

A Details of the Corpus

The CNN/Daily Mail corpus (Hermann et al., 2015) contains English news articles and associated human-written summaries and is the most popular large-scale benchmark in news summarization. We used the pre-processed dataset as in (See et al., 2017), which has 287,226 training pairs, 13,368 validation pairs, and 11,490 test pairs. We did not replace the name entities with anonymised identifiers, and used the same sentence segmentation for documents and summaries as in (Liu and Lapata, 2019). To obtain the word embedding representation, we tokenized the sentences with the sub-word algorithm from BERT (Devlin et al., 2019).

B Implementation Details of Unsupervised Analysis Model

In this section, we provide implementation details of the model in Section 4.3: an autoencoder with adversarial training strategy.

Encoding Component: Given a document representation vector \mathbf{v}_{doc} , and a sentence representation vector \mathbf{v}_{sen} as input, the encoding component (two linear layers) compresses it to a lower dimension, namely the latent feature vector \mathbf{v}_{latent} . In our setting, the hidden dimensions of \mathbf{v}_{doc} , \mathbf{v}_{sen} and \mathbf{v}_{latent} are 768, 768 and 10, respectively. \mathbf{h}_{enc} is the hidden vector, defined as:

$$\mathbf{h}_{enc} = \text{LeakyRelu}(\mathbf{W}[\mathbf{v}_{doc}; \mathbf{v}_{sen}] + \mathbf{b})$$

the latent feature vector is defined as:

$$\mathbf{v}_{latent} = \text{Sigmoid}(\mathbf{W}\mathbf{h}_{enc} + \mathbf{b})$$

where \mathbf{W} and \mathbf{b} are trainable parameters in each layer, and $;$ denotes the concatenation operation.

Decoding Component: Given a latent feature representation vector \mathbf{v}_{latent} and a document representation \mathbf{v}_{doc} as input, the decoding component (two linear layers) is targeted to reconstruct the sentence representation \mathbf{v}_{sen} .

$$\mathbf{h}_{dec} = \text{LeakyRelu}(\mathbf{W}[\mathbf{v}_{doc}; \mathbf{v}_{latent}] + \mathbf{b})$$

$$\mathbf{s}_{dec} = \mathbf{W}\mathbf{h}_{dec} + \mathbf{b}$$

where \mathbf{h}_{dec} and \mathbf{s}_{dec} are the hidden state and reconstruction output, respectively.

Adversarial Decoding Component: Given a latent feature representation vector \mathbf{v}_{latent} as input, the adversarial decoding component (one linear

layer) is targeted to reconstruct the sentence representation \mathbf{v}_{sen} .

$$\mathbf{s}_{adv} = \mathbf{W}\mathbf{v}_{latent} + \mathbf{b}$$

where \mathbf{s}_{adv} is the reconstruction output.

Training Procedure and Setting: During each training batch, there is a two-step parameter update:

1) Update the adversarial decoder with Mean Square Error (MSE) loss between \mathbf{s}_{adv} and \mathbf{v}_{sen} .

$$\text{loss}_{adv} = \text{MSE}(\mathbf{s}_{adv}, \mathbf{v}_{sen})$$

2) Update the autoencoder with MSE loss between \mathbf{s}_{adv} and \mathbf{v}_{sen} , combined with a penalty from the adversarial MSE to reduce the unnecessary information leaked from \mathbf{v}_{sen} in the encoding component. The adversarial loss is defined as:

$$\text{loss}_{adv} = \text{MSE}(\mathbf{s}_{dec}, \mathbf{v}_{sen}) - \lambda \text{MSE}(\mathbf{s}_{adv}, \mathbf{v}_{sen})$$

where $\lambda = 0.2$ in our training setting.

Adam optimizer (Kingma and Ba, 2015) was used with learning rate of $1e-3$ and weight decay of $1e-3$. Batch size was set to 64. Drop-out (Srivastava et al., 2014) of $rate = 0.1$ was applied in each linear layer. BERT parameters were fixed during training. The trainable parameter size was 0.6M. Tesla V100 with 16G memory was used for training and we used the Pytorch 1.4.0 as computational framework⁴.

C Implementation Details of Neural Selector Model

In this section, we provide implementation details of the model in Section 5.2: a neural sentence selector for extractive summarization.

