Fernández Amor, J. A., Gallego Fontenla, V. J., Gamallo Fernández, P., Lama Penín, M., Vidal Aguiar, J. C.: Refinando modelos mediante aprendizaje por refuerzo profundo para la predicción de sufijos. In: Fabra, J. (ed.) Actas de las XXI Jornadas de Ciencia e Ingeniería de Servicios (JCIS 2026). SISTEDES (2026). https://hdl.handle.net/11705/JCIS/2026/32

Refinando modelos mediante aprendizaje por refuerzo profundo para la predicción de sufijos

José A. Fernández, Pedro Gamallo-Fernandez, Victor Gallego-Fontenla, Juan C. Vidal, and Manuel Lama

Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidad de Santiago de Compostela, España
{joseandres.fernandez.amor,pedro.gamallo.fernandez,victorjose.gallego, juan.vidal,manuel.lama}@usc.es

Resumen Anticipar el comportamiento de un proceso es clave para optimizar la toma de decisiones y la eficiencia operativa. Los enfoques tradicionales de predicción de sufijos modelan la probabilidad del siguiente evento de forma local, lo que puede generar incoherencias en la traza completa. En este trabajo en desarrollo proponemos un esquema en dos fases: (i) entrenamiento supervisado de un modelo base para la predicción del siguiente evento y (ii) su refinamiento mediante aprendizaje por refuerzo profundo, donde la recompensa optimiza directamente la calidad global del sufijo generado. Los resultados experimentales muestran que el refinamiento mejora la calidad global de las predicciones, proporcionando una modelización más coherente a nivel de secuencia completa.

Keywords: Predicción de sufijos · Minería de procesos

1. Introducción

La *minería de procesos* permite extraer conocimiento a partir de logs de eventos para analizar y mejorar los procesos de negocio. Dentro de este ámbito, la *monitorización predictiva* se centra en anticipar el comportamiento futuro de instancias en ejecución, incluyendo tareas como la predicción del tiempo restante, la siguiente actividad o, como en este trabajo, la predicción de sufijos.

Dada la naturaleza secuencial del problema, la mayoría de los enfoques del estado del arte emplean arquitecturas recurrentes (p.ej., LSTM) [2], entrenadas para maximizar la probabilidad del siguiente evento dado el prefijo observado. Este tipo de entrenamiento optimiza las decisiones locales en cada paso de la secuencia. Sin embargo, cuando las predicciones se encadenan para generar un sufijo completo, los errores pueden acumularse, produciendo secuencias globalmente inconsistentes. Estas limitaciones sugieren la necesidad de mecanismos de entrenamiento capaces de optimizar la calidad de la secuencia generada como un todo, en lugar de centrarse únicamente en decisiones locales.

En este trabajo partimos de la hipótesis de que la coherencia global de los sufijos predichos puede mejorarse mediante un esquema de aprendizaje en dos

Esta obra está bajo una licencia Creative Commons “Atribución-NoComercial-SinDerivadas 4.0 Internacional”.

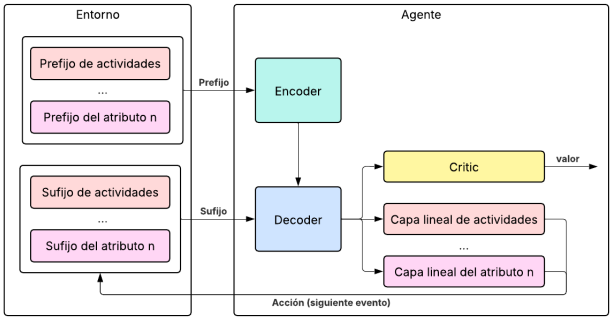

Figura 1. Arquitectura del modelo refinado

fases. En primer lugar, se entrena un modelo base mediante aprendizaje supervisado para predecir la siguiente actividad a partir del prefijo observado, lo que permite capturar dependencias temporales locales y proporciona una política inicial estable. En una segunda fase, este modelo se refina mediante aprendizaje por refuerzo profundo, optimizando directamente una función de recompensa definida sobre la secuencia generada. De este modo, el proceso de aprendizaje pasa de optimizar decisiones locales a optimizar la calidad global del sufijo completo.

