

Resumen automático de textos basado en un modelo unificado del resumen extractivo y abstractivo.

Juan José López Condori

Arequipa, Perú

2018



Resumen automático de textos basado en un modelo unificado del resumen extractivo y abstractivo.

Por
Juan José López Condori

Tesis presentada a la
Escuela Profesional de Ciencia de la Computación
de la
UNIVERSIDAD NACIONAL DE SAN AGUSTÍN
como requisito
para obtener el título profesional
de
Licenciado en Ciencia de la Computación

UNIVERSIDAD NACIONAL DE SAN AGUSTÍN
FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS
ESCUELA PROFESIONAL DE CIENCIA DE LA COMPUTACIÓN

Resumen automático de textos basado en un modelo unificado del
resumen extractivo y abstractivo.

Tesis de graduación presentado por el bachiller Juan José López Condori en el
cumplimiento de los requisitos para obtener el título profesional de Licenciado
en Ciencia de la Computación.

Arequipa, 19 de Junio del 2018

Aprobado por:

*A Dios, por todo lo que me ha dado, a
todos los profesores por sus enseñanzas
y algunos amigos.*

Índice general

Agradecimientos	VIII
Resumen	IX
1. Introducción	1
1.1. Contexto y Motivación	1
1.2. Definición del problema	2
1.3. Justificación	2
1.4. Objetivos	3
1.4.1. Objetivo General	3
1.4.2. Objetivos específicos	3
1.5. Limitaciones	3
1.6. Organización de la tesis	3
2. Marco teórico y Antecedentes	4
3. Formalismos y/o teoría propuesta	7
3.1. Combinando atenciones	7
3.2. Extractivo	8
3.2.1. PageRank	8
3.2.2. TextRank	9
3.3. Abstractivo	9
3.3.1. Modelo secuencia a secuencia con atención	10
3.3.2. Red <i>Pointer-generator</i>	11
4. Experimentación o evaluación empírica	13
4.1. Base de datos	13
4.2. Entrenamiento	13
4.3. Métricas para la evaluación	14
5. Resultados y/o Evaluaciones	15
5.1. Resultados de las oraciones extraídas	15
5.2. Resultados del resumen abstractivo	15
5.3. Resultados Modelo unificado	17
5.4. Conclusiones	18
Bibliografía	19

Índice de figuras

3.1. Estructura de la red RNN	10
3.2. Estructura de una red <i>Pointer-generator</i>	12

Índice de cuadros

5.1. Evaluación del extractor con otros modelos mediante las puntuaciones ROUGE	15
5.2. Evaluación del abstractor con otros modelos mediante las puntuaciones ROUGE	16
5.3. Comparación del resumen del modelo unificado con y el resumen realizado por una persona	17
5.4. Puntuaciones ROUGE del modelo unificado	18

Agradecimientos

Resumen

La generación automática de resúmenes de texto suele formar parte de las aplicaciones de minería de textos, la tarea es condensar automáticamente un trozo de texto a una versión más corta manteniendo los puntos importantes. Hay principalmente dos tipos de enfoques: extractivo y abstractivo.

La mayoría de estudios se enfocan en solo uno de ellos, en esta tesis se presenta un modelo que combina ambos enfoques, los resultados obtenidos muestran que el método propuesto es competitivo con los métodos de la literatura y, en varios casos, los superan.

Palabras claves— resumen automático, resumen extractivo, resumen abstractivo

Capítulo 1

Introducción

1.1. Contexto y Motivación

El desarrollo de las Tecnologías de la Información y, especialmente, Internet ha desestabilizado por completo lo que conocíamos como producción de documentos y, por ende, todo el proceso documental.

El problema ahora no es sólo la cantidad de documentos, sino otra serie de factores añadidos como la volatilidad, la falta de tipología, las autorías de nuevo tipo (colaborativas, y otras), nuevas estructuras documentales, y un largo etcétera.

Una de las cuestiones que tiene planteadas la Minería de Texto, es el resumen automático de documentos, un tema de suma importancia que recibe continuamente contribuciones científicas.

El resumen de textos es la tarea de condensar automáticamente un trozo de texto a una versión más corta manteniendo los puntos importantes. La capacidad de condensar información de texto puede ayudar a muchas aplicaciones tales como la creación de compendios de noticias, presentando resultados de búsqueda, y generación de informes.