BERT Encoding Component: Given a document \mathbf{D} containing a number of sentences $[\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_n]$ as input, the encoding component produces the sentence representation \mathbf{h}_i from each \mathbf{s} , which is a list of tokens $[\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_m]$. Here we use the average of the token-level hidden states in the last layer of BERT as \mathbf{h}_i .

$$\mathbf{h}_i = \frac{1}{m} \sum_i^m \mathbf{w}_i^{BertRep}$$

Selection Modeling Component: Given the specific control code \mathbf{v}_{ctrl} and sentence vectors $\mathbf{H} =$

⁴<https://github.com/pytorch/pytorch>

$[\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_n]$ as input, this component use a two-layer bi-directional LSTM to model the contextual information with sub-aspect conditioning. The forward and backward hidden states are concatenated as output.

$$\begin{aligned}\mathbf{u}_i^{forward} &= \text{LSTM}_f([\mathbf{h}_{i-1}; \mathbf{v}_{ctrl}]) \\ \mathbf{u}_i^{backward} &= \text{LSTM}_b([\mathbf{h}_{i+1}; \mathbf{v}_{ctrl}]) \\ \mathbf{u}_i &= [\mathbf{u}_i^{forward}; \mathbf{u}_i^{backward}]\end{aligned}$$

where the input embedding dimension, hidden size, and control code dimension is 768, 384 and 3 respectively. $;$ denotes the concatenation operation.

Output Component: A linear layer is used to produce output \mathbf{y}_i for each sentence, as the probability of being included in the generated summary.

$$\mathbf{y}_i = \text{Sigmoid}(\mathbf{W}\mathbf{u}_i + \mathbf{b})$$

Training Setting: Binary cross entropy (BCE) is used to measure the loss between the prediction y_i and the ground-truth \hat{y}_i for all time steps:

$$loss = \sum \text{BCELoss}(\hat{\mathbf{y}}_i, \mathbf{y}_i).$$

Adam optimizer (Kingma and Ba, 2015) was used with learning rate of $3e-4$ and weight decay of $1e-4$. Batch size was set to 64. Drop-out (Srivastava et al., 2014) of $rate = 0.2$ was applied in the modeling layer and output linear layer. BERT parameters were fixed during training. Lengthy documents were not truncated. The trainable parameter size was 14M (excluding the pre-trained language model). We trained the model for 20 epochs (about 12 hours) on the Tesla V100 GPU. The reported models were selected based on the best validation performance, according to the inflection point of loss value.

D Generated Summary Examples

We show summary examples from the golden groundtruth, oracle, the baseline BertEXT, and the proposed conditional neural summarizer for three news articles in Figure 10.

<p>#Document# Kevin de bruyne 's agent has denied a deal in his place for his client to leave wolfsburg in the summer but admitted that a number of clubs are keeping an eye on the belgian star . De bruyne has become one of the hottest prospects in european football after starring in the bundesliga - being linked with manchester city and bayern munich . Speaking to focus online , patrick de koster revealed that the midfielder is a wanted man but no deal for a move away is in place . <u>Kevin de bruyne (right) is wanted by a host of clubs but no transfer away from wolfsburg has been agreed</u> . He said : ' there are always interested teams and there has been some informal contact . ' people have asked me how kevin is doing , how he has been developing superbly and putting in good performances . That is only normal in situations like this . There have not been any concrete offers , though . German champions bayern , along with city and manchester united are known to have watched de bruyne but de koster is keen to speak to wolfsburg and show them the respect he feels they deserve . The belgian is one of the hottest prospects in european football after starring for his side this season ' I can assure you there have not been any talks with bayern munich . When you want to know whether bayern are interested in signing kevin , you should ask matthias sammer or michael rescike . It is true that i have worked with rescike before , but that does not mean anything now . ' we will sit down with wolfsburg first , that is a matter of respect . They supported kevin during a difficult period , when he was unhappy at chelsea . ' <u>Since leaving chelsea , de bruyne has established himself as one of the best playmakers in the bundesliga</u> .</p>
<p>#Gold# Kevin de bruyne has been linked with a transfer away from wolfsburg . His agent patrick de koster says that clubs are watching the belgian but denies a deal is in place with any team to buy the talented playmaker . De bruyne has starred for the german side this season in the bundesliga .</p>
<p>#Oracle# Kevin de bruyne 's agent has denied a deal in his place for his client to leave wolfsburg in the summer but admitted that a number of clubs are keeping an eye on the belgian star . De bruyne has become one of the hottest prospects in european football after starring in the bundesliga - being linked with manchester city and bayern munich . Kevin de bruyne (right) is wanted by a host of clubs but no transfer away from wolfsburg has been agreed</p>
<p>#BertEXT# De bruyne has become one of the hottest prospects in european football after starring in the bundesliga - being linked with manchester city and bayern munich . Speaking to focus online , patrick de koster revealed that the midfielder is a wanted man but no deal for a move away is in place . Kevin de bruyne (right) is wanted by a host of clubs but no transfer away from wolfsburg has been agreed .</p>
<p>#Ours# [0,1,0] Kevin de bruyne (right) is wanted by a host of clubs but no transfer away from wolfsburg has been agreed . German champions bayern , along with city and manchester united are known to have watched de bruyne but de koster is keen to speak to wolfsburg and show them the respect he feels they deserve . Since leaving chelsea , de bruyne has established himself as one of the best playmakers in the bundesliga .</p>

