

IDIAPers @ Causal News Corpus 2022: Extracción de causa-efecto-señal

Trillizos a través del modelo de lenguaje autorregresivo preentrenado

Martín Fajcik*, 1, 2, Muskaan Singh¹, Juan Zuluaga-Gómez^{1,3}, Esaú Villatoro-Tello^{1,4}, Sergio Burdisso^{1,5}, Petr Motlicek^{1, 2}, Pavel Smrz²

¹Idiap Research Institute, Martigny, Suiza 2Brno
University of Technology, Brno, República Checa 3Ecole
Polytechnique Fédérale de Lausanne, Suiza 4Universidad
Autónoma Metropolitana Unidad Cuajimalpa, Ciudad de México, México
5Universidad Nacional de San Luis (UNSL), San Luis, Argentina

*autor para correspondencia: martin.fajcik@vut.cz

Abstracto

En este documento, describimos nuestras presentaciones de tareas compartidas para la subtask 2 en CASE-2022, Identificación de causalidad de eventos con Casual News Corp. El desafío se centró en la detección automática de todos los tramos de causa-efecto-señal presentes en la oración de los medios de comunicación. Detectamos lapsos de causa-efecto-señal en una oración usando T5, un modelo de lenguaje autorregresivo previamente entrenado. Identificamos iterativamente todos los tripletes de intervalo de señal de causa-efecto, siempre condicionando la predicción del siguiente triplete a los previamente predichos. Para predecir el triplete en sí, consideramos diferentes relaciones causales como causa→efecto→señal. Cada componente del triplete se genera a través de un modelo de lenguaje condicionado por la oración, las partes anteriores del triplete actual y los tripletes previamente predichos. A pesar del entrenamiento en un conjunto de datos extremadamente pequeño de 160 muestras, nuestro enfoque logró un desempeño competitivo, quedando segundo en la competencia. Además, mostramos que asumiendo que el orden causa→efecto o efecto→causa logra resultados similares.¹

1. Introducción

La causalidad vincula la relación entre dos argumentos: causa y efecto (Barik et al., 2016).

La Figura 1 muestra ejemplos extraídos del Causal News Corpus (CNC) (Tan et al., 2022b). Las cláusulas de causa aparecen en amarillo, las de efecto en verde y las de señal en rosa; en lo sucesivo denominados **trillizos CES**. Como se muestra en el ejemplo, "el bombardeo generó pánico entre los aldeanos", ilustra que el evento "bombardeo" provocó el evento "pánico entre los aldeanos" denominado efecto. El vínculo entre la causa y el efecto, es decir, la palabra "creó", se denomina señal y puede expresarse explícita o implícitamente.

¹Código en <https://github.com/idiap/cncsharedtask>.

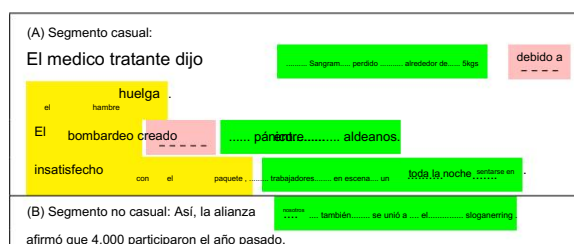


Figura 1: Ejemplos de Causal News Corpus, las causas están en amarillo, efectos en verde y señales en rosa. Si una oración tiene tanto causa como efecto, se la denomina casual (A), de lo contrario, no casual (B).

La detección y extracción automática de relaciones de causalidad desempeña un papel vital en muchos trabajos de procesamiento del lenguaje natural (PLN) para abordar la inferencia y la comprensión (Dunietz et al., 2020; Fajcik et al., 2020; Jo et al., 2021; Feder et al., 2021a). Tiene aplicaciones en varias tareas de PLN posteriores, a saber, generación de preguntas y respuestas causales, explicación del comportamiento de las redes sociales, fenómenos políticos, educación efectiva y sesgo de género en la comunidad de investigación (Tan et al., 2014; Wood Doughty et al., 2018; Sridhar y Getoor, 2019; Veitch et al., 2020; Zhang et al., 2020; Feder et al., 2021b).

En este documento, describimos nuestra metodología para la tarea compartida de detección de intervalo de señal de causa-efecto-CASO-2022 (Subtask 2). En general, nuestras principales contribuciones se enumeran a continuación:

1. Mostramos que los lapsos de causa-efecto-señal se pueden extraer mediante un modelo seq2seq generativo preentrenado simple entrenado en solo 160 instancias.
2. Desarrollamos un método para extraer todos los tripletes causales de la oración en un hombre iterativo ner.
3. Investigamos cómo los modelos de lenguaje tratan

el orden causal de la causa y el efecto se extiende para responder a la pregunta de investigación "¿debe identificarse primero la causa y solo después el efecto, o viceversa?".

4. Mostramos que se puede aplicar un algoritmo eficiente de coincidencia de la mejor subcadena F1, conocido por preguntas y respuestas, para tratar casos raros en los que un modelo de lenguaje (LM) no genera parte de la secuencia de entrada.

