

Um modelo end-to-end hierárquico para melhorar conjuntamente a síntese de texto e Classificação de Sentimentos

RESUMO: Sumarização de texto e classificação de sentimentos ambos visam capturar as principais idéias do texto, mas Níveis diferentes. Sumarização de texto é descrever o texto dentro de algumas frases, enquanto o sentimento classificação pode ser considerada como um tipo especial de sumarização que “resume” o texto em um forma ainda mais abstrata, ou seja, uma classe de sentimento. Com base nessa ideia, propomos um fim hierárquico Modelo final para aprendizagem conjunta de sumarização de texto e classificação de sentimento, onde o sentimento rótulo de classificação é tratado como o “Resumo” da saída de resumo do texto. Assim, a camada de classificação de sentimento é colocar sobre a camada de resumo de texto, e um hierárquico estrutura é derivada. Resultados experimentais na Amazon on-line conjuntos de dados de avaliações mostram que o nosso modelo atinge melhor desempenho do que o forte sistemas de linha de base em resumo resumido e classificação de sentimentos

INTRODUÇÃO:

Resumo de texto e classificação de sentimento são dois importantes tarefas no processamento de linguagem natural. Resumo de texto visa gerar um resumo com os principais pontos do texto original. Comparado com a sumarização extrativa, que seleciona um subconjunto de palavras existentes no texto original para formar o resumo, a sumarização resumida constrói uma estrutura interna representação semântica e, em seguida, usa a linguagem natural técnicas de geração para criar um resumo mais próximo o que um humano poderia expressar. Neste trabalho, nos concentramos principalmente no resumo do texto abstracto. Classificação de sentimento é atribuir um rótulo de sentimento para determinar a atitude ou a opinião dentro do texto. Também é conhecido como mineração de opinião, derivando a opinião ou a atitude de um falante. Ambos texto sumarização e classificação de sentimento visam a mineração principais idéias do texto. Sumarização de texto descreve o texto com palavras e frases de uma forma mais específica, enquanto o sentimento classificação resume o texto com rótulos em um mais maneira abstrativa.

A maioria dos modelos existentes é construída para resumir ou classificação. Para resumo de texto abstrativo, o modelo mais popular é a sequência-a-sequência modelo [Sutskever et al., 2014; Rush et al., 2015], onde gerando um breve resumo para o texto fonte longo pode ser considerado como um mapeamento entre uma longa sequência e um curto sequência. O modelo consiste em um codificador e um decodificador. O codificador codifica o texto original para um representação e o decodificador gera o resumo. Alguns modelos recentes de resumo resumido são as variantes do modelo sequência-a-sequência [Chopra et al., 2016; Veja et al., 2017]. Para a classificação do sentimento, a maioria dos recentes trabalho usa a arquitetura de rede neural [Kim, 2014; Tang et al., 2015], como LSTM ou CNN, para gerar um texto incorporação, e usar um perceptron multicamadas (MLP) para prever o rótulo da incorporação

Alguns trabalhos anteriores [Hole and Takalikal, 2013; Mane et al., 2015] propõe os modelos para produzir tanto os resumos e os rótulos de sentimento. No entanto, estes modelos treinam a parte de resumo e o sentimento parte de classificação de forma independente, e exigem ricos, artesanato

características. Há também alguns trabalhos sobre o resumo do sentimento [Titov e McDonald, 2008; Lerman et al., 2009], que visam extrair as sentenças com uma certa classe de sentimento dos textos originais. Estes O trabalho concentra-se apenas na sumarização e não melhorar a classificação do sentimento.

Neste trabalho, exploramos um primeiro passo para melhorar tanto o resumo do texto quanto a classificação do sentimento dentro uma estrutura de ponta a ponta. Propomos um end-end hierárquico modelo, que consiste em uma camada de sumarização e camada de classificação de sentimentos. A camada de compactação comprime o texto original em frases curtas e o sentimento camada de classificação "resume" os textos em um classe de sentimento. A estrutura hierárquica estabelece um vínculo entre resumo de texto e classificação de sentimento, para que as duas tarefas possam melhorar umas às outras. Depois de comprimir os textos com sumarização, será mais fácil para o classificador de sentimento para prever os rótulos de sentimento do texto mais curto. Além disso, o resumo do texto pode indicar palavras importantes e informativas, e remover o redundante e informações enganosas que são prejudiciais para prever o sentimento. A classificação do sentimento pode fornecer uma sinal de supervisão para resumos de texto e guias o componente de resumo para capturar a tendência do sentimento do texto original, que pode melhorar a coerência entre o texto curto e o texto original.

