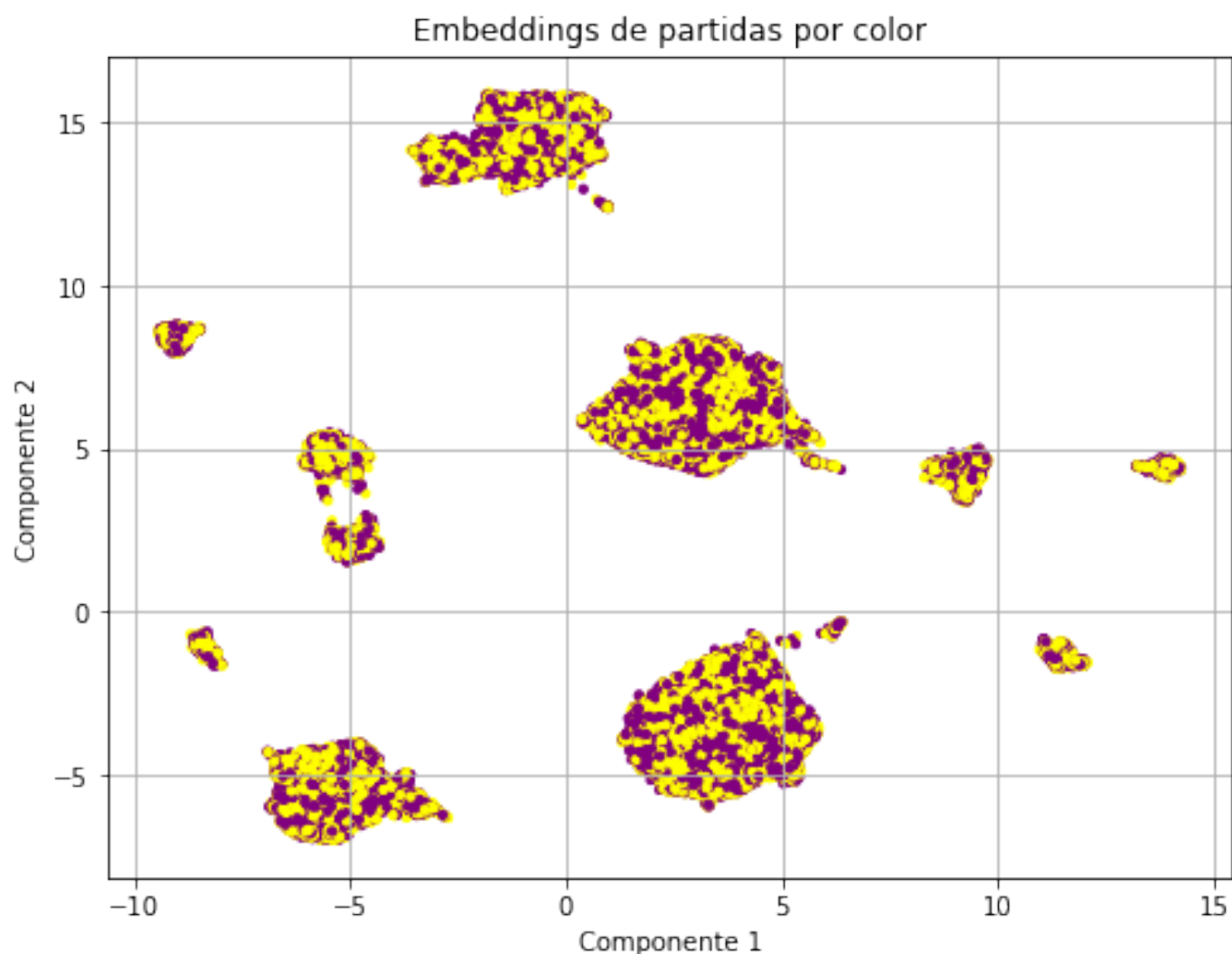
A partir de los pgn se extrajeron los movimientos del jugador de interés y se probaron 3 embeddings distintos:

Sentence-BERT: Modelo basado en BERT, preentrenado para generar embeddings semánticos. Representa los movimientos del jugador como una cadena de texto y captura relaciones lingüísticas entre jugadas.