2. Trabajo relacionado

El problema de la extracción de causalidad del texto es una tarea desafiante, ya que requiere comprensión semántica y conocimiento contextual. Hubo muchos intentos en el dominio de la lingüística para la creación de corpus para la extracción de eventos pero con un tamaño limitado, como CausalTimeBank (CTB) (Mirza et al., 2014) de noticias con 318 pares, CaTeRS (Mostafazadeh et al., 2016) de historias cortas con 488 enlaces casuales, EventStoryLine (Caselli y Vossen, 2017) de artículos de noticias en línea con 1.770 pares de eventos casuales, corpus de relación semántica PDTB-3 (Webber et al., 2019) con más de 7.000 relaciones causales y corpus CNC (Tan et al., 2022b,c) con 1957 eventos casuales con múltiples pares de eventos. En comparación con conjuntos de datos anteriores, CNC se diferencia porque se centra en oraciones de eventos, acepta argumentos que no necesitan formar una cláusula y no se limita a una lista predefinida de conectores, sino que incluye ejemplos causales en construcciones lingüísticas más variadas. El trabajo previo en este dominio se puede clasificar en términos generales en enfoques basados en el conocimiento, aprendizaje automático estadístico y enfoques basados en el aprendizaje profundo. El enfoque basado en el conocimiento utiliza patrones lingüísticos mediante la predefinición de palabras clave o manuales (García et al., 1997; Khoo et al., 2000; Radinsky et al., 2012; Beamer et al., 2008; Girju et al., 2009); ; Ittoo y Bouma, 2013; Kang et al., 2014; Khoo et al., 1998; Bui et al., 2010).

Las técnicas estadísticas (Girju, 2003; Do et al., 2011) se basan en la construcción de modelos probabilísticos sobre características extraídas a través de herramientas de PNL de terceros como Wordnet (Miller, 1994). Las técnicas de aprendizaje profundo mapean palabras y características en vectores densos de baja dimensión, lo que puede aliviar el problema de escasez de características. Los modelos de secuencia a secuencia utilizados con mayor frecuencia son redes feed-forward (Ponti y Korhonen, 2017), redes de memoria a largo plazo (Kruengkrai et al., 2017; Dasgupta et al., 2018; Martínez-Cámara et al., 2017) redes neuronales convolucionales (Jin et al., 2020; Kruengkrai et al.,

2017; Wang et al., 2016), redes neuronales recurrentes (Yao et al., 2019), unidades recurrentes cerradas (Chen et al., 2016) que incorporan información semántica y sintáctica en secuencias de palabras consecutivas locales (Yao et al., 2019). Modelo de entrenamiento no supervisado posterior como BERT (Devlin et al., 2018; Sun et al., 2019), RoBERTa (Becquin, 2020), red de convolución de gráficos (Zhang et al., 2018), redes de atención de gráficos y modelo conjunto para extracción de relación de entidad (Li et al., 2017; Wang y Lu, 2020; Zhao et al., 2021; Bekoulis et al., 2018).

En este trabajo, basamos nuestro modelo en T5 (Raffel et al., 2020), un modelo transformador de secuencia a secuencia, entrenado previamente en una combinación de objetivo de eliminación de ruido y 25 tareas supervisadas como traducción automática, aceptabilidad lingüística, resumen abstracto o respuesta a preguntas. El objetivo de eliminación de ruido no supervisado reemplaza aleatoriamente tramos de la entrada con diferentes tokens de máscara y genera contenidos de estos tramos enmascarados con el prefijo de estos tokens de máscara especiales. Además, nuestro trabajo comparte similitudes con el marco generativo basado en red de punteros (Vinyals et al., 2015) para varias sub tareas NER introducidas por Yan et al. (2021).

Por el contrario, nuestro trabajo se adapta más a escenarios de bajos recursos, ya que no se agregaron parámetros adicionales a nuestro sistema, a costa de errores, que pueden ocurrir en el paso de coincidencia de posprocesamiento.

3 Descripción del problema

Desafío de tarea compartida CASE-2022 (Tan et al., 2022a) destinado a la identificación de causalidad de eventos y extracción en corpus de noticias casuales (Tan et al., 2022b). Comprendería de dos sub tareas, a saber, la clasificación de eventos casuales (Subtarea 1) y la detección de intervalo de señal de causa-efecto (Subtarea 2)2. La subtarea 2 tiene como objetivo extraer los tramos correspondientes a los tripletes de la señal de causa-efecto (CES), como se muestra en la Figura 1. Entrenamos un modelo generativo seq2seq para abordar este desafío y extrajimos los tripletes CES mediante un procedimiento iterativo (consulte la Sección 4.1).

Las estadísticas del conjunto de datos se presentan en la Tabla 1. El número total de oraciones está dado por la columna #Sentencias, mientras que el número total de tripletes CES está en la columna #Relaciones. La columna #Signals muestra cuántas anotaciones de señales estaban presentes en el número total de trillizos CES.

²Participamos en ambas sub tareas, pero informamos sobre la Subtarea 2 en este documento. Para la Subtarea 1, remitimos al lector a nuestra publicación independiente (Burdizzo et al., 2022).

Dividir	#Oraciones	#Relaciones	#Señales
Tren	160	183	118 (64%)
Desarrollo	15	18	10 (56%)
prueba	89	119	98 (82%)

Tabla 1: Estadísticas del conjunto de datos. Ver texto para más detalles.