Avaliamos nosso modelo proposto nas revisões on-line da Amazon conjuntos de dados. Resultados experimentais mostram que nosso modelo alcança um desempenho melhor do que os sistemas de linha de base fortes tanto na sumarização quanto na classificação dos sentimentos. As contribuições deste documento estão listadas a seguir:

- Nós tratamos a classificação do sentimento como um tipo especial de sumarização e classificação de desempenho e Resumo de texto usando um modelo unificado
- Propomos uma atenção multi-visão para obter diferentes representação dos textos para resumo e sentimento classificação.
- Resultados experimentais mostram que nossos modelos superam as linhas de base fortes que treinam a sumarização e classificação do sentimento separadamente.

Modelo Proposto

Nesta seção, apresentamos nosso modelo proposto em detalhes. Na Seção 2.1, damos a formulação do problema. Nós explicamos a visão geral de nosso modelo proposto na Seção 2.2. Então nós introduzir cada componente do modelo da Seção 2.3 para Seção 2.5. Finalmente, a Seção 2.6 dá a função de perda global e os métodos de treinamento.

2.1 Formulação de Problemas

Dado um conjunto de dados de avaliações on-line que consiste em N amostras de dados, a i -ésima amostra de dados (x_i, y_i, l_i) contém um texto original x_i , um resumo y_i e um rótulo de sentimento l_i . Tanto o original x_i e o resumo y_i são sequências de palavras:

$$x_i = \{x_{i1}, x_{i2}, \dots, x_{iL_i}\}$$

$$y_i = \{y_{i1}, y_{i2}, \dots, y_{iM_i}\}$$

onde L_i e M_i denotam o número de palavras nas seqüências x_i e y_i , respectivamente. O rótulo $l_i \in \{1, 2, \dots, K\}$ denota a atitude de sentimento do conteúdo original x_i , de a classificação mais baixa 1 para a classificação mais alta K .

O modelo é aplicado para aprender o mapeamento da origem texto para o resumo de destino e o rótulo de sentimento. Para o propósito de simplicidade, (x, y, l) é usado para denotar cada par de dados no restante desta seção, onde x é a sequência de palavras de um texto original, y é a sequência de palavras do correspondente resumo e l é o rótulo de sentimento correspondente.

2.2 Visão Geral do Modelo

A figura 1 mostra a arquitetura do nosso modelo. Nosso modelo consiste em três componentes, que são o codificador de texto, o 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 bidirecional. O decodificador de resumo é um LSTM unidirecional, que gera um vetor de resumo $v(c)$ e um vetor de sentimento $v(t)$ sequencialmente com a atenção mecanismo, 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 o rótulo do sentimento. Para capturar as informações de contexto do texto original, usamos o mecanismo da estrada para alimentar a memória de contexto h como parte da entrada do

...

O brinquedo comprou!
 Cinco estrelas
 Bom brinquedo magnético
 Codificador de texto
 Decodificador Resumido
 Classificador de Sentimentos

Figura 1: A visão geral do nosso modelo.

classificador. Portanto, o classificador prevê o rótulo de acordo aos vetores de sentimento do decodificador sumário e ao memória de contexto do codificador de texto.

2.3 Codificador de Texto

O objetivo do codificador de texto de origem é fornecer uma série de representação densa do texto original para o decodificador e o classificador. Em nosso modelo, o codificador de texto original é um Rede de Memória de Longo Prazo de Curto Prazo Bidirecional (BiLSTM), que produz a memória de contexto $h = \{h_1, h_2, \dots, h_L\}$ do texto de origem x :

$$\tilde{h}_t = \tilde{f}(x_t, \tilde{h}_{t-1}) \quad (1)$$

$$\tilde{h}_t = \tilde{f}(x_t, \tilde{h}_{t+1}) \quad (2)$$

$$h_t = \tilde{h}_t + \tilde{h}_t \quad (3)$$

onde \tilde{f} e \tilde{f} são as funções para frente e para trás de LSTM por um passo de tempo, \tilde{h}_t e \tilde{h}_t são a frente e o saídas ocultas para trás respectivamente, x_t é a entrada no t -ésimo passo de tempo e L é o número de palavras na sequência x .