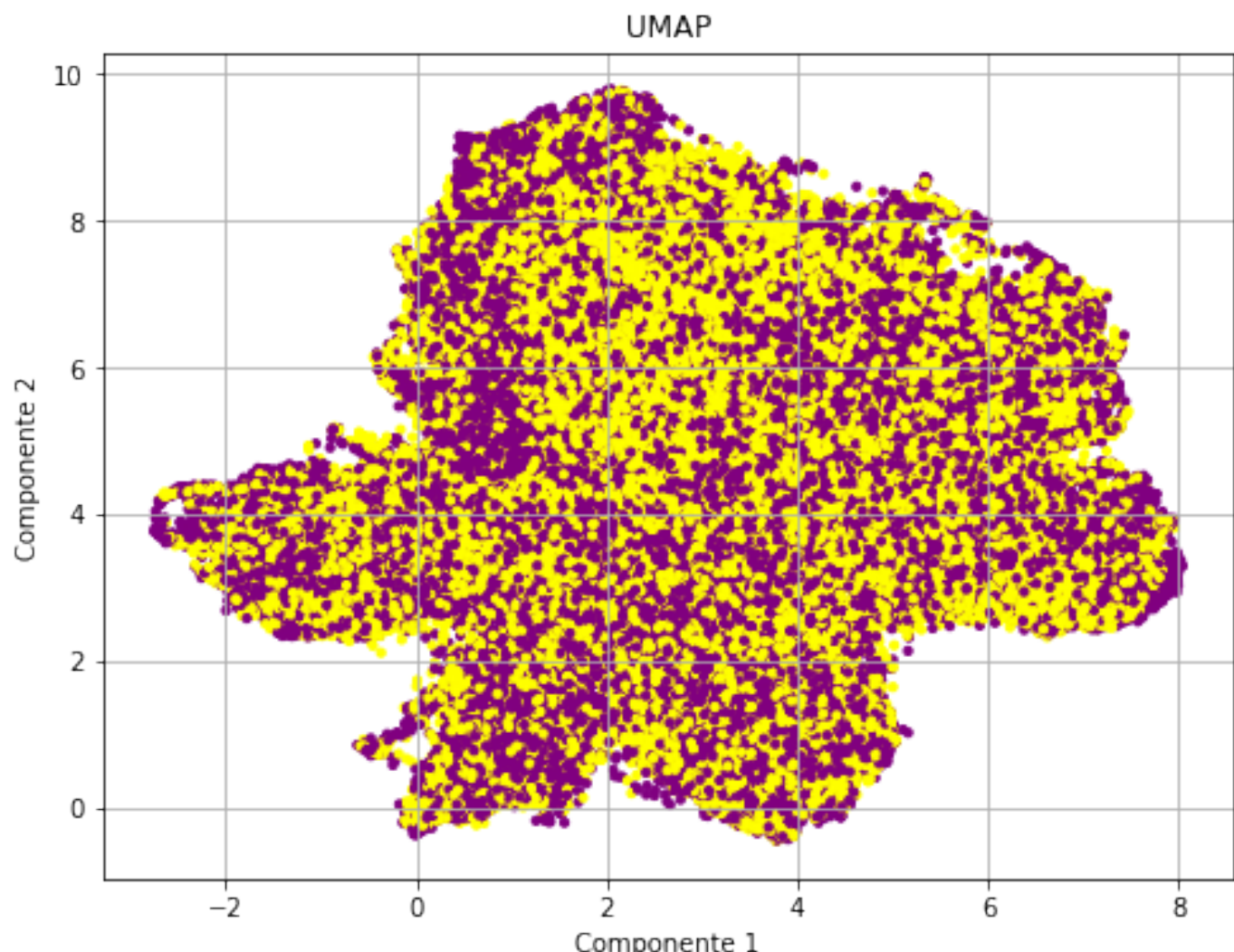


Continuación Metodología

e5-large-v2: Encoder preentrenado para recuperación semántica. Al igual que SBERT, representa las partidas como texto plano, generando vectores optimizados para comparar similitud semántica.



GNN: Arquitectura de red de grafos con pesos aleatorios reproducibles. Modela cada partida como un grafo de jugadas y genera embeddings basados en su estructura, sin entrenamiento supervisado.



Clustering para embedding del jugador

Se aplicó **K-Means** para agrupar las partidas según sus embeddings, probando distintas cantidades de clusters k , desde 2 hasta 60.

Luego, se construyó el embedding de cada jugador como un vector de longitud k que representa la distribución de sus partidas entre los clusters. Este vector actúa como una **representación latente del jugador**.

Por ejemplo, si $k = 2$ y un jugador tiene un 80 % de sus partidas en el cluster 0 y un 20 % en el cluster 1, su embedding sería: [0.8, 0.2].