Hay principalmente dos tipos de enfoques: extractivo y abstractivo. Los enfoques extractivos reúnen resúmenes directamente de el texto de origen típicamente selecciona una oración completa a la vez. En contraste, los enfoques abstractivos pueden generar nuevas palabras y frases que no son copiados del texto de origen.

La mayoría de estudios [20], [30], [26], [3], [16], [10] y [17] se enfocan solo en uno de ellos, es por esto que la presente tesis describe un método unificado del resumen extractivo y abstractivo para la generación automática de resúmenes de un texto determinado.

1.2. Definición del problema

El propósito de los resúmenes es facilitar y acelerar la identificación de los temas interesantes de entre una gran cantidad de documentos. El objetivo final es salvar un tiempo de lectura necesario para localizar la información requerida en un determinado momento. El problema es que la elaboración de resúmenes consume abundantes recursos humanos, la aplicación de ordenadores a esta tarea viene siendo estudiada desde hace varias décadas; ya en 1958 se empiezan a proponer sistemas de resumen automático [13] como solución al creciente número de textos técnicos publicados.

Pero es en la actualidad, sobre todo con el crecimiento espectacular de Internet, cuando se ha hecho más perentoria la necesidad de disponer de esta tecnología, lo que ha potenciado su desarrollo.

Una solución a este problema se plantea en la presente tesis que es el unificar dos enfoques de resumen: el extractivo y abstractivo para la generación de un resumen más informativo y legible.

1.3. Justificación

Esta tesis pretende explotar tanto las características del resumen extractivo como del resumen abstractivo, adaptando un modelo que combina atenciones en dos niveles: oración (extractivo) y palabra (abstractivo) para los datos CNN / Daily Mail especialmente en textos de noticias, además de aplicarlo para contribuir en la creación de compendios de noticias, presentando resultados de búsqueda, y generación de informes.

1.4. Objetivos

1.4.1. Objetivo General

Aplicar un modelo para unificar el resumen extractivo y abstractivo y así aprovechar sus características propias de cada uno, logrando un resumen más informativo y legible.

1.4.2. Objetivos específicos

- Comparar las distintas técnicas del resumen extractivo y abstractivo respectivamente.
- Evaluar y comparar el método que aprovechará las características del resumen extractivo y abstractivo.

1.5. Limitaciones

La tesis se limita enfocandose en textos de noticias y no discute otros tipos de textos tales como opiniones, tweets, etcétera.

1.6. Organización de la tesis

El contenido de la presente tesis está estructurada en 5 capítulos, los cuales se explican a continuación. En el capítulo 2, se explican los conceptos básicos que se utilizan en la tesis, los cuales incluyen nociones de resumen automático, el resumen extractivo y abstractivo, además se hace una descripción de los trabajos relacionados con el tema de investigación de esta tesis.

En el capítulo 3, se describe el metodo utilizado para el resumen extractivo, el método para el resumen abstractivo y el metodo que unifica ambos enfoques.

En el capítulo 4 se presentan los experimentos realizados. La base de datos utilizada también se describe en este capítulo

Por último el capítulo 5 se muestran los resultados obtenidos con distintas evaluaciones.

Capítulo 2

Marco teórico y Antecedentes

El resumen automático de textos es la tarea que consiste en la producción automática de una versión más corta (conocida como sumario o resumen) de uno o más textos-fuente [14]. El resumen debe contener la información más relevante de los textos fuente. Muchos enfoques se han desarrollado para determinar de manera automática la información que debe incluirse en el resumen. [15] y [21], los resúmenes pueden clasificarse de varias formas. En cuanto a la información que contienen, los resúmenes pueden clasificarse en tres tipos: indicativos, informativos y críticos / evaluadores. Los resúmenes indicativos contienen sólo los temas esenciales de los textos fuente, no necesariamente conteniendo detalles de resultados, argumentos y conclusiones. Por ejemplo, los índices son sumarios indicativos. Los resúmenes informativos, que son los más tradicionales, se consideran sustitutos de los textos, debiendo contener toda la información principal. Los abstractos de artículos son ejemplos de este tipo de resumen. Los sumarios críticos, además de resumir el contenido de los textos fuente, añaden crítica en función del contenido. Las reseñas de libros son ejemplos de los resúmenes críticos. La sumarización automática también puede clasificarse como monodocumento o multidocumento. En la primera clase, el resumen se genera automáticamente a partir de un único texto-fuente. En la segunda categoría, se produce un resumen a partir de un conjunto de textos sobre un mismo tema. En cuanto a la forma de composición de los resúmenes, se tiene el resumen extractivo (cuando se seleccionan segmentos textuales enteros) y el abstractivo (cuando se realizan operaciones de reescritura). A continuación, estos enfoques se

explican, así como se mencionan brevemente algunos trabajos desarrollados para estos enfoques.