Embora a rede neural convolucional (CNN) seja também uma escolha alternativa para o codificador, BiLSTM é mais popular para o aprendizado de sequência para sequência de tarefas de geração de texto incluindo resumo de texto abstracto. Além disso, de acordo Para as nossas experiências, o BiLSTM consegue um melhor desempenho em classificação de sentimento em nossos conjuntos de dados de benchmark. Nós damos os detalhes da comparação da CNN e da BiLSTM na Seção 3

2.4 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. Em nosso modelo, o decodificador de resumo consiste em um LSTM unidirecional, uma atenção multi-visão mecanismo e um gerador de palavras. O LSTM gera primeiro a saída escondida s_t condicionada na informação histórica do resumo gerado:

$$s_t = f(y_{t-1}, s_{t-1}) \quad (4)$$

onde f é a função do LSTM por um passo de tempo e y_{t-1} são as últimas palavras geradas no t -ésimo passo do tempo.

Dada a saída oculta, implementamos uma atenção multi-visão

mecanismo para recuperar as informações de resumo e a informação de sentimento da memória de contexto h do texto original. A motivação da atenção multi-visão é que o modelo deve se concentrar em diferentes partes do original texto para resumo e classificação. Para resumir, os mecanismos de atenção devem focar no informativo palavras que descrevem melhor os pontos principais. Para a classificação do sentimento, o mecanismo de atenção deve se concentrar nas palavras que contém a tendência mais sentimental, como "ótimo", "Ruim" e assim por diante. Na implementação, a atenção multi-visão gera um vetor de resumo v (c) para resumo:

$$v_t^{(c)} = \sum_{i=1}^N \alpha_{ti} h_i \quad (5)$$

$$\alpha_{ti} = \frac{e^{g(s_t, h_i)}}{\sum_{j=1}^N e^{g(s_t, h_j)}} \quad (6)$$

$$g(s_t, h_i) = \tanh(s_t^T W_t h_i) \quad (7)$$

onde W_t é uma matriz de parâmetros treinável. Semelhante ao resumo vetor, o vetor de sentimento v (t) também é gerado com o mecanismo de atenção seguindo as Equações 5, 6 e 7, mas tem diferentes parâmetros treináveis. A atenção multi-visão pode ser considerado como duas atenções globais independentes para aprender para se concentrar mais no aspecto resumido ou no aspecto do sentimento. Dado o vetor de resumo v (c) t , o gerador de palavras é usado para calcular a distribuição de probabilidade das palavras de saída em t -ésimo passo do tempo:

$$p(y_t|x) = \text{softmax}(W_g v^{(c)} t + b_g) \quad (8)$$

2.5 Summary-Aware Sentiment Classifier

Depois de decodificar as palavras até o final do resumo, o modelo coleta os vetores de sentimento de todas as etapas do tempo:

$$v^{(t)} = [v_1^{(t)}, v_2^{(t)}, \dots, v_M^{(t)}] \quad (9)$$

Então, nós concatenamos os vetores de sentimento de resumo v (t) e a representação do texto original h , e execute um maxpooling operação para obter um vetor de contexto de sentimento r , que denotamos como uma operação de rodovia na Figura 1:

$$\begin{aligned} r &= \max(v^{(t)} \oplus h) \\ &= \max([v_1^{(t)}, v_2^{(t)}, \dots, v_M^{(t)}, h_1, h_2, \dots, h_L]) \end{aligned} \quad (10)$$

onde $(+)$ denota a operação de concatenação ao longo do primeira dimensão, M é o número de palavras no resumo,

e L é o número de palavras no texto original. O sentimento vetor de contexto é então alimentado no classificador para computar a distribuição de probabilidade do rótulo de sentimento $p(l|x)$. o classificador é uma rede de feed-forward de duas camadas com RELU como a função de ativação. O rótulo com a maior probabilidade é o rótulo do sentimento previsto

2.6 Função e Treinamento de Perda Geral

A função de perda consiste em duas partes, que são a cruz perda de entropia de sumarização e de classificação de sentimento:

$$L_s = - \sum_t y_t \log p(y_t|x) \quad (11)$$

$$L_c = -l \log p(l|x) \quad (12)$$

onde y_t e l são a verdade fundamental das palavras e dos rótulos e $p(y_t|x)$ e $p(l|x)$ são a distribuição de probabilidade das palavras e rótulos, calculados pela Equação 8. Nós minimizamos conjuntamente duas perdas com otimizador de Adam [Kingma e Ba, 2014]:

$$L = L_s + \lambda L_c \quad (13)$$

onde λ é um hiper-parâmetro para equilibrar duas perdas. Montamos $\lambda = 0,5$ neste trabalho.