Evaluación del mejor embedding para partidas & selección del valor óptimo de k

- División estratificada 80/20 a nivel de jugador (train/test).
- Predicción del jugador más similar utilizando *cosine similarity* entre vectores.
- Métrica de evaluación: **Accuracy**. Se considera un acierto si el vector más cercano en test corresponde al mismo jugador en train; en caso contrario, es incorrecto.

La evaluación busca medir si el embedding generado permite identificar correctamente al jugador al que pertenece un conjunto de partidas no vistas.

Clustering para estilo

- Con el mejor embedding y valor óptimo de k , se aplica clustering para agrupar jugadores por estilo.
- Se eliminan outliers y se evalúa consistencia con división train/test a nivel jugador.
- Se busca la máxima **accuracy**: el cluster de test debe coincidir con el de train para cada jugador.
- Posteriormente, se analiza e interpreta cada cluster para describir los estilos identificados.

Resultados

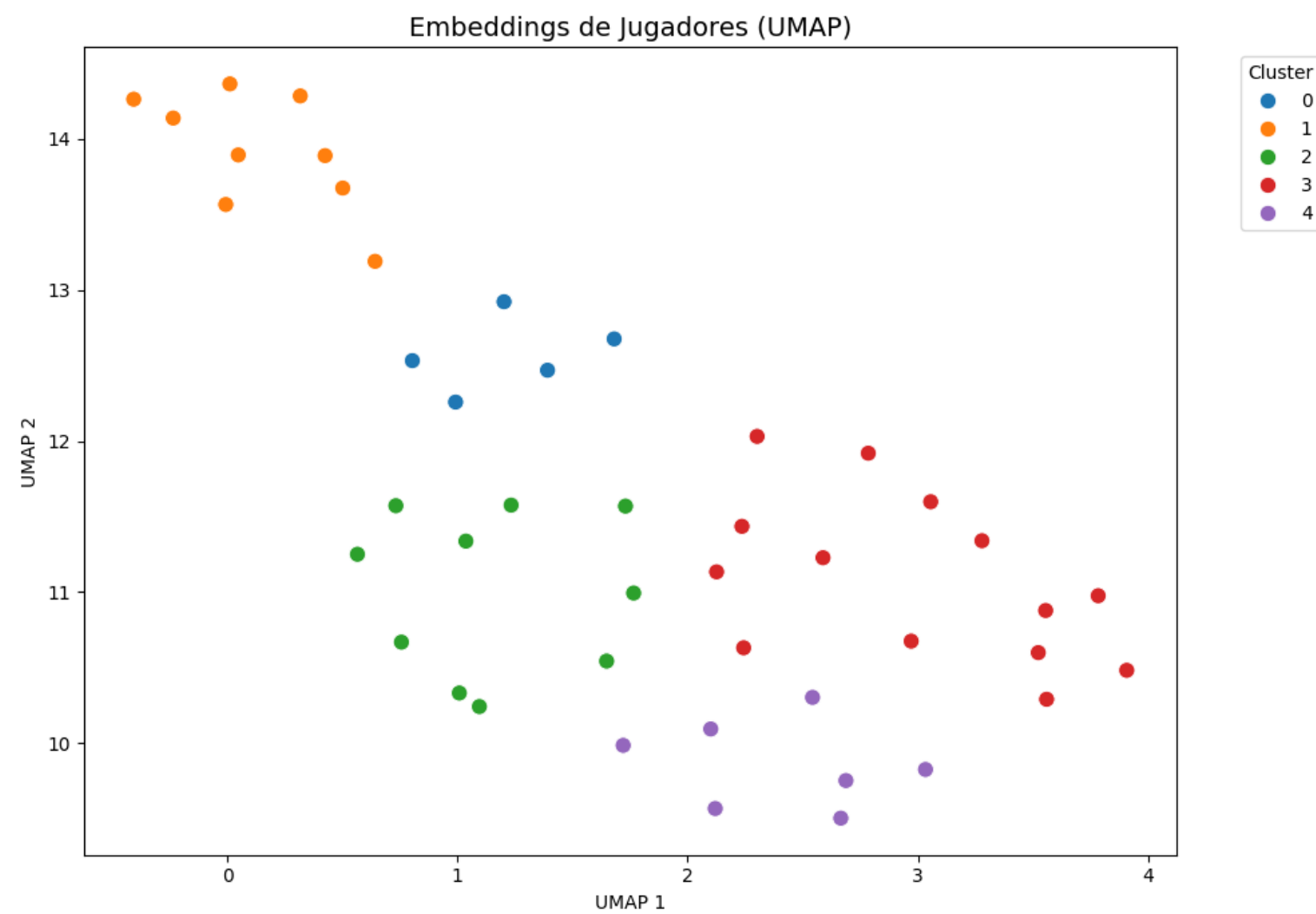
Embeddings de Partidas

Modelo	$k_{\text{ópt}}$	Accuracy (%)	Tamaño (MB)	Tiempo (H)
SBERT	30	69.4	1200	4
e5-large-v2	24	85.7	1200	6
GNN	56	95.9	157	0.8

