

Supplemental Materials

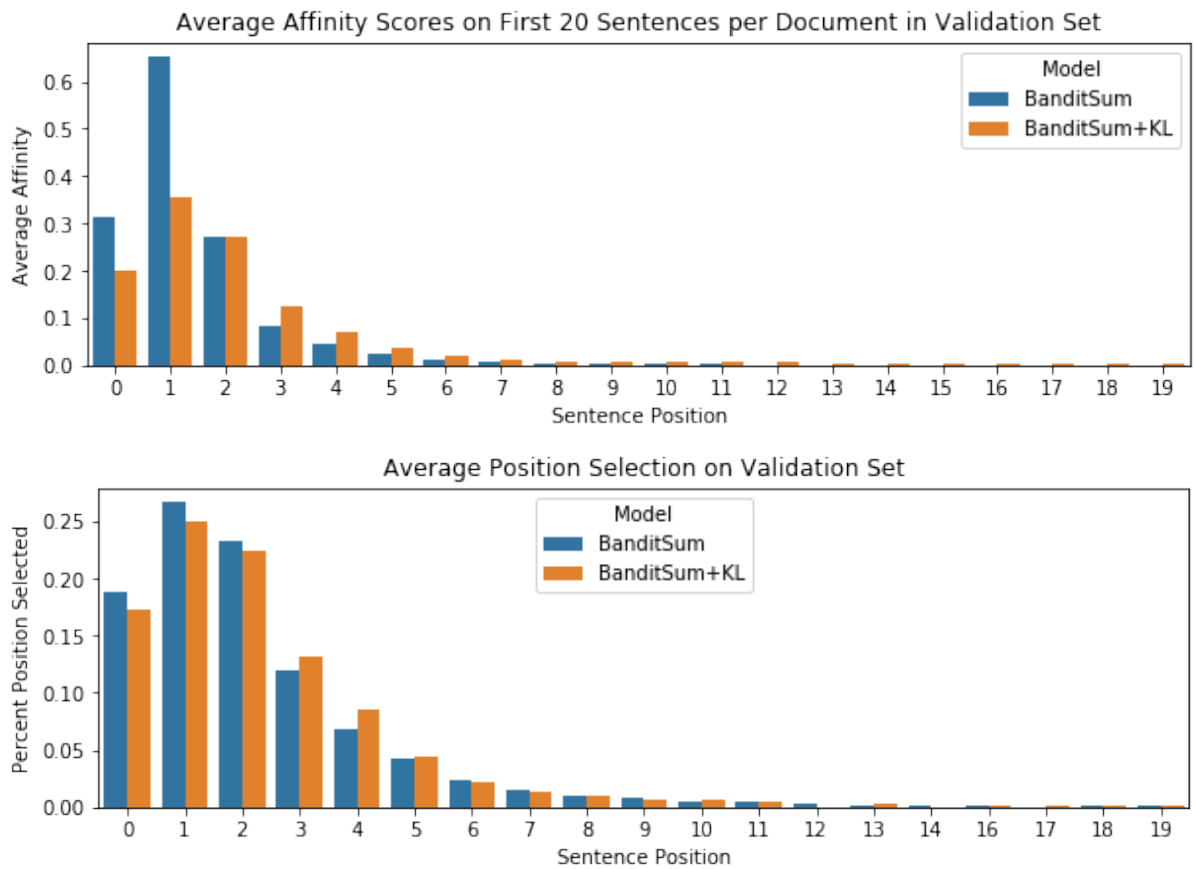


Figure 1: (Top) Average affinity scores for the BanditSum and BanditSum+KL models. Note that the original model has far higher affinity on average for the first two sentence positions, while our proposed model has a more even distribution. (Bottom) Average position selected for the same models. We observe similar trends as the affinity score distribution.

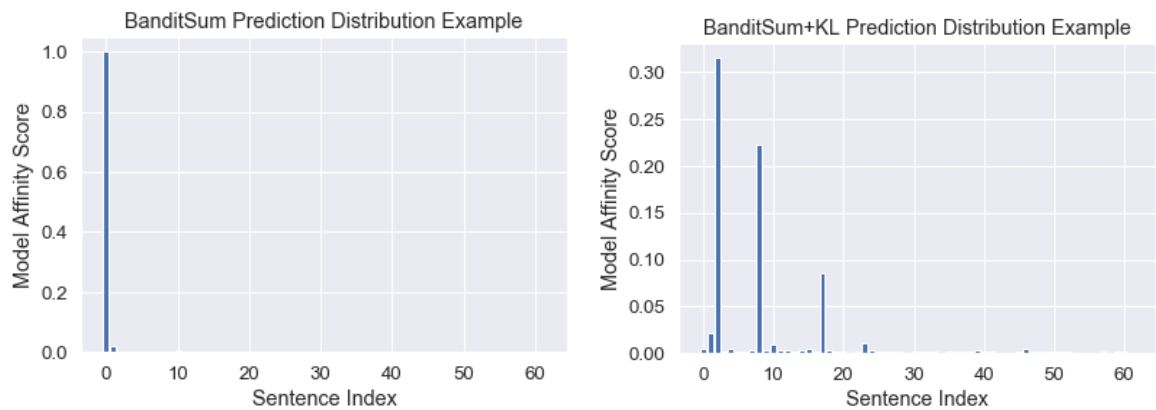


Figure 2: Example of the BanditSum prediction distribution (left) vs. the BanditSum+KL prediction distribution (right) on the same given article from the validation set. The original formulation is more prone to lead bias compared to our proposed method, as demonstrated here. The article in question is presented on the next page.