3 experimentos

Nesta seção, avaliamos nosso modelo na Amazônia on-line conjunto de dados de revisão, que contém as resenhas on-line, resumos, e etiquetas de sentimento. Nós primeiro introduzimos os conjuntos de dados, métricas de avaliação e detalhes experimentais. Então, nós comparamos nosso modelo com vários sistemas de linha de base populares. Finalmente, nós fornecemos a análise e a discussão de nossas modelo.

3.1 Conjuntos de Dados

Conjunto de dados de revisão do Amazon SNAP (SNAP): este conjunto de dados faz parte do Stanford Network Analysis Project (SNAP) 1, e é fornecido por Ele e McAuley [2016]. O conjunto de dados consiste em comentários da Amazon e contém análises e metadados de produtos da Amazon, incluindo 142,8 milhões de avaliações abrangendo Maio de 1996 - Julho de 2014. Inclui conteúdo de revisão, produto, informações do usuário, classificações e resumos. Nós emparelharemos cada revisão conteúdo com o resumo e sentimento correspondentes rótulo. Seleccionamos três domínios de análises de produtos para construir três conjuntos de dados de referência, que **são Brinquedos e Jogos, Esportes & Outdoors, e Movie & TV**. Seleccionamos as primeiras 1.000 amostras de cada conjunto de dados como o conjunto de validação, os seguintes 1.000 amostras como o conjunto de testes e o restante como o conjunto de treinamento.

3.2 Métrica de Avaliação

Para resumo resumido, nossa métrica de avaliação é Pontuação ROUGE [Lin e Hovy, 2003], que é popular para avaliação de sumarização. As métricas comparam um automaticamente produziu resumo com os resumos de referência, por computação de unidades lexicais sobrepostas, incluindo unigram, bigram, trigram, e subsequência mais longa comum (LCS). Segue trabalho prévio [Rush et al., 2015; Hu et al., 2015], nós use ROUGE-1 (unigrama), ROUGE-2 (bi-grama) e ROUGEL (LCS) como as métricas de avaliação no relatório experimental resultados.

Para classificação de sentimento, a métrica de avaliação é perlabel precisão. Avaliamos a precisão de ambos os sentimentos de cinco classes, dos quais o sentimento é classificado em 5 classes, e sentimento de duas classes, do qual o sentimento é positivo ou negativo.

3.3 Detalhes Experimentais

Parâmetros do modelo

Os vocabulários são extraídos dos conjuntos de treinamento e os O conteúdo da fonte e os resumos compartilham os mesmos vocabulários. Nós ajustamos os hiper-parâmetros baseados no desempenho nos conjuntos de validação.

Nós limitamos o vocabulário para 50.000 palavras mais frequentes que aparecem no conjunto de treinamento. Nós definimos a palavra incorporação e o tamanho oculto para 256, 512 e 512 para Brinquedos, Esportes e Conjuntos de dados de filmes, respectivamente. A palavra incorporação é aleatória inicializado e aprendido do zero. O codificador é um LSTM bidirecional de camada única, o decodificador é uma camada única LSTM unidirecional, eo classificador é um feedforward de duas camadas rede com uma dimensão oculta de 512. O tamanho do lote é 64, e usamos dropoutwith probabilidade $p = 0,2, 0,05, 0,0$ para conjuntos de dados de brinquedos, esportes e filmes, respectivamente.

Treinamento de Modelo

Usamos a otimização de Adam [Kingma e Ba, 2014] método para treinar o modelo. Para os hiper-parâmetros de Adam otimizador, definimos a taxa de aprendizagem $\alpha = 0,0003$, dois momentos parâmetros $\beta_1 = 0,9$ e $\beta_2 = 0,999$ respectivamente, e $\eta = 1 \times 10^{-8}$. Seguindo Sutskever et al. [2014], nós treinamos o modelo para um total de 10 épocas, e começar a reduzir pela metade o aprendizado taxa a cada meia época após 5 épocas. Nós clipamos os gradientes [Pascanu et al., 2013] para a norma máxima de 10,0.

