

	Rouge-1	Rouge-2	Rouge-L
Lead-3	39.2	15.7	35.5
(Nallapati et al. 2016)	35.4	13.3	32.6
SummaRuNNer-abs	37.5	14.5	33.4
SummaRuNNer	39.6±0.2*	16.2±0.2*	35.3±0.2

Table 3: Performance comparison of abstractive and extractive models on the entire CNN/Daily Mail test set using **full-length F1** variants of Rouge. SummaRuNNer is able to significantly outperform the abstractive state-of-the-art as well as the Lead-3 baseline (on Rouge-1 and Rouge-2).

	Rouge-1	Rouge-2	Rouge-L
Lead-3	43.6	21.0	40.2
LReg	43.8	20.7	40.3
ILP	45.4	21.3	42.8
TGRAPH	48.1	24.3*	-
URANK	48.5*	21.5	-
Cheng <i>et al</i> '16	47.4	23.0	43.5
SummaRuNNer-abs	44.8	21.0	41.2
SummaRuNNer	46.6 ±0.8	23.1 ±0.9	43.03 ±0.8

Table 4: Performance of various models on the **DUC 2002** set using the **limited length recall** variants of Rouge at **75 words**. SummaRuNNer is statistically within the margin of error at 95% C.I. with respect to (Cheng and Lapata 2016), but both are lower than state-of-the-art results.

4.7 Results on the Out-of-Domain DUC 2002 corpus

We also evaluated the models trained on the DailyMail corpus on the out-of-domain DUC 2002 set as shown in Table 4. SummaRuNNer is again statistically on par with the model of (Cheng and Lapata 2016). However, both models perform worse than graph-based TGRAPH (Parveen, Ramsi, and Strube 2015) and URANK (Wan 2010) algorithms, which are the state-of-the-art models on this corpus. Deep learning based supervised models such as SummaRuNNer and that of (Cheng and Lapata 2016) perform very well on the domain they are trained on, but may suffer from domain adaptation issues when tested on a different corpus such as DUC 2002. Graph based unsupervised approaches, on the other hand, may be more robust to domain variations.

5 Qualitative Analysis

In addition to being a state-of-the-art performer, SummaRuNNer has the additional advantage of being very interpretable. The clearly separated terms in the classification layer (see Eqn. 6) allow us to tease out various factors responsible for the classification of each sentence. This is illustrated in Figure 2, where we display a representative document from our validation set along with normalized scores from each abstract feature responsible for its final classification. Such visualization is especially useful in explaining to the end-user the decisions made by the system.

We also display a couple of example documents from the Daily Mail and DUC corpora highlighting the sentences chosen by SummaRuNNer and comparing them with the gold summary in Table 5. The examples demonstrate quali-

Document: @entity0 have an interest in @entity3 defender @entity2 but are unlikely to make a move until january . the 00 - year - old @entity6 captain has yet to open talks over a new contract at @entity3 and his current deal runs out in 0000 . @entity3 defender @entity2 could be targeted by @entity0 in the january transfer window @entity0 like @entity2 but do n't expect @entity3 to sell yet they know he will be free to talk to foreign clubs from january . @entity12 will make a 0million offer for @entity3 goalkeeper @entity14 this summer . the 00 - year - old is poised to leave @entity16 and wants to play for a @entity18 contender . @entity12 are set to make a 0million bid for @entity2 's @entity3 team - mate @entity14 in the summer

Gold Summary: @entity2 's contract at @entity3 expires at the end of next season . 00 - year - old has yet to open talks over a new deal at @entity16 . @entity14 is poised to leave @entity3 at the end of the season

Document: today , the foreign ministry said that control operations carried out by the corvette spiro against a korean-flagged as received ship fishing illegally in argentine waters were carried out " in accordance with international law and in coordination with the foreign ministry " . the foreign ministry thus approved the intervention by the argentine corvette when it discovered the korean ship chin yuan hsing violating argentine jurisdictional waters on 00 may the korean ship , which had been fishing illegally in argentine waters , was sunk by its own crew after failing to answer to the argentine ship 's warnings . the crew was transferred to the chin chuan hsing , which was sailing nearby and approached to rescue the crew of the sinking ship

Gold Summary: the korean-flagged fishing vessel chin yuan hsing was scuttled in waters off argentina on 00 may 0000 . adverse weather conditions prevailed when the argentine corvette spiro spotted the korean ship fishing illegally in restricted argentine waters . the korean vessel did not respond to the corvette 's warning . instead , the korean crew sank their ship , and transferred to another korean ship sailing nearby . in accordance with a uk-argentine agreement , the argentine navy turned the surveillance of the second korean vessel over to the british when it approached within 00 nautical miles of the malvinas (falkland) islands .

Table 5: Example documents and gold summaries from Daily Mail (top) and DUC 2002 (bottom) corpora. The sentences chosen by SummaRuNNer for extractive summarization are highlighted in bold.

tatively that SummaRuNNer performs a reasonably good job in identifying the key points of the document.

6 Conclusion

In this work, we propose a very interpretable neural sequence model for extractive document summarization that allows intuitive visualization, and show that it is better performing than or is comparable to the state-of-the-art deep learning models.

We also propose a novel abstractive training mechanism to eliminate the need for extractive labels at training time, but this approach is still a couple of Rouge points below our extractive training on most datasets. We plan to further explore combining extractive and abstractive approaches as part of our future work. One simple approach could be to pre-train the extractive model using abstractive training. Further, we plan to construct a joint extractive-abstractive model where the predictions of our extractive component form stochastic intermediate units to be consumed by the abstractive component.