4 Metodología

4.1 Entrenamiento del modelo de

lenguaje Utilizamos T5 (Raffel et al., 2020), un modelo de lenguaje basado en un transformador autorregresivo entrenado previamente en una combinación de tareas supervisadas y no supervisadas que requieren comprensión del lenguaje. El modelo se condiciona $n \times 3$ veces para cada ejemplo, ya que puede haber n tripletes CES en una oración (hasta $n = 4$ tripletes en datos de entrenamiento). Cada vez, condicionamos el modelo de lenguaje 3 veces para cada ejemplo y su triplete CES correspondiente, generando un componente de triplete diferente (causa, efecto y señal) para aprender a generar el triplete CES completo. Como estos tripletes no están ordenados, muestreamos uniformemente una ruta aleatoria entre ellos (por ejemplo, 2-3-1-4, para una muestra con cuatro tripletes) durante el entrenamiento. Solo entrenamos con tantos trillizos como estén disponibles en los datos de entrenamiento. Ahora describimos el formato de entrada, más ilustrado en el Apéndice B.

En primer lugar, el codificador del modelo se condiciona con tokens de oraciones <frase> seguido del historial de trillizos CES ya generados para este ejemplo (vacío si no hubo ninguno) como

```
<frase> _historia : <historia>.
```

El historial siempre se antepone con _history: tokens. El contenido de la historia son los tripletes ya generados. Cada parte del triplete se antepone con su correspondiente _causa:, o _efecto:, o _señal: secuencia. Al mismo tiempo, el decodificador del modelo tiene el prefijo _causa: secuencia. En este caso, la probabilidad de secuencia de causa se maximiza.

En segundo lugar, el modelo está condicionado con tokens de oración <oración> y tokens de causa

```
<causa>, precedido de _causa: token como
```

```
<frase> _causa : <causa>
_historia : <historia>.
```

Esta vez, se solicita al decodificador el prefijo _effect: y se maximiza la probabilidad de secuencia de efecto.

En tercer lugar, el modelo está condicionado con fichas de oración <oración>, fichas de causa <causa> y fichas de efecto <efecto> con _efecto: ficha antepuesta como

```
<frase> _causa : <causa> _efecto
: <efecto> _historia : <historia>.
```

De forma análoga, se solicita al decodificador con _signal: prefijo y se maximiza la probabilidad de secuencia de señal. Como es posible que la señal no siempre sea parte del triplete CES, dejamos que el modelo genere un token _empty en estos casos.

4.2 Detalles experimentales

Usamos la pérdida de entropía cruzada (CE) para entrenar el T5. En primer lugar, promediamos la pérdida de CE en tokens, luego en puts por ejemplo (para todos los tripletes de CES) y luego en mini lotes. Utilizamos la búsqueda voraz para generar las secuencias. En el tiempo de inferencia, siempre generamos 4 tripletes CES para cada oración, ya que ese es el máximo que observamos en los datos de entrenamiento.

Como no restringimos la decodificación, la secuencia generada no tiene que coincidir con cierta subcadena en la entrada. Sin embargo, la tarea extractiva requiere la inserción de etiquetas alrededor de una causa, efecto o intervalo de señal dentro de la oración de entrada. Por lo tanto, asignamos las secuencias generadas a la oración de entrada a través de la coincidencia F1. En particular, para cada secuencia generada, encontramos la subcadena más similar en la entrada, donde la similitud se mide a través de la puntuación F1 a nivel de token. Utilizamos una técnica de coincidencia F1 eficiente, que elimina una parte significativa del espacio de búsqueda, que se presenta en el Apéndice C.1 de Fajcik et al. (2021) basan nuestra implementación en las bibliotecas PyTorch (Paszke et al., 2019), Transformers (Wolf et al., 2020) y usan AdamW (Loshchilov y Hutter, 2017) para la optimización. Ajustamos los hiperparámetros a través de Hy perOpt (Bergstra et al., 2015) e informamos los hiperparámetros exactos en el Apéndice A.

4.3 Métricas de evaluación

En esta sección, describimos las métricas que usamos para evaluar el sistema.

F1: la puntuación de F1 fue la principal evaluación oficial cumplida en el reto. Se calcula sobre las etiquetas B e I en secuencia siguiendo el esquema de etiquetado BIO para cada ejemplo y cada componente de triplete CES por separado, utilizando sequeval⁴. La F1 es entonces

³ Implementado en <https://shorturl.at/kxEVW>.
⁴ <https://github.com/chakki-works/sequeval>.

Sistema	CE Causa Efecto Señal General
Basal 2.2 T5-NoHistory	.181 67.7±2 T5-ECS :168 75.9±5 71.3±4
76.1±5 73.5±2 T5-CES	.183 81.0±4 67.8±2 66.7±5 73.0±2 T5-
CESLARGE	.159 73.5±8 74.1±4 77.2±7 74.8±2

Tabla 2: Resultados principales, en términos de Cross-Entropy (CE) y F1, con ± desviaciones estándar en los datos de desarrollo.

promediado primero a través de ejemplos de conjuntos de datos, obteniendo F1 para cada componente (Causa F1, Efecto F1, Señal F1). El F1 general se calcula como un promedio ponderado de los ejemplos de componentes por su frecuencia.

CE: es una entropía cruzada de token promedio, calculada como se describe en la Sección 4.2.

ES Acc: es una precisión de señal vacía, es decir, una precisión del modelo que predice que no hay amplitud de señal en el triplete CES cuando se le da una causa y efecto dorados.