3.4 Linhas de Base

Para resumo resumido, nosso modelo de linha de base é o modelo sequência-a-sequência para resumo resumido, seguindo o trabalho anterior [Hu et al., 2015]. Nós denotamos o modelo de sequência a sequência sem o mecanismo de atenção como S2S, e que com o mecanismo de atenção como S2S-att.

Para classificação de texto, comparamos nosso modelo com dois modelos de linha de base: **BiLSTM** e **CNN**. Para as duas linhas de base modelos, o modelo BiLSTM usa um LSTM bidirecional com a dimensão de 256 em cada direção e usa pool máximo em todos os estados ocultos do LSTM para obter a incorporação de sentenças vetor e, em seguida, usa uma camada de saída MLP com 512 ocultos estados para produzir o resultado da classificação. O modelo CNN usa o mesmo esquema, mas substitui o BiLSTM por 1 camada de rede convolucional. Durante o treinamento usamos 0.2 desistentes no MLP. Usamos Adam como o otimizador, com um aprendizado taxa de 0,001, e um tamanho de lote de 64. Para BiLSTM, nós também cortar a norma de gradientes para ser 5.0. Nós procuramos hiperparâmetros em uma ampla gama e encontrar o conjunto acima mencionado de Os hiperparâmetros produzem a maior precisão.

Os modelos de linha de base acima exploram apenas parte dos dados anotados. dados (resumos ou rótulos de sentimento). Por mais justa comparação, também implementamos um modelo conjunto de S2S-att e BiLSTM (S2S-att + BiLSTM) e ambos os rótulos anotados de resumos e sentimentos são usados para treinar esta linha de base modelo. Nós comparamos nosso modelo com este modelo, a fim para analisar as melhorias do nosso modelo, dando exatamente os mesmos dados anotados. Neste modelo de linha de base, S2S-att e BiLSTM compartilham o mesmo codificador, e o S2S-att produz o resumo com um decodificador LSTM, enquanto o BiLSTM prevê o rótulo de sentimento com um MLP. Sintonizamos o hiper-parâmetro no conjunto de validação. Nós definimos a palavra incorporação e tamanho oculto para 256, 512 e 512. O tamanho do lote é 64, e a taxa de desistência é $p = 0,15, 0,05, 0,1$ para Brinquedos, Esportes e Conjuntos de dados de filmes, respectivamente.

Toys & Games	RG-1	RG-2	RG-L
S2S [Hu et al., 2015]	14.05	2.47	15.75
S2S-att [Hu et al., 2015]	16.23	4.27	16.01
S2S-att + BiLSTM	16.32	4.43	16.27
HSSC (this work)	18.44	5.00	17.69

Sports & Outdoors	RG-1	RG-2	RG-L
S2S [Hu et al., 2015]	13.38	2.59	13.18
S2S-att [Hu et al., 2015]	15.70	3.61	15.53
S2S-att + BiLSTM	15.75	3.64	15.68
HSSC (this work)	17.85	4.77	17.59

Movie & TV	RG-1	RG-2	RG-L
S2S [Hu et al., 2015]	10.98	2.34	10.77
S2S-att [Hu et al., 2015]	12.17	3.08	11.77
S2S-att + BiLSTM	12.33	3.22	11.92
HSSC (this work)	14.52	4.84	13.42

Tabela 1: Comparação entre o nosso modelo e a sequência-sequência linha de base para sumarização abstractiva na Amazônia Conjuntos de testes SNAP. Os conjuntos de testes incluem três domínios: Toys & Gamse, Esportes e Ar Livre e Filmes e TV. RG-1, RG-2 e RG-L denotam ROUGE-1, ROUGE-2 e ROUGE-L, respectivamente

3.5 Resultados

Denotamos nossa análise e resumo hierárquico Modelo de classificação como HSSC.

