

RESUMO DAS TÉCNICAS CITADAS NO ARTIGO A HIERARCHICAL END – TO-END MODEL FOR JOINTLY IMPROVING TEXT SUMMARIZATION AND SENTIMENT CLASSIFICATION.

PHELIPE AUGUSTO TISONI
FRANK COELHO DE ALCANTARA

Este documento apresenta a lista de técnicas citadas no artigo A Hierarchical End-to-End Model for Jointly Improving Text Summarization and Sentiment Classification, escrito por Shuming Ma, Xu Sun, Junyang Lin, Xuancheng Ren. Também existem algumas sugestões de leituras e trabalhos relacionados para futuras pesquisas.

RESUMO DO ARTIGO: Apresenta um modelo de classificação e sumarização hierárquica que melhora conjuntamente tanto a síntese de texto quanto a classificação de sentimentos, comparando dados fornecidos pela Amazon e demonstrando através de testes experimentais a eficiência do modelo em relação à outros.

IMPORTANTE LEMBRAR: Sumarização de texto e classificação de sentimentos, ambos visam capturar as principais ideias do texto, mas em níveis de compreensão diferentes. A sumarização de texto descreve o texto dentro de algumas frases, como um breve resumo da ideia, de forma que pode ser entendida pela linguagem natural. A classificação de sentimentos pode ser considerada como algo mais abstrato, pois ela identifica se a frase foi boa ou ruim, se teve a intenção de demonstrar alegria, tristeza, raiva. Em outras palavras, ela classifica e rotula os sentimentos através de classes e labels.

TÉCNICAS UTILIZADAS:

No artigo foi considerada principalmente a forma de resumo abstrato, classificando o sentimento e atribuindo uma label para determinar a atitude ou a opinião de uma pessoa. Após gerar um resumo de um texto, forma-se um mapeamento entre uma long sequence e uma short sequence.

O modelo consiste em um codificador que transforma o texto original em uma representação para a linguagem natural. E também em um decodificador que gera um resumo do texto. Destacando que o trabalho descrito no artigo concentra-se apenas na sumarização e não na melhora da classificação do sentimento.

Foi proposto um modelo hierárquico end-to-end que consiste em uma camada de sumarização e outra de classificação de sentimentos. A camada de sumarização comprime o texto original em frases curtas e a camada de classificação de sentimento “resume” os textos em uma classe de sentimentos. Essa estrutura hierárquica estabelece um vínculo entre resumo de texto e classificação de sentimento, para que as duas tarefas possam melhorar umas as outras. Após a compressão do texto com a sumarização, torna-se mais fácil para o classificador de sentimentos prever as labels de sentimento do texto resumido. Além disso, o resumo de texto pode indicar palavras importantes e informativas, além de remover palavras redundantes e informações enganosas ou falsas que são prejudiciais para a previsão do sentimento.

O modelo proposto foi avaliado através de datasets on-line da Amazon, demonstrando através de resultados experimentais que ele alcançou um desempenho melhor do que os sistemas de linha de base fortes tanto na sumarização quanto na classificação de sentimentos, contribuindo com:

- A classificação do sentimento como um tipo especial de sumarização e classificação de desempenho e resumo de texto usando um modelo unificado
- Foi proposto uma atenção multi-visão para obter diferentes representações dos textos para resumo e sentimento classificação.
- Resultados experimentais mostram que o modelo supera as linhas de base fortes que treinam a sumarização e classificação do sentimento separadamente.

ARQUITETURA DO MODELO:

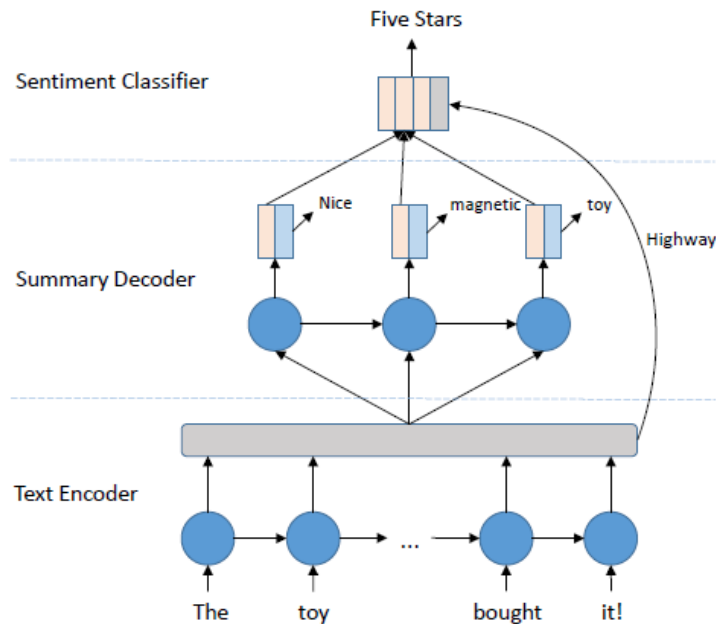


Figure 1: The overview of our model.