4.4 Modelo de línea de base

Como modelo de referencia, utilizamos el modelo proporcionado por los organizadores de CASE-2022 para la Subtarea 2: un generador aleatorio que muestrea de manera uniforme una causa, un efecto y un intervalo de señal5 de la oración. Esta línea de base garantiza que la causa y el efecto no se superpongan.

5 Resultados y Discusión

Ahora reportamos los resultados obtenidos al promediar al menos diez desempeños medidos de 10 puntos de control entrenados con diferentes semillas6 en 4 variantes . Nosotros estudiamos diferentes de nuestro sistema. El sistema T5- CES es nuestro modelo vainilla descrito en 4.1, basado en la base T5. System T5-CESLARGE es el mismo modelo basado en T5-large. A diferencia de T5-CES, el sistema T5-ECS invierte el orden de generación al generar el primer efecto y causa, seguido de la señal (suponiendo un orden causal efecto→causa→señal, de ahí el sufijo ECS). Por último, estudiamos el efecto del condicionamiento del modelo sobre la historia de los tripletes ya generados. Eliminamos el historial de la entrada en todo momento durante el entrenamiento y predecimos los cuatro tripletes CES idénticos para cada ejemplo en el tiempo de prueba.

Nuestros resultados de la ablación están disponibles en la Tabla 2.

En primer lugar, el modelo sin historial en la entrada funciona significativamente peor, lo que valida nuestra hipótesis de que el modelo puede aprender a disminuir la probabilidad de los tripletes ya contenidos dentro del modelo.

5 Disponible en <https://shorturl.at/msY04>.
Las predicciones de conjuntos de 6 desarrolladores de nuestro mejor modelo de base t5 están disponibles en <https://shorturl.at/bjVZ9>.

Sistema	Desarrollo F1	Desarrollo1 F1 ES	Desarrollo2 Cuenta	Desarrollo Prueba F1
T5-ECS 77,7 80,9 71,1	1 T5-CESLARGE 78,3		82	43,4
77,4 80,0 T5-CES 77,5	79,6 73,3		70	43,7
			70	48,8

Tabla 3: Puntos de control principales enviados a la tabla de clasificación.

entrada, incluso a partir de solo 160 muestras. En segundo lugar, observamos una tendencia general de que en Causa F1 T5- CES supera a T5-ECS y en Efecto F1, T5-ECS supera a T5-CES. Esto lleva a la hipótesis de que cualquiera que sea la parte del triplete, causa o efecto, que se genere primero, el modelo de lenguaje funciona mejor en su caso. En tercer lugar, observamos que el modelo grande logró los mejores resultados en promedio. También logró nuestro mejor desempeño en un solo punto de control en el conjunto de desarrollo (78.3 en F1 general). Sin embargo, dado el tamaño de la muestra del conjunto de desarrolladores, las diferencias entre T5-CES, T5-ECS y T5-CESLARGE difícilmente pueden considerarse significativas.

A continuación, presentamos nuestros resultados en el conjunto de prueba en la Tabla 3. Enviamos puntos de control con la mejor puntuación general de F1 en el conjunto de desarrollo (Dev F1) a la tabla de clasificación mientras variamos los tipos de modelo. Observamos una caída significativa en el rendimiento de los datos de prueba. Como la anotación en los datos de prueba no es publicado en el momento de redactar este informe, se desconocen las causas de esta caída del rendimiento. Suponemos que el tamaño podría haber sido causado por un cambio de covariable en los datos de prueba, como lo respaldan las estadísticas de #Signals en la Tabla 1.

Además, incluimos estadísticas adicionales (Dev0 F1, Dev1 F1, Dev ES Acc) para nuestros mejores puntos de control. Esperábamos que el rendimiento en el subconjunto de desarrollo con dos tripletes (Dev2 F1) por ejemplo fuera peor que en el subconjunto de desarrollo con un triplete por oración (Dev1 F1). En cuanto al rendimiento, este no siempre parece ser el caso. Tras el análisis manual , encontramos que el modelo a menudo fallaba en la segunda ronda de extracción de tripletes. Encontramos 2 alucinaciones LM de 18 muestras de desarrollo en la ronda de segunda generación.

6 Velocidad de inferencia

Al medir la velocidad de inferencia en el conjunto de prueba, utilizamos una estación de trabajo GPU 2080Ti basada en Intel i5. La inferencia de 4 tripletes CES sin procesamiento posterior por ejemplo de 1 oración tomó 1,46 segundos en promedio. El tiempo de ejecución del posprocesamiento fue insignificante, con un promedio de 0,025 segundos por ejemplo de oración.

7. Conclusión

En este trabajo, hemos analizado nuestras presentaciones del 2.º lugar de CASE-2022 en la Subtarea 2. Mostramos que un modelo generativo podría extraer tripletes de señales de causa-efecto a nivel competitivo utilizando solo 160 muestras anotadas. Investigamos supuestos causales sobre el orden de generación de causa y efecto para responder a la pregunta de investigación "¿debería identificarse primero la causa y solo luego el efecto, o viceversa?" y descubrió que si bien la F1 general no cambiará significativamente, el componente que se generó primero logró un mejor rendimiento en promedio (Causa primero logró una mejor Causa-F1 y Efecto primero Efecto-F1 respectivamente). Finalmente, mostramos que la diferencia F1 entre el subconjunto dev con 1 o 2 tripletes causales por oraciones no es significativa.