<p>Article (truncated):</p> <p>-lrb- cnn -rrb- on the 6th of april 1996 , san jose clash and dc united strode out in front of 31,683 expectant fans at the spartan stadium in san jose , california . the historic occasion was the first ever major league soccer match – a brave new dawn for the world ’s favorite sport in a land its charms had yet to conquer . summarizing the action for espn , commentator ty keough eagerly described the momentous ” birth of a new era for american soccer . ” looking back at footage from that balmy evening now it ’s hard not to feel a certain nostalgia . baggy shirts , questionable hairstyles and strange rule adaptations to make games more exciting were all part of the formative mls experience . countdown clocks were employed to provide drama at the end of each half . even more bizarrely, tied games were settled by shootouts that saw attacking players run with the ball from 35-yards out before attempting to beat the opposing goalkeeper . as the mls prepares to mark the beginning of its 20th season , it ’s hard to comprehend just how much the league has progressed in the intervening period . long gone is the desire to tamper with the rules of the game for a start . attendances are higher than ever before while the number of teams involved has doubled from 10 in the 1996 campaign to 20 in 2015 . a further four are set to be added by 2020 . on top of this , the new season is the first of a new domestic tv and media rights deal with fox , espn and univision worth \$ 700 million over eight years . this figure may pale beside the \$ 5.1 billion recently paid by uk broadcasters for the english premier league , the richest football league in the world , but it represents a tripling in value of the previous mls deal . according to phil rawlins , co-primary owner and president of the new mls franchise , orlando city soccer club , ” the industry and the game itself has moved on dramatically ” in the u.s. . he believes what would equal 50 years growth in most other industries has been experienced in the first two decades of the mls . rawlins ’ club is a prime example of this rapid transformation . he describes players being pushed out of changing facilities because of a schedule clash with a yoga class not so long ago . this weekend 60,000 fans are expected to witness orlando city ’s opening weekend fixture against new york city , another new club making their mls bow .</p>
<p>Reference Summary:</p> <p>the 20th mls season begins this weekend . league has changed dramatically since its inception in 1996 . some question whether rules regarding salary caps and transfers need to change .</p>
<p>BanditSum:</p> <p>-lrb- cnn -rrb- on the 6th of april 1996 , san jose clash and dc united strode out in front of 31,683 expectant fans at the spartan stadium in san jose , california . the historic occasion was the first ever major league soccer match – a brave new dawn for the world ’s favorite sport in a land its charms had yet to conquer . a further four are set to be added by 2020 .</p>
<p>BanditSum+KL:</p> <p>summarizing the action for espn , commentator ty keough eagerly described the momentous “ birth of a new era for american soccer . ” long gone is the desire to tamper with the rules of the game for a start . this weekend 60,000 fans are expected to witness orlando city ’s opening weekend fixture against new york city , another new club making their mls bow .</p>

Table 1: Example from validation set. While BanditSum opts for a similar summary as the lead-3 baseline, adding the mixed ROUGE loss attenuates the severity of this effect and allows it to choose more appropriate sentences occurring later in the article.

D_{early} and D_{late} Word Distribution

Dataset	Most Frequent Summary Words
D_{early}	labour , leader , ed , miliband , police, david , women, attack, manchester, nhs , family, brown , million, father, city
D_{late}	man , city , league , police, england, manchester , united , film, sterling , champions , scored , passengers, fire, liverpool , chelsea

Table 2: Top 15 most frequent words in the D_{early} and D_{late} subsets, with stop words removed. In D_{early} , we bold words seemingly related to UK politics, and in D_{late} , we do the same for UK soccer. Manchester United is a soccer team in the UK, Man City is an abbreviation for the same team, and Raheem Sterling is a player on this team.

Normalized Entropy on Validation Set

Model	Normalized Entropy
BanditSum	0.158
BanditSum+entropy	0.332
BanditSum+KL	0.440

Table 3: Normalized entropy scores on the validation set for BanditSum, BanditSum+entropy and BanditSum+KL. The entropy regularizer encourages the model to uniformly distribute its affinity scores across all sentences, whereas our KL-based ROUGE loss encourages a prediction distribution similar to sentence-level ROUGE scores. In this sense, the KL-based ROUGE loss acts as a ‘directed’ entropy regularizer, which encourages the model to converge towards a stochastic policy. This is beneficial for tasks containing inputs with a slightly different distribution from each other (Ahmed et al., 2019), such as summarization, where the distribution among distinct documents is different.

References

Zafarali Ahmed, Nicolas Le Roux, Mohammad Norouzi, and Dale Schuurmans. 2019. Understanding the impact of entropy in policy learning. *International Conference on Machine Learning (ICLM)*.