

NOMBRE: Benjamín Farías Valdés

N.ALUMNO: 22102671



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

IIC3692 — Tópicos Avanzados en Inteligencia Artificial — 2' 2022

## Lectura 20

### Crítica

#### COMET: Commonsense Transformers for Automatic Knowledge Graph Construction

El paper propone un modelo generativo de conocimiento basado en sentido común, denominado *COMET*. Este modelo aprende a generar descripciones adecuadas y diversas basadas en sentido común (según lo percibimos los humanos), permitiendo construir/extender grafos denominados *KBs* (*Knowledge Bases*).

El punto fuerte de esta propuesta es su capacidad para utilizar modelos pre-entrenados y adaptarlos a esta nueva tarea de generación de conocimiento basado en sentido común. En pocas palabras, *COMET* es un *framework* que permite aplicar *fine-tuning* sobre modelos pre-entrenados en modelamiento del lenguaje natural (como *GPT*). Me parece que esto es un enfoque bastante innovador, ya que es el primero que intenta usar modelos de lenguaje natural en vez de buscar reglas lógicas sobre los datos. La razón de porqué esto funciona, en mi opinión, viene del hecho de que el sentido común está presente en una gran parte de nuestro lenguaje, por lo que el conocimiento aprendido por modelos como *GPT-3* de cierta forma contiene reglas lógicas implícitas que pueden ser aprovechadas para generar ejemplos nuevos que sigan el formato de las *KBs*.

Los experimentos realizados muestran una gran capacidad para generar tuplas de sentido común, dado que incluso los humanos encontraron que alrededor del 70% de los candidatos eran plausibles. Algo que encontré interesante es que los casos en los que falla parece ser debido a que alguna parte del sujeto desvía el significado de la tupla, por ejemplo en una tupla se habla sobre una persona apagando una fogata, y el modelo responde que el efecto de esto es que la persona se quema. Lo más probable es que el hecho de que haya una fogata de cierta forma eclipsó el peso de la acción de apagarla, por lo que el modelo se desvía y asocia el fuego con quemarse. Otro resultado importante de los experimentos, es que existe una considerable diferencia de rendimiento cuando se usa el modelo pre-entrenado desde el inicio, respecto a usando el mismo modelo pero sin pre-entrenar. Esta diferencia era esperable, dado que la premisa de utilizar este enfoque era justamente que el conocimiento de los modelos del lenguaje pre-entrenados sería clave en la generación de ejemplos de valor con sentido común.

Algo que me gustaría ver es la aplicación de *COMET* con otros modelos pre-entrenados aparte de *GPT*, dado que en el artículo sólo se probó con ese. Siento que sólo probando con múltiples modelos distintos se podrá realmente decir que esta propuesta es agnóstica a los modelos. Otro tipo de aplicación que sería interesante es utilizar estas mismas ideas pero en el dominio de las imágenes, generando, por ejemplo, secuencias de imágenes que cuentan una pequeña historia que tenga sentido común. En esos casos se podrían

utilizar modelo pre-entrenados que hayan sido entrenados con videos o secuencias de imágenes relacionadas temporalmente.

Una limitación posible en términos de verificar la calidad del conocimiento generado por *COMET* es que tener personas realizando esta tarea no es muy práctico. En el paper se presentan otras métricas para determinar la calidad de los resultados, pero se hace énfasis en la evaluación de los humanos e incluso se indica que esta es una forma muy viable de evaluar el modelo para trabajo futuro, lo que en mi opinión está bien pero puede resultar bastante ineficiente si se trata de demasiados datos generados. Quizás hay buen lugar a investigar sobre métricas automatizadas que sean tan buenas como la evaluación humana para estas tareas, dado que también se mencionan ciertos métodos que están cerca de lograrlo.

En general encontré bastante buena la propuesta, y aprendí bastante sobre una tarea que no había visto en otros trabajos del área. Creo que este es sólo el primer paso para este tipo de enfoques en generación de contenido basado en sentido común, y existen varias vías de trabajo futuro que parecen atractivas, por ejemplo aplicar esto en otros dominios de los datos (ej. imágenes), probar con múltiples modelos, e investigar mejores métricas para verificar la calidad de forma automática.