<p>#Document# It 's not intended to be a safety video . But this stomach-churning clip of a longboarder crashing downhill at high speedcertainly serves as a warning to wear a helmet . <u>Footageshows the skater confidently sailing down a concrete slope in santa monica ,los angeles , before he loses balance and falls headfirst</u> . For a short moment he scrapes along the ground on his stomach . He then rolls over several times . Finally he comes to a stop and as he lies on his side , he clutches hishead in pain . The skater 's friends then rush over tocheck if he 's okay . Going , going : thisstomach-churning clip of a longboarder crashing downhill at high speedcertainly serves as a warning to wear a helmet Footageshows the skater confidently sailing down a concrete slope in santa monica ,los angeles , before he loses balance and falls headfirst Painful : for a short moment he scrapes along the floor on his stomach. <u>As they survey the damage , blood is seen on the groundand pouring from the boarder 's face</u> . The filmer laterwrote on the video-sharing site newswire that the injury required a total ofseven stitches . ' We were very happy that he olyreceived minor injuries , being road rash and two deep cuts in his face . ' thecameraman added . The incident took place on april 1 on asunny day in the morning . Luckily there was no one inthe skater 's path and he was the only one injured . Ouch: finally he comes to a standstill and as he lies on his side , he clutches hishead in apparent pain Bloody : the filmer later revealedthat the injury required a total of seven stitches Warning for next time : luckily there was no one in the skater 's pathand he was the only one injured .</p>
<p>#Gold# Footage showsthe skater confidently sailing down a concrete slope in santa monica , losangeles , before he loses balance and falls headfirst. As his friendsgo to investigate , blood is seen pouring from his face .</p>
<p>#Oracle# Footage shows the skater confidently sailing down aconcrete slope in santa monica , los angeles , before he loses balance andfalls headfirst. As they survey the damage , blood is seen on theground and pouring from the boarder 's face .</p>
<p>#BertEXT# Footage shows the skater confidently sailing down a concrete slope insanta monica , los angeles , before he loses balance and falls headfirst . For a short moment he scrapes along on hisstomach . He then rolls over several times .</p>
<p>#Ours# [1,0,0] Footage shows the skaterconfidently sailing down a concrete slope in santa monica , los angeles ,before he loses balance and falls headfirst . As theysurvey the damage , blood is seen on the ground and pouring from the boarder 'sface . The filmer later wrote on the video-sharing sitenewswire that the injury required a total of seven stitches .</p>