Resumo Sumitivo

Primeiro, comparamos nosso modelo com a sequência-a-sequência linha de base nos conjuntos de teste do Amazon SNAP. Nós relatamos o Pontuação ROUGE F1 do nosso modelo e os modelos de referência os conjuntos de teste. Como mostrado na Tabela 1, nosso modelo de HSSC tem um grande margem sobre os modelos S2S e S2s-att em todos os três conjuntos de testes, o que mostra que a supervisão do sentimento rótulos melhora a representação do texto original. Além disso, dados exatamente os mesmos dados anotados (resumo + rótulo de sentimento), nosso modelo HSSC ainda tem uma melhoria sobre a linha de base S2S-att + BiLSTM, que indica que O HSSC aprende uma melhor representação para sumarização. No geral, A HSSC alcança o melhor desempenho em termos de ROUGE-1, ROUGE-2 e ROUGE-L nos três modelos básicos nos três conjuntos de teste.

A tarefa de resumo nos textos de revisão on-line é muito mais difícil e complicar, então os escores ROUGE no O conjunto de dados SNAP é menor do que outros conjuntos de dados de resumo, como DUC. Os documentos nos conjuntos de dados do DUC são originalmente do site news, assim os textos são formais, e os resumos

no DUC são selecionados manualmente e bem escritos. O SNAP o conjunto de dados é construído com as revisões na amazon e as resenhas originais e os resumos correspondentes são informais e cheios de barulho

Toys & Games	5-class	2-class
CNN	70.5	90.2
BiLSTM	70.7	90.9
BiLSTM + S2S-att	70.9	90.9
HSSC (this work)	71.9	91.8

Sports & Outdoors	5-class	2-class
CNN	72.0	91.5
BiLSTM	71.9	91.6
BiLSTM + S2S-att	72.1	91.9
HSSC (this work)	73.2	92.1

Movie & TV	5-class	2-class
CNN	66.9	86.0
BiLSTM	67.8	86.2
BiLSTM + S2S-att	68.0	86.6
HSSC (this work)	68.9	88.4

Tabela 2: Comparação entre o nosso modelo e a sequência-sequência linhas de base para a classificação do sentimento na Amazônia Conjuntos de testes SNAP. Os conjuntos de testes incluem três domínios: Brinquedos e Jogos, Esportes e Ar Livre e Filmes e TV. 5 classes e 2 classes denotam a precisão do sentimento de cinco classes e o sentimento de duas classes, respectivamente.

Classificação de Sentimento

Nós comparamos nosso modelo com duas classificações populares de sentimento métodos, que são CNN e BiLSTM, na Amazônia Conjuntos de testes SNAP. Nós relatamos a precisão do sentimento de cinco grãos e sentimento de duas classes nos conjuntos de teste. Como mostrado em Tabela 2, BiLSTM tem uma pequena melhora em relação à CNN linha de base, mostrando que o BiLSTM tem um melhor desempenho representam os textos sobre esses conjuntos de dados. Portanto, selecionamos BiLSTM como o codificador do nosso modelo. HSSC obtém uma melhor desempenho sobre os dois modelos de linha de base amplamente utilizados em todos os dos conjuntos de testes, principalmente devido ao benefício de dados e melhor representação. Além disso, o HSSC supera linha de base S2S-att + BiLSTM, mostrando que as informações do decodificador resumido ajuda a prever o sentimento rótulos. No geral, o HSSC alcança o melhor desempenho termos de precisão de 5 classes e precisão de 2 classes ao longo dos três modelos de linha de base nos três conjuntos de teste

Realizamos testes de significância com base no teste t. o testes de significância sugerem que a HSSC tem um impacto melhoria em todas as linhas de base, com $p < 0,001$ em

todas as métricas ROUGE para sumarização em três referências conjuntos de dados, p 0,005 para classificação do sentimento em ambos os Brinquedos E conjuntos de dados de jogos e filmes e TV, e p 0,01 para o sentimento classificação nos conjuntos de dados Sports & Outdoors.

3.6 Estudo de ablação

Para analisar o efeito de cada componente, removemos os componentes de multi-view and rodovia em ordem, e avaliar o desempenho do modelo de descanso. Primeiro removemos o atenção multi-visão. Como mostrado na Tabela 3, o modelo sem atenção multi-view tem uma queda de desempenho em ambos Precisão de 5 classes e ROUGE-L. Pode-se concluir que a atenção multi-visão melhora o desempenho de ambas as resumo e classificação de sentimentos. Nós ainda remove a parte da rodovia e encontre o componente beneficia não só a classificação do sentimento, bot também o resumo resumido. O benefício vem principalmente o fato de que o gradiente do classificador de sentimentos pode ser propagada diretamente para o codificador, para que aprenda melhor representação do texto original para classificação e sumarização.

