**Sum-QE: a BERT-based Summary Quality Estimation Model**

Stratos Xenouleas¹  
Marianna Apidianaki²  
Prodromos Malakasiotis¹  
Ion Androutsopoulos¹

---

**A**  
BlackBox Summarizer

Oh look! A super box that makes a summary out of many documents. I won’t have to read tons of articles any more!!!

Ah... you are right! What am I gonna do? I know what quality a summary must have. Can you show me?

There they are!

Q1: Grammaticality  
Q2: Non Redundancy  
Q3: Referential Clarity  
Q4: Focus  
Q5: Structure & Coherence

---

**B**  

(a) Single Task (S-1)  
(b) Multi-Task-1 (M-1)  
(c) Multi-Task-5 (M-5)

---

**C**  
I can see you are using multi-task learning. Why is that?

It will help me learn richer representations and make better predictions, especially when the qualities are highly correlated.

See in the heatmaps around us how qualities are correlated.

---

**D**  

And how well do you correlate with humans?

Pretty well actually! Come take a closer look at the table on the right:

---

**E**  

DUC-05  
DUC-06  
DUC-07

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.77 ± (± 0.42)</td>
<td>4.41 ± (± 0.20)</td>
<td>3.58 ± (± 0.60)</td>
<td>3.54 ± (± 0.78)</td>
<td>3.71 ± (± 0.31)</td>
</tr>
<tr>
<td>4.23 ± (± 0.26)</td>
<td>3.11 ± (± 0.52)</td>
<td>3.20 ± (± 0.66)</td>
<td>3.00 ± (± 0.47)</td>
<td>3.15 ± (± 0.41)</td>
</tr>
<tr>
<td>2.99 ± (± 0.50)</td>
<td>3.15 ± (± 0.41)</td>
<td>3.29 ± (± 0.46)</td>
<td>3.24 ± (± 0.59)</td>
<td></td>
</tr>
</tbody>
</table>

---

**F**  

I’m glad I could help! What are your plans now?

Well with your help I managed to predict linguistic quality.

I was wondering if I could learn how to predict content related aspects without human references.

---

See you around!