2. Propuesta

Nuestro enfoque se basa en un esquema de aprendizaje en dos fases. En la primera etapa se entrena un modelo base para la predicción del siguiente evento basado en la arquitectura propuesta en [3] con una cabeza de salida para cada atributo mediante aprendizaje supervisado. Este entrenamiento permite capturar las dependencias temporales locales del proceso y proporciona un modelo inicial capaz de generar predicciones plausibles. En la segunda etapa, el modelo previamente entrenado se refina mediante aprendizaje por refuerzo profundo.

2.1. Refinamiento

Una vez entrenado el modelo base, comienza el refinamiento por aprendizaje por refuerzo. La figura 1 muestra el esquema general del modelo en esta etapa. El modelo previamente entrenado se utiliza como inicialización del componente generador de acciones del algoritmo Proximal Policy Optimization (PPO), lo que permite partir de una política ya entrenada mediante supervisión.

El *entorno* contiene la información disponible en cada momento de la predicción: el prefijo observado y el sufijo generado. En cada iteración, el agente selecciona una *acción*, que consiste en predecir el siguiente evento de la secuencia. Esta predicción se incorpora al sufijo generado, actualizando así el estado del entorno. Con el objetivo de evaluar la calidad de la secuencia completa y no únicamente de cada decisión individual, se define una recompensa diferida: se asigna

Fernández Amor, J. A., Gallego Fontenla, V. J., Gamallo Fernández, P., Lama Penín, M., Vidal Aguiar, J. C.: Refinando modelos mediante aprendizaje por refuerzo profundo para la predicción de sufijos. In: Fabra, J. (ed.) Actas de las XXI Jornadas de Ciencia e Ingeniería de Servicios (JCIS 2026). Sistedes (2026). https://hdl.handle.net/11705/JCIS/2026/32

Esta obra está bajo una licencia Creative Commons “Atribución-NoComercial-SinDerivadas 4.0 Internacional”.

Tabla 1. Comparativa entre el modelo base y el modelo refinado mediante aprendizaje por refuerzo, evaluados mediante la similitud normalizada de Damerau–Levenshtein.

Model	BPI12	BPI12A	BPI12C	BPI12O	BPI12W	BPI12WC	BPI13CP	BPI13I	EnvPerm	Helpdesk	Nasa	Sepsis
Modelo refinado	**0,427**	**0,646**	**0,463**	**0,606**	**0,396**	**0,324**	**0,706**	0,593	**0,849**	**0,901**	**0,513**	**0,416**
Modelo base	0,410	0,580	0,415	0,573	0,376	0,317	0,701	**0,600**	0,845	0,889	0,488	0,405

Tabla 2. Resultados del test bayesiano jerárquico entre el modelo base (B) y el modelo refinado (R) utilizando un ROPE = [-0.01, 0.01]

Model	BPI12	BPI12A	BPI12C	BPI12O	BPI12W	BPI12WC	BPI13CP	BPI13I	EnvPerm	Helpdesk	Nasa	Sepsis
$B < R$	**0,797**	**0,984**	**0,979**	**0,906**	**0,966**	0,422	0,234	0,015	0,036	**0,697**	**0,905**	**0,603**
$B = R$	0,196	0,015	0,019	0,087	0,034	**0,560**	**0,757**	**0,874**	**0,963**	0,302	0,090	0,365
$B > R$	0,007	0,001	0,001	0,007	0,001	0,018	0,009	0,111	0,001	0,001	0,005	0,032

una recompensa nula en los pasos intermedios y, una vez finalizado el episodio, se devuelve como recompensa la similitud normalizada de Damerau–Levenshtein entre el sufijo generado y el sufijo real.

Dado que PPO sigue un esquema *actor–critic*, el modelo generativo previamente entrenado actúa como el *actor*, encargado de seleccionar las acciones (esto es, predecir el siguiente evento de la secuencia), mientras que se introduce un segundo componente encargado de estimar el valor del estado. En nuestra implementación, el *crítico* consiste en una capa lineal que toma como entrada la salida del *decoder*, interpretada como la representación del estado actual.

Sin embargo, dado que este componente se inicializa sin entrenamiento previo, sus estimaciones iniciales del valor del estado son esencialmente aleatorias. Si ambos componentes se entrenasen conjuntamente desde el inicio, estas estimaciones inestables podrían introducir señales de actualización ruidosas y degradar el conocimiento previamente adquirido durante el entrenamiento supervisado, produciendo un fenómeno de *olvido catastrófico*. Para mitigar este efecto, los parámetros del actor se congelan durante 160 actualizaciones (8 rollouts), permitiendo que el crítico aprenda primero estimaciones más estables antes de comenzar el proceso conjunto de optimización.