O modelo consiste em três componentes, que são: Codificador de texto, Decodificador de resumo e o Classificador de sentimento. O codificador de texto comprime o texto original x na memória de contexto h com um LSTM (Long Short-Term Memory) bidirecional. O decodificador de resumo é um LSTM unidirecional, que gera um vetor de resumo $v(c)$ e um vetor de sentimento $v(t)$ sequencial com mecanismo de atenção, consultando a memória de contexto h . Os vetores de resumo $v(c)$ são usados para gerar o resumo com um gerador de palavras. Os vetores de sentimento $v(t)$ de todas as etapas de tempo são coletados e, em seguida, alimentados no classificador de sentimento para prever a label de sentimento. Para capturar as informações de contexto do texto original, foi usado o mecanismo highway para alimentar a memória de contexto h como parte da entrada do texto.

DECODIFICADOR RESUMIDO COM ATENÇÃO MULTI-VISÃO: O objetivo do decodificador de resumo é gerar uma série de palavras sumárias e fornece as informações resumidas para o classificador de sentimento. No modelo do artigo, o decodificador de resumo consiste em um LSTM unidirecional, um mecanismo de atenção multi-visão e um gerador de palavras. O LSTM gera primeiro as hidden outputs s_t condicionada na informação do resumo gerado.

CLASSIFICADOR DE SENTIMENTO CONSCIENTE DO SUMÁRIO: Após decodificar as palavras até o final do resumo, o modelo coleta os vetores de sentimento de todas as etapas do tempo, concatenando os vetores de sentimento e a representação do texto original. Executando uma operação de maxpooling¹ para obter um vetor de contexto de sentimento.

¹ Maxpooling é um processo de discretização baseado em amostra. O objetivo é reduzir a amostragem de uma representação de entrada (imagem, matriz de saída de camada oculta, etc.), reduzindo sua dimensionalidade e permitindo suposições sobre os recursos contidos nas sub-regiões categorizadas. (RICCO, 2017)

TÉCNICAS CITADAS:

SEQUENCE-TO-SEQUENCE: “A maioria dos modelos existentes é construída para resumir ou classificar algo. Para resumo de texto abstrativo, o modelo mais popular é o sequence-to-sequence model.” (SUSTSKEVER *et al.*, 2014; RUSH *et al.*, 2015).

VARIANTES DO SEQUENCE-TO-SEQUENCE: “Alguns modelos recentes de sumarização abstrata são variantes do modelo sequence-to-sequence” (CHOPRA *et al.*, 2016; VEJA *et al.*, 2007).

ARQUITETURA DE REDE NEURAL: “Para a classificação do sentimento, a maioria dos trabalhos recentes usam a arquitetura de rede neural” (KIM, 2014; TANG *et al.*, 2015). São citadas a LSTM (Long Short-Term Memory) e a CNN (Convolutional Neural Network) para gerar um texto incorporado. E o perceptron multicamadas (multi-layer perceptron - MLP) para prever a label da incorporação.

PRODUZIR TANTO RESUMO QUANTO SENTIMENTOS: “Alguns trabalhos anteriores propõem os modelos para produzir tanto os resumos quanto as labels de sentimentos” (HOLE e TAKALIKAR, 2013; MANE *et al.*, 2015). Porém o autor diz que estes modelos treinam a parte de resumo e sentimento de forma independente, exigindo muito trabalho para obter bons resultados.

RESUMO DE SENTIMENTOS: “Há também alguns trabalhos sobre o sentiment summarization que visam extrair as sentenças de uma certa classe de sentimento dos textos originais (TITOV e MCDONALD, 2008; LERMAN *et al.* 2009).