Resumen Extractivo: La sumarización extractiva genera un resumen seleccionando los segmentos más representativos (usualmente sentencias) de los textos fuente, sin hacer ningún cambio en los segmentos. La idea de este enfoque es extraer las oraciones que contengan mucha información y novedad. En la selección de las oraciones más relevantes, los métodos extractivos utilizan un mecanismo de los rangos para obtener las sentencias con las mejores puntuaciones, es decir, las más importantes.

Se usan redes neuronales para mapear oraciones en vectores en [20] , [30] y seleccionar oraciones basadas en esos vectores. [7] , [26] y [28], además de aplicar redes neuronales recurrentes para leer el artículo. y obtener las representaciones de las oraciones y el artículo para seleccionar oraciones.

Se puede utilizar información lateral (es decir, leyendas de imágenes y títulos) para ayudar al clasificador de oraciones a elegir oraciones.[9].

Un estudio reciente [18] combina redes neuronales recurrentes con redes convolucionales gráficas para computar la importancia de cada oración. Mientras que algunos métodos de resumen extractivo obtienen altas puntuaciones de ROUGE, todos ellos sufren de baja legibilidad.

Resumen Abstractivo: A diferencia del enfoque extractivo, el resumen abstractivo no sólo selecciona las sentencias de los textos fuente, analiza los documentos y automáticamente genera nuevas sentencias. Este enfoque intenta producir nuevos textos a partir de los fragmentos originales identificables como importantes. Con esta característica, este enfoque puede solucionar el problema de falta de cohesión de los enfoques extractivos. A pesar de no ser una cuestión nueva, hay relativamente pocos trabajos sobre la sumarización abstracta.

En [3] utiliza un codificador basado en la atención para leer el texto de entrada y generar el resumen. Basados en ellos, [17] emplea un codificador automático variacional y [27] utiliza un modelo de secuencia a secuencia más potente. Además, [27] crea un nuevo conjunto de datos de resumen de nivel de artículo llamado *CNN / Daily Mail* adaptando el conjunto de datos de preguntas y respuestas de DeepMind [10].

Se cambia el método de entrenamiento tradicional para optimizar directamente las

métricas de evaluación (por ejemplo, BLEU y ROUGE) en [16] .[8], [1] y [29] combinan redes de punteros [22] en sus modelos para tratar con palabras fuera de vocabulario (OOV). Para disminuir frases repetidas en el resumen generado. [24] y [2] restringen sus modelos de prestar atención a la misma palabra.

La investigación [4] aplica el modelo de secuencia a secuencia convolucional y diseña varias tareas nuevas para el resumen. [12] logra una alta puntuación de legibilidad en la evaluación humana utilizando redes adversas generativas.

Capítulo 3

Formalismos y/o teoría propuesta

El modelo combina un método extractivo basado en el algoritmo de TextRank [19] y un modelo abstractivo [1] que usa *Pointer-generator* combinando atención a nivel de oración de la primera y la atención a nivel de palabra de la segunda.

La entrada tanto del extractor como del abstractor es una secuencia de palabras $w = [w_1, w_2, \dots, w_m, \dots]$, donde m es el índice de palabras. La secuencia de palabras también forma una secuencia de oraciones $s = [s_1, s_2, \dots, s_n, \dots]$, donde n es el índice de oraciones. La palabra m^{th} se mapea en la $n(m)^{th}$ oración, donde $n(\Delta)$ es la función de mapeo. La salida del extractor es la atención a nivel de oración $\beta = [\beta_1, \beta_2, \dots, \beta_n, \dots]$, donde β_n es la probabilidad de que la n^{th} oración ha sido extraída en el resumen.

Por otro lado, nuestro abstractor basado en la atención calcula la atención a nivel de la palabra $\alpha^t = [\alpha_1^t, \alpha_2^t, \dots, \alpha_m^t, \dots]$, dinámicamente mientras se genera la palabra t^{th} en el resumen. La salida del abstractor es el texto de resumen. $y = [y^1, y^2, \dots, y^t, \dots]$, donde y^t es t^{th} palabra en el resumen.