Agradecimientos

Este trabajo fue apoyado por CRITERIA, un proyecto de la UE, financiado bajo el programa Horizon 2020, acuerdo de subvención no. 101021866 y el Ministerio de Educación, Juventud y Deportes de la República Checa a través de e-INFRA CZ (ID:90140). Esaú Villatoro-Tello, fue apoyado parcialmente por Ildiap, SNI CONACyT y UAM Cuajimalpa México.

Referencias

Biswanath Barik, Erwin Marsi y Pinar Øzturk. 2016.

Extracción de causalidad de eventos de la literatura de ciencias naturales

Brandon Beamer, Alla Rozovskaya y Roxana Girju. 2008. Extracción automática de relaciones semánticas con generación de límites múltiples. En AAAI, páginas 824–829.

Guillermo Becquin. 2020. Gbe en fincausal 2020, tarea 2: extracción de causalidad basada en intervalos para documentos financieros. En Actas del 1er Taller Conjunto sobre Procesamiento de Narrativas Financieras y Resumen Financiero MultiLing, páginas 40–44.

Giannis Bekoulis, Johannes Deleu, Thomas Demeester y Chris Develder. 2018. Entrenamiento adversario para la extracción de relaciones y entidades conjuntas multicontexto. preimpresión de arXiv arXiv:1808.06876.

James Bergstra, Brent Komer, Chris Eliasmith, Dan Yamins y David D Cox. 2015. Hyperopt: una biblioteca de Python para la selección de modelos y la optimización de hiperparámetros. Ciencia computacional y descubrimiento, 8(1):014008.

Quoc-Chinh Bui, Breannán Ó Nualláin, Charles A Boucher y Peter Sloot. 2010. Extracción causal

relaciones sobre la resistencia a los medicamentos del VIH de la literatura. Bioinformática BMC, 11(1):1–11.

Sergio Burdisso, Juan Zuluaga-Gomez, Martin Fajcik, Esaú Villatoro-Tello, Muskaan Singh, Petr Motlicek, and Pavel Smrz. 2022. IDIAPers @ corpus de noticias causales 2022: Identificación eficiente de relaciones causales a través de un enfoque de pocas tomas basado en indicaciones. En el 5º Taller sobre Desafíos y Aplicaciones de la Extracción Automatizada de Eventos Sociopolíticos del Texto (CASE @ EMNLP 2022). Asociación de Lingüística Computacional.

Tommaso Caselli y Piek Vossen. 2017. El corpus de la historia del evento: un nuevo punto de referencia para la extracción de relaciones causales y temporales. En Proceedings of the Events and Stories in the News Workshop, páginas 77–86.

Jifan Chen, Qi Zhang, Pengfei Liu, Xinpeng Qiu y Xuan-Jing Huang. 2016. Detección de relación de discurso implícito a través de una arquitectura profunda con red de relevancia cerrada. En Actas de la 54.ª Reunión Anual de la Asociación de Lingüística Computacional (Volumen 1: Documentos extensos), páginas 1726–1735.

Tirthankar Dasgupta, Rupsa Saha, Lipika Dey y Abir Naskar. 2018. Extracción automática de relaciones causales del texto utilizando redes neuronales profundas informadas lingüísticamente. En Actas de la 19.ª Reunión Anual SIGDial sobre Discurso y Diálogo, páginas 306–316.

Jacob Devlin, Ming-Wei Chang, Kenton Lee y Kristina Toutanova. 2018. Bert: Preentrenamiento de transformadores bidireccionales profundos para la comprensión del lenguaje. preimpresión de arXiv arXiv:1810.04805.

Quang Do, Yee Seng Chan y Dan Roth. 2011. Identificación de causalidad de eventos mínimamente supervisados. En Actas de la Conferencia de 2011 sobre métodos empíricos en el procesamiento del lenguaje natural, páginas 294–303.

Jesse Dunietz, Gregory Burnham, Akash Bharadwaj, Owen Rambow, Jennifer Chu-Carroll y David Ferrucci. 2020. Para probar la comprensión de la máquina, comience definiendo la comprensión. preimpresión de arXiv arXiv:2005.01525.

Martin Fajcik, Martin Docekal, Karel Ondrej y Pavel Smrz. 2021. [R2-D2: una línea de base modular para responder preguntas de dominio abierto](#). En Hallazgos de la Asociación de Lingüística Computacional: EMNLP 2021, páginas 854–870, Punta Cana, República Dominicana. Asociación de Lingüística Computacional.

Martin Fajcik, Josef Jon, Martin Docekal y Pavel Smrz. 2020. [BUT-FIT en SemEval-2020 tarea 5: Detección automática de declaraciones contrafácticas con profundos modelos de representación del lenguaje pre-entrenados](#). En Actas del Decimocuarto Taller sobre Evaluación Semántica, páginas 437–444, Barcelona (en línea). Comité Internacional de Lingüística Computacional.