REDE DE MEMÓRIA DE LONGO / CURTO PRAZO: Foi usado para produzir a memória de contexto uma Rede de Memória de Longo / Curto Prazo Bidirecional (BiLSTM – Bi-Directional Long Short-term Memory Network) no codificador de texto original. Afirmando ser mais popular para o aprendizado de sequence-to-sequence de tarefas de geração de texto incluindo resumo de

texto abstrato. Também possuindo um melhor desempenho em classificação de sentimentos feitos na experiência.

OTIMIZADOR DE ADAM: “Nós minimizamos conjuntamente duas perdas com o otimizador de Adam” (KINGMA e BA, 2014). “Utilizamos a otimização de Adam para treinar o modelo” (KINGMA e BA, 2014).

CONJUNTO DE DADOS DE REVISÃO DA AMAZON (SNAP): Este conjunto de dados faz parte do Stanford Network Analysis Project (SNAP) 1 e é fornecido por Ele e McAuley (2016). Este dataset consiste em comentários e análises de metadados de produtos da Amazon, incluindo 142,8 milhões de avaliações, abrangendo Maio de 1996 a Julho de 2014. Ele inclui conteúdos de revisão, produtos, informações do usuário, classificação e resumos.

ROUGE SCORE: “Para a sumarização abstrata, nossa métrica de avaliação é pontuada por ROUGE que é popular para avaliação de sumarização.” (LIN e HOVY, 2003).

COMPUTAÇÃO DE UNIDADES LEXICAIS SOBREPOSTAS: “Incluem unigram, bigram, trigram e subsequência mais longest commun subsequence (LCS). Seguindo o trabalho prévio, foram usados ROUGE-1 (unigram), ROUGE-2 (bigram) e ROUGEL (LCS) como as métricas de avaliação no relatório experimental de resultados.” (RUSH *et al.*, 2015; HU *et al.*, 2015).

TREINAMENTO DE ÉPOCAS: “Definimos a taxa de aprendizagem $\alpha = 0,0003$, dois parâmetros para os momentos $\beta_1 = 0,9$ e $\beta_2 = 0,999$ respectivamente, e $q = 1 \times 10^{-8}$. Seguindo Sutskever et al. [2014], nós treinamos o modelo para um total de 10 épocas, reduzindo pela metade a taxa de aprendizado a cada meia época, após o início de 5 épocas. Nós clipamos os gradientes [Pascanu et al., 2013] para a norma máxima de 10,0.”

LINHAS DE BASE: “Para a sumarização abstrata, nosso modelo de linha de base é o sequence-to-sequence model [Hu et al., 2015]. Nós denotamos o modelo de sequence-to-sequence sem o mecanismo de atenção S2S,

utilizando o mecanismo S2S-att. Para classificação de texto, comparamos nosso modelo com outros dois modelos de linha de base: **BiLSTM** e **CNN**. O modelo BiLSTM usa um LSTM bidirecional com a dimensão de 256 em cada direção e usa max pooling em todos os estados ocultos do LSTM para obter a incorporação de sentenças em um vetor e, em seguida, usa uma camada de saída MLP com 512 hidden states para produzir o resultado da classificação. O modelo CNN usa o mesmo esquema, mas substitui o BiLSTM por 1 camada de rede convolucional (CNN). Durante o treinamento usamos 0.2 dropout no MLP. Usamos Adam como o otimizador, com uma taxa de aprendizado de 0,001 e um tamanho de batch de 64. Para BiLSTM, nós também cortamos a norma de gradientes para ser 5.0. Nós procuramos hiperparâmetros para encontrar o conjunto mencionado, produzindo uma maior precisão.”

ESTUDO DE ABLAÇÃO: “Para analisar o efeito de cada componente do modelo, nós removemos os componentes de multi-view and highway em ordem, avaliando o desempenho”.

VISUALIZAÇÃO DA ATENÇÃO MULTI-VISÃO: “Apresentamos a pontuação do mapa de calor da atenção em três exemplos. A atenção multi-visão permite a representação do texto da opinião, do sentimento e da visão resumida. Para analisar se a atenção multi-visão capta as informações de sentimento e as informações de resumo, nós fizemos o mapa de calor da atenção da opinião, do sentimento e da atenção da visão resumida, respectivamente. Tomamos a média dos escores de atenção nas saídas do decodificador em todas as etapas do tempo, sendo marcadas as pontuações mais altas com cores escuras e baixa pontuação com cores claras”.