3.1. Combinando atenciones

El método para unificar los enfoques se basan en [5] que muestra que el mecanismo de atención es muy importante para tareas de PNL (Procesamiento de lenguaje natural). Por lo tanto, combinar explícitamente la atención al nivel de oración β_n y a nivel de

palabra α_m^t por una multiplicación escalar y renormalización. La atención de la palabra actualizado α'_m es:

$$\alpha'_m = \frac{\alpha_m^t \times \beta_{n(m)}}{\sum_m \alpha_m^t \times \beta_{n(m)}}$$

3.2. Extractivo

Para nuestro extractor(resumen extractivo) usaremos el algoritmo de *TextRank* que se base en el algoritmo de *PageRank*.

3.2.1. PageRank

El PageRank, presentado por Brin y Page en [6], es la medida de la importancia de una página web en internet utilizada por el buscador Google. Si consideramos Internet como un grafo dirigido, en el que los nodos son las páginas web y la existencia de una arista entre dos nodos denota que hay un enlace del primero hacia el segundo, se puede extraer un índice de relevancia de cada nodo sin más que considerar la topología del grafo (aunque Google realmente tiene en cuenta muchos más factores). Este índice de relevancia es el PageRank, y se basa, como sustrato teórico, en la existencia de un navegador aleatorio que va visitando páginas web pulsando los enlaces que ve en ellos, o bien se va a una página totalmente distinta. Esto se traduce en lo siguiente: si llamamos $PR(V)$ al PageRank de la página V , y llamamos $In(V)$ al conjunto de las páginas que tienen enlaces hacia V (en la terminología de grafos, los vértices origen de las aristas que tienen como destino V), y $Out(V)$ el conjunto de los vértices hacia los que tiene enlace V (los vértices destino de las aristas que tienen como origen a V) es

$$PR(V) = (1 - d) + d \sum_{W \in In(V)} \frac{PR(W)}{|Out(W)|}$$

donde d es un factor de amortiguación que tiene como objetivo incluir en el modelo la probabilidad de que haya un salto aleatorio de un vértice del grafo a cualquier otro. En el contexto de la navegación en Internet, dicho factor representa la probabilidad de

que un usuario acceda a una página a través de un enlace situado en la página actual, siendo por tanto $(1-d)$ la probabilidad de que dicho usuario salte a una página aleatoria no enlazada con la página actual. En la definición original de PageRank se recomienda un valor de 0,85 [6] para el factor d . El PageRank de cada página web puede calcularse mediante un sencillo algoritmo iterativo y representa una distribución de probabilidad sobre las páginas web.

3.2.2. TextRank

El TextRank [19], es una adaptación del PageRank a un problema muy distinto al anterior: la extracción de palabras clave y resúmenes de textos en el ámbito del lenguaje natural. Una palabra o frase (en general, un vértice de un grafo) tiene un ranking mayor o menor dependiendo de la influencia que sobre ésta tengan las demás; a este ranking se le denomina TextRank. En este contexto, una palabra o frase puede influir más o menos sobre otra, dependiendo de su similitud; por esta razón, y a diferencia del PageRank, la influencia tiene una cierta ponderación: un vértice V_i influye en un vértice V_j con un peso w_{ij} (por tanto, el grafo es ponderado). La expresión del PageRank anterior se convierte en el TextRank en

$$TR(V_i) = (1 - d) + d \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{v_k \in Out(V_j)} w_{jk}} TR(V_j)$$

Por otra parte, la influencia de una palabra o frase sobre otra es recíproca, de ahí que el grafo que describe la situación sea no dirigido ($w_{ij} = w_{ji}$, $In(V) = Out(V)$), lo que simplifica mucho los cálculos.

3.3. Abstractivo

La segunda parte es una abstracción. que lee el artículo; luego, genera un resumen palabra por palabra. Utilizamos la red de puntero-generator. propuesto por [1] y se combina con el extractor mediante la combinación de atenciones a nivel de oración y a nivel de palabra, explicados en la sección 3.1. Antes de explicar a profundidad

sobre la red puntero-generador (*Pointer-generator*), analizaremos el modelo secuencia a secuencia con atención.

3.3.1. Modelo secuencia a secuencia con atención

La red neuronal recurrente (RNN), un tipo de red neuronal que puede realizar cálculos en datos secuenciales (por ejemplo, secuencias de palabras), se ha convertido en el enfoque estándar para muchas tareas de procesamiento del lenguaje natural. En particular, el modelo de secuencia a secuencia con atención. Su estructura y funcionamiento se detalla en la Figura 3.3.1.