- Sólo se usaron 49 jugadores para esta parte, ya que no tenía partidas válidas.
- El mejor modelo en todo sentido fue GNN y el k óptimo fue 56.

Cluster de estilos

- Se eliminaron 4 jugadores más, considerados outliers por Kmeans (principalmente por la baja cantidad de partidas), quedando con un total de 45 jugadores.
- Se obtuvo 100% de accuracy para el cluster con $k=5$



Con ayuda de una LLM se definieron los cluster de la siguiente manera:

- Cluster 0 – Vanguardia del Ajedrez Moderno:**
Ejemplos: Alekhine, Bogoljubow, Capablanca, Nimzowitsch, Rubinstein.
- Cluster 1 – Escuela Clásica y Fundacional:**
Ejemplos: Lasker, Steinitz, Tarrasch, Staunton, Chigorin.
- Cluster 2 – Maestros Contemporáneos del Siglo XXI:**
Ejemplos: Carlsen, Anand, Kramnik, Caruana, Topalov.
- Cluster 3 – Escuela Posicional Soviética:**
Ejemplos: Botvinnik, Karpov, Petrosian, Smyslov, Gelfand.
- Cluster 4 – Genios Creativos y Agresivos:**
Ejemplos: Fischer, Kasparov, Tal, Bronstein, Spassky.

Sistema Recomendador

El sistema recibe un archivo PGN, el nombre del usuario, un número n (jugadores a recomendar) y un número m (partidas a recomendar por jugador).

- Se lee el PGN y se generan embeddings con GNN para cada partida del usuario.
- Cada partida se clasifica usando el modelo K-Means de partidas previamente entrenado, y se construye el embedding del usuario.
- Se asigna al usuario un estilo mediante el modelo K-Means de jugadores, junto con una descripción del cluster correspondiente.
- Se calcula la similitud (coseno) entre el embedding del usuario y el de todos los jugadores, recomendando los n más similares.
- Adicionalmente, se promedian los embeddings de las partidas del usuario y se recomiendan las m partidas más similares (una por jugador recomendado).

Ejemplo

```
make_recommendations("felipepollo", "lichess/liches/felipepollo/2025-06-29.pgn", 1, 1)

...
Leyendo partidas...
Utilizando 388 partidas de felipepollo
Generando embeddings: 100% 106/106 [08:01:00-00, 95.01it/s]

Tu estilo es:
Cluster 3 - "Escuela Posicional Soviética"
Descripción: Jugadores de la tradición soviética. Muy técnicos, con enfoque estratégico y profundo conocimiento del medio juego.
Estilo: Posicional, estructural, metódico.
Ejemplos: Botvinnik, Karpov, Petrosian, Smyslov, Gelfand

Jugadores similares recomendados:
(1) Korchnoi con 0.81 de similitud

Partidas recomendadas:
[Event "4th Piatomsky Bank Chess Cup"]
[Site "Odessa URS"]
[Date "2008-05-31"]
[Round "2"]
[White "Nica,V"]
[Black "Korchnoi,V"]
[Result "1/2-1/2"]
[WhiteToMove "2/2"]
[BlackToMove "2/2"]
[ECO "B07"]

1. Bf3 a6 2. d4 d5 3. c4 Bf6 4. Nc3 c5 5. e3 Nc6 6. cxd5 Bxd5 7. Bd3 cxd4 8. exd4 Bb4 9. Bd2 Re7 10. a3 Bf6 11. Bc3 Bxc3 12. bxc3 c5 13. dxc5 Bxc5 14. Bc2
```

Referencias

[1] Reimers, N., Gurevych, I. (2019). *Sentence BERT: Sentence Embeddings using Siamese BERT Networks*. EMNLP.

[2] Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., Majumder, R., Wei, F. (2024). *Text Embeddings by Weakly Supervised Contrastive Pre training*. arXiv:2212.03533.

[3] Veličković, P. (2023). *Everything is Connected: Graph Neural Networks*. Patterns, 4(9), 100639.