Toys & Games	5-class	RG-L
w/o Multi-View	70.9	16.47
w/o Highway	70.1	16.06
HSSC (Full Model)	71.9	17.69

Sports & Outdoors	5-class	RG-L
w/o Multi-View	72.0	16.36
w/o Highway	71.5	15.73
HSSC (Full Model)	73.2	17.59

Movie & TV	5-class	RG-L
w/o Multi-View	68.1	12.34
w/o Highway	67.7	12.01
HSSC (Full Model)	68.9	13.42

Tabela 3: Estudo de ablação 5-classe denota a precisão de cinco grãos sentimento, e RG-L denota ROUGE-L para a sumarização.

3.7 Visualização da Atenção Multi-Visão

Como mostra a Tabela 4, apresentamos o mapa de calor da atenção pontuações de três exemplos. A atenção multi-visão permite o modelo para representar o 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 damos o mapa de calor do atenção da opinião do sentimento e a atenção da visão resumida, respectivamente. Tomamos a média dos escores de atenção no

saídas do decodificador em todas as etapas do tempo e marque as pontuações mais altas com cor profunda e baixa pontuação com cor clara. A partir de a tabela, concluímos que a atenção vista-sentimento se concentra mais sobre as palavras sentimentais, e. "Melhor", "poderoso", "Ótimo", "divertido" e "confortável". A visão resumida atenção concentra-se nas palavras informativas que melhor descrevem a opinião dos autores, e. "Eu acho que isso é um do melhor filme "e" um ótimo livro, muito divertido ". Além disso, a atenção da opinião do sentimento se concentra mais no indivíduo palavras, enquanto a visão sumária presta mais atenção ao seqüências de palavras. Além disso, a atenção do ponto de vista do sentimento e a atenção da visão resumida compartilha o foco no informativo palavras, mostrando o benefício da atenção

3.7 Visualização da Atenção Multi-Visão

Como mostra a Tabela 4, apresentamos o mapa de calor da atenção pontuações de três exemplos. A atenção multi-visão permite o modelo para representar o 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 damos o mapa de calor do atenção da opinião do sentimento e a atenção da visão resumida, respectivamente. Tomamos a média dos escores de atenção no saídas do decodificador em todas as etapas do tempo e marque as pontuações mais altas com cor profunda e baixa pontuação com cor clara. A partir de a tabela, concluímos que a atenção vista-sentimento se concentra mais sobre as palavras sentimentais, e. "Melhor", "poderoso", "Ótimo", "divertido" e "confortável". A visão resumida atenção concentra-se nas palavras informativas que melhor descrevem a opinião dos autores, e. "Eu acho que isso é um do melhor filme "e" um ótimo livro, muito divertido ". Além disso, a atenção da opinião do sentimento se concentra mais no indivíduo palavras, enquanto a visão sumária presta mais atenção ao seqüências de palavras. Além disso, a atenção do ponto de vista do sentimento e a atenção da visão resumida compartilha o foco no informativo palavras, mostrando o benefício da atenção

4 Trabalhos Relacionados

Rush et al. [2015] primeiro propõe uma sumarização baseada em resumos modelo, que usa um atento codificador CNN para comprimir textos e um modelo de linguagem de rede neural para gerar sumários. Chopra et al. [2016] explora uma recorrência estrutura para resumo resumido. Para lidar com problema de vocabulário, Nallapati et al. [2016] propõe um modelo gerador-ponteiro para que o decodificador seja capaz de gerar palavras em textos de origem. Gu et al. [2016] também resolve isso questão incorporando mecanismo de cópia, permitindo que partes do os resumos a serem copiados do conteúdo de origem. Vejo 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 mídias sociais chinesas resumo resumido de texto. Chen et al. [2016] introduz um modelo neural baseado em distração, que força a atenção mecanismo para 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 o resumos previstos.