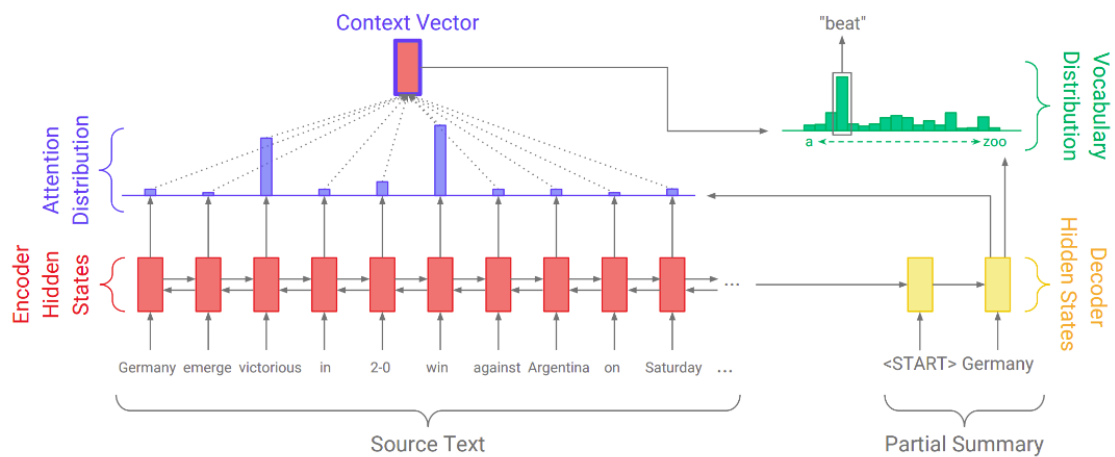


Figura 3.1: Estructura de la red RNN

En la Figura 3.3.1 el texto de origen es un artículo de noticias que comienza a con *Alemania salió victoriosa en la victoria 2-0 contra Argentina el sábado.*, y estamos en el proceso de producir el resumen abstracto *Alemania derrotó a Argentina 2-0*. Primero, el **codificador RNN** lee el texto fuente palabra por palabra, produciendo una secuencia de **estados ocultos del codificador** (Hay flechas en ambas direcciones porque el codificador es bidireccional).

Una vez que el codificador ha leído todo el texto de origen, el **decodificador RNN** comienza a generar una secuencia de palabras que deben formar un resumen. En cada paso, el decodificador recibe como entrada la palabra anterior del resumen (en el

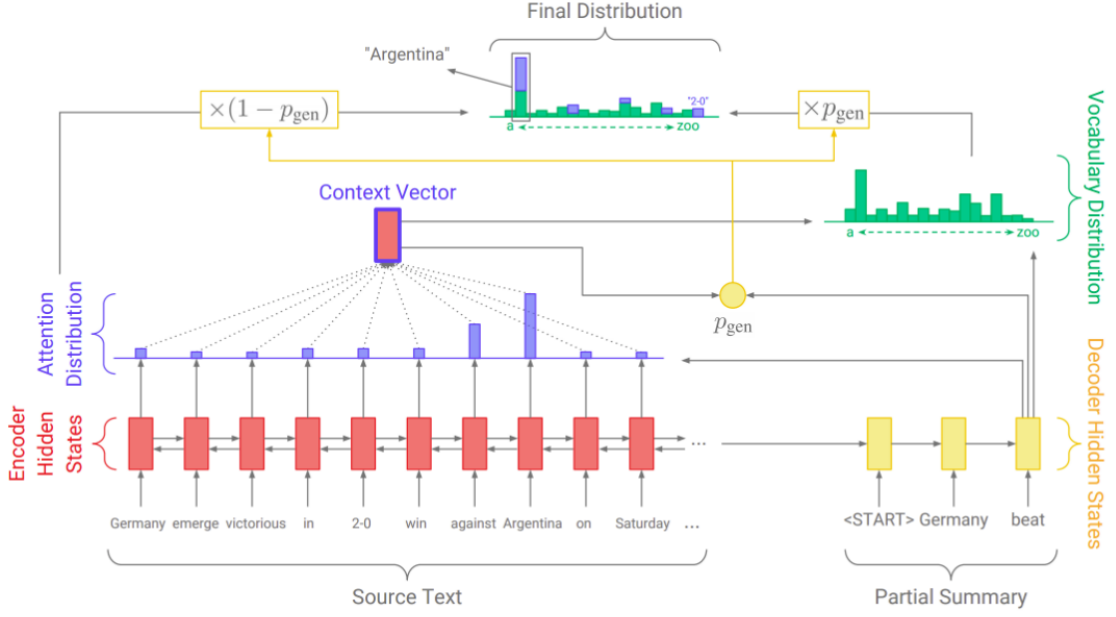
primer paso, este es un token $\langle START \rangle$ especial que es la señal para comenzar a escribir) y lo usa para actualizar el **estado oculto del decodificador**. Esto se utiliza para calcular la distribución de la atención, que es una distribución de probabilidad sobre las palabras en el texto de origen. Intuitivamente, la **distribución de la atención** le dice a la red dónde buscar para ayudarla a producir la siguiente palabra. En el diagrama de arriba, el decodificador ha producido hasta ahora la primera palabra Alemania(Germany), y se concentra en las palabras fuente *win* y *Victorious* para generar la siguiente palabra *beat*.