<p>#Document# Downaugusta way they say the azaleas are in full bloom , which is more than can besaid for england 's justin rose . <u>A bruising floridaswing last month saw the englishman fall outside the world 's top 10</u> . For a player who has been virtually a fixture in the topfive for the last three years it was certainly a dent to the ego , with themasters now just around the corner . Rose 's solution tohis miserable form -- three missed cuts and a 55th-place finish at the cadillacchampionship in four pga tour starts -- was the time-honoured one . For the past two weeks , the 34-year-old has spent longhours on the practice ground . <u>Justin rose hit 17 out of18 greens in regulation and signed for a 69 at the shell houston open</u> . In the first round of the shell houston open on thursdaythere were encouraging signs his decline will prove temporary . <u>Rose hit 17 out of 18 greens in regulation and signed for a 69 , the samescore as his playing partner , the ever - consistent jordan spieth</u> . ' It 's certainly a welcome return to the sixties , forit had been a while , ' said rose , smiling . On a daywhen american scott piercy went round in 63 and phil mickelson enjoyed his bestround in months with a 66 , it was hardly surprising the only reporter waitingto talk to rose was this one . But under the radar isnever a bad place to be going to the masters . The boomand bust years that characterised the first half of rose 's career meant therewas never going to be any feelings of panic following his unusually poor run inthe sunshine state . ' There 's no doubt i lost my gamethere but the florida swing can be unforgiving if you 're slightly off , ' hesaid . ' Over the past two weeks i feel like i 've done some good work and whether i finish well or not here i feel like i 'm going inthe right direction again . ' Basically i was gettingahead of the ball at impact , and shots were going left or right , the ironswere not solid and the new putter was not working . So we've corrected the faults and i 've gone back to the old putter . ' <u>Phil mickelson enjoyed his best round in months with a 66on thursday</u> . Does he pay much attention to the worldrankings ? ' You notice , for sure , ' he said . ' i 'm very proud of the fact i 've been in the world 'stop five for practically the whole of the last three years . It 's a nice ego thing , so by the end of the year i 'm hoping there won't be any slippage . ' Rose might be the only player in masters historyto have led or had a share of the lead after every round bar the last one . His best chance so far of finishing on top after 72 holescame in 2007 , when he stood on the 17th hole tied for the lead only to run up a fatal double-bogey six . <u>Paul casey celebrated his last-gaspmasters invitation with two eagle threes in a fine round of 68</u> . ' The masters has probably been less on my mind this year because i amtrying to find some form , ' he admitted . ' But i thinkthe fact i 've had a number of great rounds there will always stand me in goodstead . Regardless of what happens here , i feelcomfortable on that course and know i can manage my game even if it 's not 100per cent . You draw off the energy of the place . ' Mickelson has certainly done that over the years andperhaps the veteran lefty , a three-time masters champion , is gearing himselfup for another run at the green jacket . ' It was a goodstart to the tournament and now i 'm looking for three more good rounds , ' hesaid . ' This is a big week for me . I felt the game was close last week . The onlything missing was chipping and short game . ' Paul casey , like mickelson another former winner of this event , celebrated his last-gaspmasters invitation with a fine round of 68 notable for two eagle threes . <u>In the afternoon wave , padraig harrington and leewestwood both made good starts to play their first six holes in two under</u> .</p>
<p>#Gold# Justin rose bounced back from floridamisery by carding 69 in houston . Three-time masters champion phil mickelson enjoyed return to form . Paul casey celebrated last-gaspmasters invitation with fine round of 68 .</p>
<p>#Oracle# Justin rose hit 17out of 18 greens in regulation and signed for a 69 at the shell houston open . Mickelson has certainly done that over the yearsand perhaps the veteran lefty , a three-time masters champion , is gearinghimself up for another run at the green jacket . Paul casey celebrated his last-gasp masters invitation with two eaglethrees in a fine round of 68 .</p>
<p>#BertEXT# A bruising florida swing lastmonth saw the englishman fall outside the world 's top 10 . Rose hit 17 out of 18 greens in regulation and signed for a 69 , the samescore as his playing partner , the ever - consistent jordan spieth . Phil mickelson enjoyed his best round inmonths with a 66 on thursday .</p>
<p>#Ours# [1,1,0] Rose hit 17 out of 18 greens in regulation and signed for a 69 , the same score as his playing partner , the ever - consistent jordan spieth . Phil mickelson enjoyed his best round in months with a 66 on thursday . Paul casey celebrated his last-gasp masters invitation with two eagle threes in a fine round of 68 . In the afternoon wave , padraig harrington and lee westwood both made good starts to play their first six holes in two under .</p>

Figure 10: Three news article examples. Oracle summaries are underlined, summaries from a baseline model are highlighted in blue, summaries from our model with specified control codes are in orange, and their overlaps are in purple.