- Amir Feder, Katherine A Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E Roberts, et al. 2021a. Inferencia causal en el procesamiento del lenguaje natural : Estimación, predicción, interpretación y más allá. preimpresión de arXiv arXiv:2109.00725.
- Amir Feder, Nadav Oved, Uri Shalit y Roi Reichart. 2021b. Causalidad: Explicación del modelo causal a través de modelos de lenguaje contrafáctico. *Lingüística computacional* , 47(2):333–386.
- Daniela García et al. 1997. Coatis, un sistema de PNL para localizar expresiones de acciones conectadas por vínculos de causalidad. En *Conferencia Internacional sobre Ingeniería del Conocimiento y Gestión del Conocimiento*, páginas 347–352 . Springer.
- Roxana Girju. 2003. Detección automática de relaciones causales para la respuesta a preguntas. En *Actas del taller ACL 2003 sobre resumen multilingüe y respuesta a preguntas*, páginas 76–83.
- Roxana Girju, Preslav Nakov, Vivi Nastase, Stan Szpakowicz, Peter Turney y Deniz Yuret. 2009. Clasificación de relaciones semánticas entre nominales. *Recursos lingüísticos y evaluación*, 43(2):105–121.
- Ashwin Ittoo y Gosse Bouma. 2013. Aprendizaje mínimamente supervisado de relaciones causales específicas del dominio utilizando un corpus de dominio abierto como base de conocimiento. *Ingeniería de datos y conocimiento*, 88: 142–163.
- Xianxian Jin, Xinzhi Wang, Xiangfeng Luo, Subin Huang y Shengwei Gu. 2020. Extracción de causalidad implícita y entre oraciones del corpus chino. En *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, páginas 739–751. Saltador.
- Yohan Jo, Seojin Bang, Chris Reed y Eduard Hovy. 2021. Clasificar relaciones argumentativas utilizando mecanismos lógicos y esquemas de argumentación. *Transactions of the Association for Computational Linguistics*, 9:721–739.
- Ning Kang, Bharat Singh, Chinh Bui, Zubair Afzal, Erik M van Mulligen y Jan A Kors. 2014. Extracción basada en el conocimiento de eventos adversos de medicamentos del texto biomédico. *Bioinformática BMC*, 15(1):1–8.
- Christopher SG Khoo, Syin Chan y Yun Niu. 2000. Extraer conocimiento causal de una base de datos médica utilizando patrones gráficos. En *Actas de la 38.ª reunión anual de la asociación de lingüística computacional* , páginas 336–343.
- Christopher SG Khoo, Jaklin Kornfilt, Robert N Oddy y Sung Hyon Myaeng. 1998. Extracción automática de información de causa-efecto del texto del periódico sin inferencia basada en el conocimiento. *Informática literaria y lingüística*, 13(4):177–186.
- Canasai Kruengkrai, Kentaro Torisawa, Chikara Hashimoto, Julien Kloetzer, Jong-Hoon Oh y Masahiro Tanaka. 2017. Mejora del reconocimiento de causalidad de eventos con múltiples fuentes de conocimiento de fondo utilizando redes neuronales convolucionales de varias columnas . En *Actas de la Conferencia AAAI sobre Inteligencia Artificial*, volumen 31.
- Fei Li, Meishan Zhang, Guohong Fu y Donghong Ji. 2017. Un modelo de articulación neural para la extracción de entidades y relaciones de textos biomédicos. *Bioinformática BMC*, 18(1):1–11.
- Ilya Loshchilov y Frank Hutter. 2017. Regularización de la caída de peso desacoplada. preimpresión de arXiv arXiv:1711.05101.
- Eugenio Martínez-Cámara, Vered Shwartz, Iryna Gurevych e Ido Dagan. 2017. Desambiguación neuronal de marcadores léxicos causales según el contexto. En *IWCS 2017—12.ª Conferencia Internacional sobre Semántica Computacional—Artículos breves*.
- Jorge A. Miller. 1994. [WordNet: Una base de datos léxica para inglés](#). En *Tecnología del lenguaje humano: actas de un taller realizado en Plainsboro, Nueva Jersey, del 8 al 11 de marzo de 1994*.
- Paramita Mirza, Rachele Sprugnoli, Sara Tonelli, and Manuela Speranza. 2014. Anotando la causalidad en el corpus tempeval-3. En *Actas del taller de EACL 2014 sobre enfoques computacionales de la causalidad en el lenguaje (CAToCL)*, páginas 10–19.
- Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen y Lucy Vanderwende. 2016. Caters: esquema de relación causal y temporal para la anotación semántica de estructuras de eventos. En *Actas del Cuarto Taller sobre Eventos*, páginas 51–61.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai y Soumith Chintala. 2019. [Pytorch: Un estilo imperativo, aprendizaje profundo de alto rendimiento biblioteca](#). En H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox y R. Garnett, editores, *Advances in Neural Information Processing Systems* 32, páginas 8024–8035. Curran Associates, Inc.
- Edoardo Ponti y Anna-Leena Korhonen. 2017. Las características relacionadas con eventos en las redes neuronales feedforward contribuyen a identificar relaciones causales implícitas en discurso.
- Kira Radinsky, Sagie Davidovich y Shaul Markovitch. 2012. Aprendizaje de la causalidad para la predicción de eventos noticiosos. En *Actas de la 21.ª conferencia internacional sobre la World Wide Web*, páginas 909–918.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Explorando los límites