A continuación, la distribución de atención se usa para producir una suma ponderada de los estados ocultos del codificador, conocido como **vector de contexto**. El vector de contexto puede considerarse como lo que se ha leído del texto de origen en este paso del decodificador. Finalmente, el vector de contexto y el estado oculto del decodificador se utilizan para calcular **la distribución del vocabulario**, que es una distribución de probabilidad sobre todas las palabras de un gran vocabulario fijo (generalmente decenas o cientos de miles de palabras). La palabra con la mayor probabilidad (en este paso, tiempo) se elige como salida y el decodificador pasa al siguiente paso.

La capacidad del decodificador para generar libremente palabras en cualquier orden, incluidas las palabras como el ritmo que no aparecen en el texto de origen, hace que el modelo de secuencia a secuencia sea una solución potencialmente poderosa para el resumen abstractivo.

3.3.2. Red *Pointer-generator*

Una red de puntero-generador es una red híbrida que puede elegir copiar palabras de la fuente a través de apuntar (*pointing*), al tiempo que conserva la capacidad de generar palabras a partir del vocabulario fijo.

Figura 3.2: Estructura de una red *Pointer-generator*

La Figura 3.3.2 muestra el tercer paso del decodificador, cuando hasta el momento hemos generado el resumen parcial de Alemania (Germany). Como antes, calculamos una **distribución de atención** y una **distribución de vocabulario**. Sin embargo, también calculamos la **probabilidad de generación**, p_{gen} , que es un valor escalar entre 0 y 1. Esto representa la probabilidad de generar una palabra del vocabulario, frente a la copia de una palabra de la fuente. La probabilidad de generación p_{gen} se utiliza para el peso y combina la **distribución vocabulariorio** P_{vocab} (que utilizamos para la generación) y la **distribución de la atención** a (que usamos para señalar las palabras de origen w_i) en la distribución final P_{final} a través de la siguiente fórmula:

$$P_{\text{final}}(w) = p_{\text{gen}} P_{\text{vocab}}(w) + (1 - p_{\text{gen}}) \sum_{i:w_i=w} a_i$$

Esta fórmula solo dice que la probabilidad de producir palabra w es igual a la probabilidad de generarlo a partir del vocabulario (multiplicado por la probabilidad de generación) más la probabilidad de señalarlo en cualquier lugar que aparezca en el texto de origen (multiplicado por la probabilidad de copia).

Capítulo 4

Experimentación o evaluación empírica

Presentamos el conjunto de datos y detalles de implementación del método evaluado en nuestros experimentos.

4.1. Base de datos

Evaluamos nuestros modelos en el conjunto de datos de CNN / Daily Mail ([10], [28],[27]) que contiene noticias en los sitios web de CNN y Daily Mail. Cada artículo en este conjunto de datos está emparejado con un resumen de oraciones múltiples escrito por el hombre.

Exactamente: 287,113 pares de entrenamiento, 13,368 pares de validación y 11,490 pares de prueba.

4.2. Entrenamiento

Luego de recibir el ranking de oraciones de nuestro extractor al aplicar *TextRank*, entrenamos nuestro abstractor, configuramos el vocabulario con un tamaño a 50k para el texto de origen y destino. La red con una dimensión oculta de 200. Se limita la

longitud del texto de origen a 450 y la longitud del resumen a 100.