-
- (1) i saw this movie 11 times in the theater and i think that this is one of the best movies ever made and the best movie made about christ and his passion . god bless all those responsible for the creation of this powerful film .
- (2) my daughter , who is now 8 years old , received this as a christmas gift when she was 2 . it has been ready many times , and since been passed along to my son who is now 4 . my children enjoy the tactile quality of the monkeys faces . it is helpful learning counting when there is something they can feel . i have always enjoyed reading the sing song story . it does not take long to read , and after all these years i pretty much have it memorized . a great book , very fun .
- (3) this mattress is too narrow to be comfortable . you fit on it fine but because of the air , i found that it was a balancing act to switch positions . i tried more and less air to no effect . i think if you sleep on your back and stay in that position it would be fine but unfortunately that is not how i sleep . the strong vinyl smell does go away after airing out though .
-

(a) Sentiment view of the original text.

-
- (1) i saw this movie 11 times in the theater and i think that this is one of the best movies ever made and the best movie made about christ and his passion . god bless all those responsible for the creation of this powerful film .
- (2) my daughter , who is now 8 years old , received this as a christmas gift when she was 2 . it has been ready many times , and since been passed along to my son who is now 4 . my children enjoy the tactile quality of the monkeys faces . it is helpful learning counting when there is something they can feel . i have always enjoyed reading the sing song story . it does not take long to read , and after all these years i pretty much have it memorized . a great book , very fun .
- (3) this mattress is too narrow to be comfortable . you fit on it fine but because of the air , i found that it was a balancing act to switch positions . i tried more and less air to no effect . i think if you sleep on your back and stay in that position it would be fine but unfortunately that is not how i sleep . the strong vinyl smell does go away after airing out though .
-

(b) Summary view of the original text.

TRABALHOS RELACIONADOS:

Rush et al. [2015] propõe uma sumarização baseada em summarization models, que usa um codificador attentive CNN para comprimir textos e um modelo de linguagem de rede neural para gerar sumários.

Chopra et al. [2016] explora uma estrutura recorrente para abstractive summarization. Para lidar com problemas de vocabulário,

Nallapati et al. [2016] propõe um modelo gerador de ponteiro para que o decodificador seja capaz de gerar palavras em textos de origem.

Gu et al. [2016] também resolve essa questão incorporando um mecanismo de cópia, permitindo que partes dos resumos sejam copiados do conteúdo de origem.

See et al. [2017] discute ainda mais esse problema e incorpora o modelo ponteiro-gerador com o mecanismo de cobertura

.

Hu et al. [2015] constrói um grande corpus de texto sumarizado de mídias sociais chinesas.

Chen et al. [2016] introduz um modelo neural baseado em distração, que força o mecanismo de atenção a se concentrar nas partes diferentes das entradas de origem.

Ma et al. [2017] propõe um modelo neural para melhorar a relevância semântica entre o conteúdo de origem e os resumos previstos.

SUMARIZAÇÃO E CLASSIFICAÇÃO DE SENTIMENTOS: “Existem alguns trabalhos relacionados com a sumarização e classificação de sentimentos. Buraco e Takalikar [2013] e Mana et al. [2015] propõem os modelos para produzir tanto o resumos quanto as labels de sentimento. No entanto, esses modelos treinam a parte do resumo e da classificação de sentimento de forma independentes, exigindo altos recursos de hand-craft. Cao et al. [2017] propõe um modelo para treinar o gerador de resumo e o classificador de texto em conjunto, mas apenas melhora o desempenho do resumo do texto. Titov e McDonald [2008] propõe um método de sumarização de sentimentos para extrair o resumo dos textos, dada a classe de sentimento. Lerman et al. [2009] constrói um novo sumarizador para treinar um modelo de SVM de classificação sobre o conjunto de julgamentos e preferências humanas, ele melhora o desempenho da sumarização de sentimento. Diferente de todos esses trabalhos, o modelo apresentado pelo artigo melhora tanto o resumo do texto quanto a classificação do sentimento, e não requer nenhum recurso de hand-craft.