- de transferencia de aprendizaje con un transformador unificado de texto a texto. *J. Mach. Aprender. Res.*, 21(140):1–67.
- Dhanya Sridhar y Lise Getoor. 2019. Estimación de los efectos causales del tono en los debates en línea. preimpresión de arXiv arXiv:1906.04177.
- Cong Sun, Zhihao Yang, Ling Luo, Lei Wang, Yin Zhang, Hongfei Lin y Jian Wang. 2019. Un enfoque de aprendizaje profundo con representaciones de palabras contextualizadas profundas para la extracción de interacción químico-proteína de la literatura biomédica. *Acceso IEEE*, 7:151034–151046.
- Chenhao Tan, Lillian Lee y Bo Pang. 2014. El efecto de la redacción en la propagación de mensajes: experimentos naturales controlados por temas y autores en Twitter. preimpresión de arXiv arXiv:1405.1438.
- Fiona Anting Tan, Ali Hürriyetoglu, Tommaso Caselli, Nelleke Oostdijk, Hansi Hettiarachchi, Tadashi Nomoto, Onur Uca y Farhana Ferdousi Liza. 2022a. Identificación de causalidad de eventos con corpus de noticias causales: tarea compartida 3, CASO 2022. En *Actas del 5.º taller sobre desafíos y aplicaciones de la extracción automatizada de eventos sociopolíticos del texto (CASO 2022)*, en línea. Asociación de Lingüística Computacional.
- Fiona Anting Tan, Ali Hürriyetoglu, Tommaso Caselli, Nelleke Oostdijk, Tadashi Nomoto, Hansi Hettiarachchi, Iqra Ameer, Onur Uca, Farhana Ferdousi Liza y Tiancheng Hu. 2022b. *El corpus de noticias causales: anotación de relaciones causales en oraciones de eventos de noticias*. En *Proceedings of the Language Resources and Evaluation Conference*, páginas 2298–2310, Marsella, Francia. Asociación Europea de Recursos Lingüísticos.
- Fiona Anting Tan, Xinyu Zuo y See-Kiong Ng. 2022c. *Unicausal: Benchmark y modelo unificados para la minería de texto causal*.
- Victor Veitch, Dhanya Sridhar y David Blei. 2020. Adaptación de incrustaciones de texto para inferencia causal. En *Conferencia sobre Incertidumbre en Inteligencia Artificial*, páginas 919–928. PMLR.
- Oriol Vinyals, Meire Fortunato y Navdeep Jaitly. 2015. Redes de punteros. *Avances en los sistemas de procesamiento de información neuronal*, 28.
- Jue Wang y Wei Lu. 2020. Dos son mejor que uno: entidad conjunta y extracción de relaciones con codificadores de secuencia de tabla. preimpresión de arXiv arXiv:2010.03851.
- Linlin Wang, Zhu Cao, Gerard De Melo y Zhiyuan Liu. 2016. Clasificación de relaciones vía cnns de atención multinivel. En *Actas de la 54.ª Reunión Anual de la Asociación de Lingüística Computacional (Volumen 1: Documentos extensos)*, páginas 1298–1307.
- Bonnie Webber, Rashmi Prasad, Alan Lee y Aravind Joshi. 2019. *Manual de anotaciones de Penn Speech Treebank 3.0*. Filadelfia, Universidad de Pensilvania, 35:108.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest y Alexander Rush. 2020. *Transformadores: procesamiento de lenguaje natural de última generación*.
- En *Actas de la Conferencia de 2020 sobre métodos empíricos en el procesamiento del lenguaje natural: demostraciones del sistema*, páginas 38 a 45, en línea. Asociación de Lingüística Computacional.
- Zach Wood-Doughty, Ilya Shpitser y Mark Dredze. 2018. Desafíos del uso de clasificadores de texto para la inferencia causal. En *Actas de la Conferencia sobre métodos empíricos en el procesamiento del lenguaje natural*. Conferencia sobre métodos empíricos en el procesamiento del lenguaje natural, volumen 2018, página 4586. NIH Public Access.
- Hang Yan, Tao Gui, Junqi Dai, Qipeng Guo, Zheng Zhang y Xipeng Qiu. 2021. Un marco generativo unificado para varias sub tareas. En *Actas de la 59.ª Reunión Anual de la Asociación de Lingüística Computacional y la 11.ª Conferencia Internacional Conjunta sobre Procesamiento del Lenguaje Natural (Volumen 1: Documentos extensos)*, páginas 5808–5822.
- Liang Yao, Chengsheng Mao y Yuan Luo. 2019. Redes convolucionales gráficas para clasificación de texto. En *Actas de la conferencia AAAI sobre inteligencia artificial*, volumen 33, páginas 7370–7377.
- Justine Zhang, Sendhil Mullainathan y Cristian Danescu-Niculescu-Mizil. 2020. Cuantificación de los efectos causales de las tendencias conversacionales. *Procedimientos de la ACM sobre interacción humano-computadora*, 4 (CSCW2): 1–24.
- Yuhao Zhang, Peng Qi y Christopher D. Manning. 2018. La convolución de gráficos sobre árboles de dependencia podados mejora la extracción de relaciones. preimpresión de arXiv arXiv:1809.10185.
- Shan Zhao, Minghao Hu, Zhiping Cai y Fang Liu. 2021. Modelado de interacciones intermodales densas para la extracción conjunta de entidad-relación. En *Actas de la Vigésima Novena Conferencia Internacional sobre Conferencias Conjuntas Internacionales sobre Inteligencia Artificial*, páginas 4032–4038.