4.3. Métricas para la evaluación

Para evaluar los resultados, se utilizó la medida ROUGE-N [11], que compara un resumen producido automáticamente con un resumen de referencia (generalmente humano), computando el número de N-gramas en común.

Capítulo 5

Resultados y/o Evaluaciones

En este capítulo, se presentarán los resultados obtenidos en las evaluaciones, en la sección 5.1 detallaremos los resultados de nuestro extractor por por separado y en la 5.2 de nuestro abstractor, finalmente detallamos los resultados de el modelo unificado.

5.1. Resultados de las oraciones extraídas

Para evaluar si el extractor utilizando *TextRank* obtiene suficiente información para el abstractor, utilizamos puntajes ROUGE [11].

Metodo	ROUGE-1	ROUGE-2	ROUGE-L
Lead ([23])	45.9	18.0	42.1
Mead ([25])	44.5	20.0	41.0
TextRank ([19])	47.0	19.5	44.07

Cuadro 5.1: Evaluación del extractor con otros modelos mediante las puntuaciones ROUGE

5.2. Resultados del resumen abstractivo

Utilizamos ROUGE-1 de longitud completa, ROUGE-2 y puntuaciones de ROUGE-L para evaluar los resúmenes generados de nuestro abstractor basado en *Pointer-generator* [1].

Metodo	ROUGE-1	ROUGE-2	ROUGE-L
HierAttn ([27])	32.75	12.21	29.01
DeepRL ([29])	39.87	15.82	36.90
GAN ([12])	39.92	17.65	36.71
Pointer-generator	39.53	17.28	36.38

Cuadro 5.2: Evaluación del abstractor con otros modelos mediante las puntuaciones ROUGE

5.3. Resultados Modelo unificado

<p>Artículo: The search for a comic book artist missing in the Cayman Islands since Thursday is now being called a recovery mission.</p> <p>Norman Lee, an artist for DC and Marvel comics, went missing while snorkeling with his wife off the eastern coast of Grand Cayman, CNN affiliate WCVB reported. Strong currents hindered the search, which lasted until Friday evening, Cayman 27 reported. It is unlikely that we will make any recovery at this stage, Chief Inspector Brad Ebanks told Cayman 27. Lee, 47, of Weymouth, Massachusetts, was known and for his work on Wolverine Annual, Supergirl, Starman and other comic book titles. Tributes flooded his Facebook page and Twitter from friends, fans and colleagues who knew him from art school and comic conventions. I cannot express how shaken I am that I will never get the chance to see that smile again, and it saddens me that this world has lost a wonderful man in Norman Lee. To his wife Jan, and his family and all his friends and fans that loved him, my sincerest condolences, friend and fellow graphic artist Chris Kinniery said on Facebook. I'm so sorry to hear about Norman Lee's disappearance. My condolences go out to his family. ... He was an amazing talent in the industry and it was always a pleasure to work with him,"freelance artist</p>
<p>Referencia: comic book artist norman lee went missing in the cayman islands on thursday . authorities called off search on friday evening .</p>
<p>Modelo unificado: norman lee , an artist for dc and marvel comics , went missing while snorkeling with his wife off the eastern coast of grand cayman . lee , 47 , of weymouth , massachusetts , was known and for his work on wolverine annual , supergirl and other comic book titles .</p>

Cuadro 5.3: Comparación del resumen del modelo unificado con y el resumen realizado por una persona

Metodo	ROUGE-1	ROUGE-2	ROUGE-L
Modelo Unificado	40.13	19.68	36.39

Cuadro 5.4: Puntuaciones ROUGE del modelo unificado

5.4. Conclusiones

En la presente tesis utilizó un modelo que combina atenciones en dos niveles, a nivel de oración referente al resumen extractivo y a nivel de palabra referente al resumen abstractivo, aprovechando cada una de sus características.

Se logra un modelo competitivo con los métodos de la literatura y, en varios casos, una calificación ROUGE mejor que otros estudios, siendo el resumen más informativo y legible sobre el conjunto de datos CNN / Daily Mail en una evaluación humana sólida.