3. Evaluación

Para evaluar el modelo se utilizaron los logs y métricas propuestos en [2], que incluyen una amplia variedad de datasets reales con distintas características, así como un esquema de evaluación basado en la similitud normalizada de Damerau–Levenshtein, que mide la calidad global de la traza generada.

Fernández Amor, J. A., Gallego Fontenla, V. J., Gamallo Fernández, P., Lama Penín, M., Vidal Aguiar, J. C.: Refinando modelos mediante aprendizaje por refuerzo profundo para la predicción de sufijos. In: Fabra, J. (ed.) Actas de las XXI Jornadas de Ciencia e Ingeniería de Servicios (JCIS 2026). SISTEDES (2026). https://hdl.handle.net/11705/JCIS/2026/32

Esta obra está bajo una licencia Creative Commons “Atribución-NoComercial-SinDerivadas 4.0 Internacional”.

Fernández Amor, J. A., Gallego Fontenla, V. J., Gamallo Fernández, P., Lama Penín, M., Vidal Aguiar, J. C.: Refinando modelos mediante aprendizaje por refuerzo profundo para la predicción de sufijos. In: Fabra, J. (ed.) Actas de las XXI Jornadas de Ciencia e Ingeniería de Servicios (JCIS 2026). Sistedes (2026). https://hdl.handle.net/11705/JCIS/2026/32

Los resultados de la comparativa se detallan en la Tabla 1. Para comparar ambos modelos, se aplicó un test estadístico jerárquico bayesiano [1]. Los resultados de la tabla 2 muestran que el modelo refinado presenta una probabilidad posterior elevada de superar al modelo base ($P(R > B) > 0{,}90$) en la mayoría de los datasets. En particular, en BPIC12, BPIC12A, BPIC12C, BPIC12O, BPIC12W, Helpdesk y Sepsis, la probabilidad de mejora es superior al 0.95, indicando evidencia estadística fuerte a favor del refinamiento mediante aprendizaje por refuerzo.

En algunos casos (p.ej., BPIC13CP, BPIC13I y EnvPermit), la mayor parte de la probabilidad se concentra en valores cercanos a cero, lo que sugiere una diferencia prácticamente nula entre ambos modelos. No se observa evidencia de que el modelo base supere al refinado en ningún dataset, ya que $P(B > R)$ es sistemáticamente cercana a cero, incluso en BPIC13I, que es el único dataset donde la media del modelo base fue ligeramente superior.

Estos resultados sugieren que el aprendizaje por refuerzo mejora de forma consistente la calidad global de las predicciones de sufijo, manteniendo estabilidad en aquellos contextos donde la mejora es marginal.

4. Conclusiones

Este trabajo propone un esquema de refinamiento mediante aprendizaje por refuerzo profundo para mejorar la predicción de sufijos en minería de procesos. Los resultados muestran que optimizar directamente la calidad global de la secuencia permite mejorar de forma consistente el modelo base, con evidencia estadística robusta en la mayoría de los datasets analizados. Como trabajo futuro, se plantea aplicar esta estrategia a modelos predictivos del estado del arte para evaluar su capacidad de generalización y su adecuación a las particularidades de la minería de procesos.

Agradecimientos Este trabajo ha recibido financiación de la Agencia Estatal de Investigación (acreditación PID2023-149549NB-I00 y PDC2025-166312-I00), de la Xunta de Galicia-Consellería de Educación, Ciencia, Universidades e Formación (acreditación Centro de Investigación de Galicia 2024-2027 ED431G-2023/04) y de la Unión Europea (Fondo Europeo de Desarrollo Regional–FEDER).

Referencias

1. Gamallo-Fernandez, P., Rama-Maneiro, E., Vidal, J.C., Lama, M.: Verona: A python library for benchmarking deep learning in business process monitoring. SoftwareX **26**, 101734 (2024). `https://doi.org/https://doi.org/10.1016/j.softx.2024.101734`
2. Rama-Maneiro, E., Vidal, J.C., Lama, M.: Deep learning for predictive business process monitoring: Review and benchmark. IEEE Transactions on Services Computing **16**(1), 739–756 (2023). `https://doi.org/10.1109/TSC.2021.3139807`
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. CoRR **abs/1706.03762** (2017), `http://arxiv.org/abs/1706.03762`

Esta obra está bajo una licencia Creative Commons “Atribución-NoComercial-SinDerivadas 4.0 Internacional”.