REFERÊNCIAS DO ARTIGO:

- [Cao et al., 2017] Ziqiang Cao, Wenjie Li, Sujian Li, and Furu Wei. Improving multi-document summarization via text classification. In AAAI 2017, pages 3053–3059, 2017.
- [Chen et al., 2016] Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. Distraction-based neural networks for modeling documents. In IJCAI 2016, New York, NY, July 2016. AAAI.
- [Cheng and Lapata, 2016] Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. In ACL 2016, 2016.
- [Chopra et al., 2016] Sumit Chopra, Michael Auli, and Alexander M. Rush. Abstractive sentence summarization with attentive recurrent neural networks. In NAACL HLT 2016, pages 93–98, 2016.
- [Gu et al., 2016] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O. K. Li. Incorporating copying mechanism in sequence-to-sequence learning. In ACL 2016, 2016.
- [He and McAuley, 2016] Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In WWW 2016, pages 507–517, 2016.
- [Hole and Takalikar, 2013] Vikrant Hole and Mukta Takalikar. Real time tweet summarization and sentiment analysis of game tournament. International Journal of Science and Research, 4(9):1774–1780, 2013.
- [Hu et al., 2015] Baotian Hu, Qingcai Chen, and Fangze Zhu. LCSTS: A large scale chinese short text summarization dataset. In EMNLP 2015, pages 1967–1972, 2015.
- [Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. In EMNLP 2014, pages 1746–1751, 2014.
- [Kingma and Ba, 2014] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. CoRR, abs/1412.6980, 2014.
- [Lerman et al., 2009] Kevin Lerman, Sasha Blair-Goldensohn, and Ryan T. McDonald. Sentiment summarization: Evaluating and learning user preferences. In EACL 2009, pages 514–522, 2009.
- [Lin and Hovy, 2003] Chin-Yew Lin and Eduard H. Hovy. Automatic evaluation of summaries using n-gram cooccurrence statistics. In HLT-NAACL 2003, 2003.
- [Ma et al., 2017] Shuming Ma, Xu Sun, Jingjing Xu, Houfeng Wang, Wenjie Li, and Qi Su. Improving semantic relevance for sequence-to-sequence learning of chinese socialmedia text summarization. In ACL 2017, pages 635–640, 2017.

[Ma et al., 2018] Shuming Ma, Xu Sun, Wei Li, Sujian Li, Wenjie Li, and Xuancheng Ren. Query and output: Generating words by querying distributed word representations for paraphrase generation. In NAACL 2018, 2018.

[Mane et al., 2015] Vinod L Mane, Suja S Panicker, and Vidya B Patil. Summarization and sentiment analysis from user health posts. In Pervasive Computing (ICPC), 2015 International Conference on, pages 1–4. IEEE, 2015.

[Nallapati et al., 2016] Ramesh Nallapati, Bowen Zhou, Cícero Nogueira dos Santos, Caglar Gulcehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence rnns and beyond. In CoNLL 2016, pages 280–290, 2016.

[Pascanu et al., 2013] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In ICML 2013, pages 1310–1318, 2013.

[Rush et al., 2015] Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In EMNLP 2015, pages 379–389, 2015.

[See et al., 2017] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In ACL 2017, pages 1073–1083, 2017.

[Sun et al., 2017a] Xu Sun, Xuancheng Ren, Shuming Ma, and Houfeng Wang. meprop: Sparsified back propagation for accelerated deep learning with reduced overfitting. In ICML 2017, pages 3299–3308, 2017.

[Sun et al., 2017b] Xu Sun, Bingzhen Wei, Xuancheng Ren, and Shuming Ma. Label embedding network: Learning label representation for soft training of deep networks. CoRR, abs/1710.10393, 2017.

[Sutskever et al., 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In NIPS 2014, pages 3104–3112, 2014.

[Takase et al., 2016] Sho Takase, Jun Suzuki, Naoaki Okazaki, Tsutomu Hirao, and Masaaki Nagata. Neural headline generation on abstract meaning representation. In EMNLP 2016, pages 1054–1059, 2016.

[Tang et al., 2015] Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In EMNLP 2015, pages 1422–1432, 2015.

[Titov and McDonald, 2008] Ivan Titov and Ryan T. Mc-Donald. A joint model of text and aspect ratings for sentiment summarization. In ACL 2008, pages 308–316, 2008.

[Xu et al., 2018a] Jingjing Xu, Xu Sun, Xuancheng Ren, Junyang Lin, Binzhen Wei, and Wei Li. Dp-gan: Diversity-promoting generative adversarial network for generating informative and diversified text. CoRR, abs/1802.01345, 2018.

[Xu et al., 2018b] Jingjing Xu, Xu Sun, Qi Zeng, Xiaodong Zhang, Xuancheng Ren, Houfeng Wang, and Wenjie Li. Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. In ACL 2018, 2018

REFERÊNCIA ADICIONAL:

[Ricco Jay, 2017]. What is max pooling in convolutional neural networks?. Disponível em: <<https://www.quora.com/What-is-max-pooling-in-convolutional-neural-networks>>. Acesso em: 31 jul. 2018.