Bibliografía

- [1] ABIGAIL SEE, P. J. L., AND MANNING., C. D. Get to the point: Summarization with pointer-generator networks. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (2017), 1073–1083.
- [2] ABIGAIL SEE, P. J. L., AND MANNING., C. D. Get to the point: Summarization with pointer-generator networks. *In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (2017), 1073–1083.
- [3] ALEXANDER M RUSH, S. C., AND WESTON, J. A neural attention model for abstractive sentence summarization. *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (2015), 379–389.
- [4] ANGELA FAN, D. G., AND AULI, M. Controllable abstractive summarization.
- [5] ASHISH VASWANI, NOAM SHAZEER, N. P.-J. U. L. J. A. N. G. U. K., AND POLOSUKHIN., I. Attention is all you need. in advances in neural information processing systems. 6000–6010.
- [6] BRIN, S., P. L. The anatomy of a large-scale hypertextual web search engine.
- [7] CHENG, J., AND LAPATA, M. Neural summarization by extracting sentences and words. *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (2016), 484–494.
- [8] JIATAO GU, ZHENG DONG LU, H. L., AND LI, V. O. Incorporating copying mechanism in sequence-to-sequence learning. *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (2016), 1631–1640.
- [9] JSHASHI NARAYAN, NIKOS PAPASARANTOPOULOS, M. L., AND COHEN, S. B. Neural extractive summarization with side information.
- [10] KARL MORITZ HERMANN, TOMAS KOCISKY, E. G. L. E. W. K. M. S., AND BLUNSOM, P. Teaching machines to read and comprehend. in advances in neural information processing systems. 1693–1701.
- [11] LIN, C.-Y. Rouge: A package for automatic evaluation of summaries. 74–81.
- [12] LINQING LIU, YAO LU, M. Y. Q. Q. J. Z., AND LI, H. Generative adversarial network for abstractive text summarization. *Proceddings of the 2018 Association for the Advancement of Artificial Intelligence* (2017).

- [13] LUHN, H. The automatic creation of literature abstracts. *IBM Journal of Research and Development* (1958), 159–165.
- [14] MANI, I. Automatic summarization . natural language processing.
- [15] MANI, I., M. M. Introduction. in : Advances in automatic text summarization.
- [16] MARC?AURELIO RANZATO, SUMIT CHOPRA, M. A., AND ZAREMBA, W. Sequence level training with recurrent neural networks. 1693–1701.
- [17] MIAO, Y., AND BLUNSOM, P. Language as a latent variable: Discrete generative models for sentence compression. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (2016), 319–328.
- [18] MICHIIHIRO YASUNAGA, RUI ZHANG, K. M. A. P. K. S., AND RADEV, D. Graph-based neural multi-document summarization. In *Proceedings of the 21st Conference on Computational Natural Language Learning* (2017), 452–462.
- [19] MIHALCEA, R., T. P. Textrank: Bringing order into texts.
- [20] MIKAEL KAGEBACK, OLOF MOGREN, N. T., AND DUBHASHI, D. Extractive summarization using continuous vector space models. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)* (2014), 31–39.
- [21] NENKOVA, A.; MCKEOWN, K. Automatic summarization. foundations and trends in information retrieval.
- [22] ORIOL VINYALS, M. F., AND JAITLEY, N. Pointer networks. In *Advances in Neural Information Processing Systems* (2015), 2692–2700.
- [23] PARVEEN, D., R. H.-M., AND STRUBE, M. Topical coherence for graph-based extractive summarization.
- [24] QIAN CHEN, XIAODAN ZHU, Z. L. S. W., AND JIANG, H. Distraction-based neural networks for modeling documents. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (2016).
- [25] RADEV, D. R., A. T. B.-G. S. B. J. C. A. D. S. D. E. H. A. L. W. L. D. E. A. Mead-a platform for multidocument multilingual text summarization.
- [26] RAMESH NALLAPATI, B. Z., AND MA, M. Neural architectures for extractive document summarization.
- [27] RAMESH NALLAPATI, BOWEN ZHOU, C. G.-B. X. Abstractive text summarization using sequence-to-sequence rnns and beyond. 280–290.
- [28] RAMESH NALLAPATI, F. Z., AND ZHOU, B. A recurrent neural network based sequence model for extractive summarization of documents. *Proceedings of the 31st AAAI conference* (2017).

-
- [29] ROMAIN PAULUS, C. X., AND SOCHER, R. A deep reinforced model for abstractive summarization. *In Proceedings of the 2018 International Conference on Learning Representations* (2018).
 - [30] YIN, W., AND PEI, Y. Optimizing sentence modeling and selection for document summarization. *In Proceedings of the 24th International Joint Conference on Artificial Intelligence* (2015), 1383–1389.