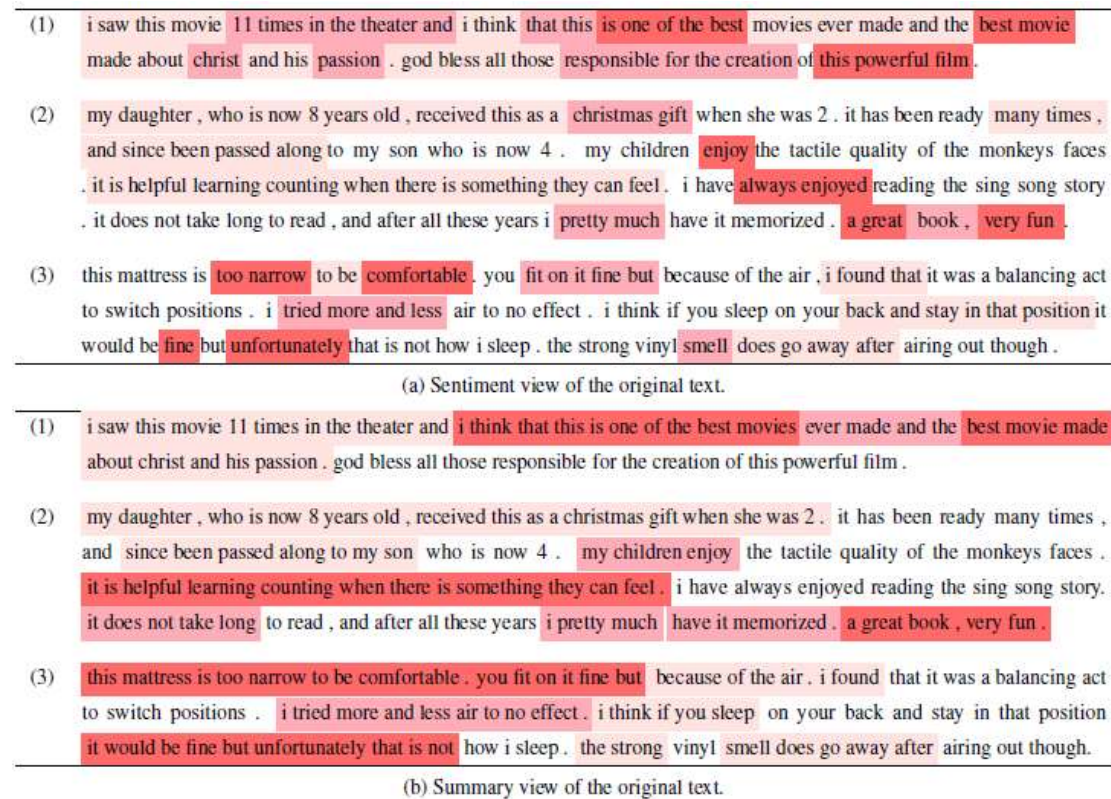


Tabela 4: Visualização da atenção de múltiplas visualizações. Acima está o mapa de calor da atenção da opinião do sentimento, e abaixo está o mapa de calor do atenção de visão resumida. Cores mais profundas significam maiores pontuações de atenção.

Há algum trabalho relacionado com a sumarização e classificação de sentimentos. Buraco e Takalikar [2013] e Mana et al. [2015] propõem os modelos para produzir tanto o resumos e os rótulos de sentimento. No entanto, esses modelos treinar a parte de resumo e a classificação de sentimento parte de forma independente, e exigem recursos ricos hand-craft . Algum trabalho melhorou a sumarização com a ajuda de classificação. 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 uma sumarização de sentimentos método para extrair o resumo dos textos dado o sentimento classe. Lerman et al. [2009] constrói um novo sumariador por treinando um modelo de SVM de classificação sobre o conjunto de preferências humanas julgamentos e melhora o desempenho do sentimento sumarização. Diferente de todos esses trabalhos, nosso modelo melhora tanto o resumo do texto quanto a classificação do sentimento, e não requer nenhum recurso de artesanato.

5. Conclusões

Neste trabalho, propomos um modelo para gerar tanto o sentimento rótulos e os resumos humanos, na esperança de resumir as opiniões dos rótulos de sentimento grosseiro para as sequências de granulação fina. Avaliamos nossa proposta modelo em vários conjuntos de dados de avaliações on-line. Resultados experimentais mostram que nosso modelo alcança um desempenho melhor do que os sistemas de linha de base tanto no resumo resumido quanto classificação de sentimento.

Agradecimentos

Este trabalho foi apoiado em parte pela National Natural Science Fundação da China (No. 61673028), National High Technology Programa de Pesquisa e Desenvolvimento da China (863 Programa, No. 2015AA015404), e do National Thousand Programa Jovens Talentos. Xu Sun é o autor correspondente deste artigo.