Hiperparámetro Tasa de aprendizaje	
del valor 0,0002 abandono oculto	
atención abandono	.1436
peso decaimiento	.4719
tamaño del	.0214
minilote proporción	8
de calentamiento 0,1570 constante	
del programador (sin disminución de lr) 10 000 pasos	
gradiente	máximos
máximo norma 1	

Tabla 4: Ajuste de hiperparámetro utilizado en este trabajo.

A Hiperparámetros

En la Tabla 4, informamos los hiperparámetros exactos utilizados al ajustar T5. La proporción de calentamiento, la disminución del peso y los abandonos están en el rango (0,1) (por ejemplo, 0,4719 significa 47,19 %).

B Ejemplo de Entradas

El formato de entrada y el formato de etiqueta para un solo ejemplo de entrenamiento, una oración con 2 tripletes CES, se ilustran en la Figura. 2.

ENTRADA DEL CODIFICADOR: " Creo que los productores de cine independientes tienen la responsabilidad de documentar lo que los principales medios de comunicación no pudo informar sobre . Pero en la víspera de las protestas segundo aniversario , Chan reclama todo Hong Kong _ s grandes cines se niegan a mostrar su película , el resultado , sospecha que es , , de autocensura progresiva como las empresas evitan ofender a Beijing . historia: **PREFIJO DEL DECODIFICADOR:** causa: **OBJETIVO DEL DECODIFICADOR :** causa: las empresas evitan ofender a Beijing

ENTRADA DEL CODIFICADOR: " Creo que los productores de cine independientes tienen la responsabilidad de documentar lo que los principales medios de comunicación no informaron . Pero en la Víspera de las protestas Hong Kong , Chan reclama todo el segundo aniversario de Los principales cines se niegan a proyectar su película y , como resultado , sospecha que las empresas , , de autocensura progresiva como evitan ofender a Beijing . causa: los negocios se rehúyen de ofender a la historia de Beijing : **PREFIJO DEL DECODIFICADOR:** efecto: **OBJETIVO DEL DECODIFICADOR:** efecto: autocensura progresiva

ENTRADA DEL CODIFICADOR: " Creo que los productores de cine independientes tienen la responsabilidad de documentar lo que los principales medios de comunicación no informaron . Pero en la víspera de la protesta, los principales segundo aniversario , Chan afirma que todo Hong Kong tiene una autocensura cines se niegan a proyectar su película , el resultado , sospecha . , , progresiva como las empresas evitan ofender a Beijing . causa: el negocio es tímido lejos de ofender a Beijing efecto: progresiva autocensura historia del barco : **PREFIJO DEL DECODIFICADOR:** señal: **OBJETIVO DEL DECODIFICADOR:** señal: como

ENTRADA DEL CODIFICADOR: " Creo que los productores de cine independientes tienen la responsabilidad de documentar lo que los principales medios de comunicación no informaron . Pero en la víspera del segundo aniversario de la protesta , Chan afirma que todos los principales cines de Hong Kong se niegan a mostrar su película , el resultado , sospecha _ , , de autocensura progresiva como las empresas evitan ofender a Beijing . historia : causa : las empresas rehúyen ofender Efecto de Beijing : señal de autocensura progresiva : como **PREFIJO DEL DECODIFICADOR:** causa: **OBJETIVO DEL DECODIFICADOR:** causa: autocensura progresiva

ENTRADA DEL CODIFICADOR: " Creo que los productores de cine independientes tienen la responsabilidad de documentar lo que los principales medios de comunicación no informaron . Pero en la víspera de las protestas segundo aniversario , Chan reclama todo Hong Kong _ s grandes cines se niegan a mostrar su película , el resultado , sospecha que es , , de autocensura progresiva como las empresas evitan ofender a Beijing . causa : autocensura progresiva historia : causa : las empresas evitan ofender a Beijing efecto : autocensura progresiva señal : como **PREFIJO DEL DECODIFICADOR:** efecto: **OBJETIVO DEL DECODIFICADOR:** efecto: todo Hong Kong Los principales cines se niegan a proyectar su película

ENTRADA DEL CODIFICADOR: " Creo que los productores de cine independientes tienen la responsabilidad de documentar lo que los principales medios de comunicación no informaron . Pero en la víspera de las protestas segundo aniversario , Chan reclama todo Hong Kong _ s grandes cines se niegan a mostrar su película , el resultado , sospecha que es , , de autocensura progresiva a medida que las empresas evitan ofender a Beijing . causa : autocensura progresiva efecto : todos los principales cines de Hong Kong se niegan a mostrar su s historia cinematográfica : causa : negocios tímidos Lejos de ofender el efecto de Beijing : señal de autocensura progresiva: como **PREFIJO DEL DECODIFICADOR:** señal: **OBJETIVO DEL DECODIFICADOR:** señal: el resultado , él sospecha , , de

Figura 2: Ejemplo de entradas tokenizadas para una oración con dos tripletes CES anotados. Frases "ENCODER ENTRADA". El "PREFIJO DEL DECODIFICADOR" y el "OBJETIVO DEL DECODIFICADOR" no forman parte de la entrada y se incluyen solo con fines ilustrativos. Las secuencias especiales (_causa:, _efecto:, _señal:, _historia:) utilizadas entre partes concatenadas de la entrada están en negrita.