Supplemental Material for the Paper: A Principled Framework for Evaluating Summarizers: Comparing Models of Summary Quality against Human Judgments

Maxime Peyrard and **Judith Eckle-Kohler**
Research Training Group AIPHES and UKP Lab
Computer Science Department, Technische Universität Darmstadt
www.aiphes.tu-darmstadt.de, www.ukp.tu-darmstadt.de

A Supplemental Material

A.1 Proof of (θ, O) decomposition theorem

We propose here a rigorous proof of the (θ, O) decomposition theorem. We first repeat the notations and the theorem statement and then propose a proof.

Notation Let $D = \{s_i\}$ be a document collection considered as a set of sentences. A summary S is a subset of D, we note $S \in \mathcal{P}(D)$.

θ is an objective function defined in the paper by:

$$\begin{array}{ccccc} \theta & : & \mathcal{P}(D) & \rightarrow & \mathbb{R} \\ & & S & \mapsto & \theta(S) \end{array} \quad (1)$$

O is an operator which outputs a summary from a document collection D and a given θ:

$$\begin{array}{ccccc} O & : & \Theta \times \mathcal{D} & \rightarrow & \mathcal{S} \\ & & (\theta, D) & \mapsto & S^* \end{array} \quad (2)$$

Suppose c is the length constraint, then O produces S^* by solving the following optimization problem:

$$\begin{aligned} S^* &= \underset{S}{\operatorname{argmax}}\, \theta(S) \\ len(S) &= \sum_{s \in S} len(s) \leq c \end{aligned} \quad (3)$$

We define an extractive summarizer σ as a set function which takes a document collection $D \in \mathcal{D}$ and outputs a summary $S_{D,\sigma} \in \mathcal{P}(D)$.

$$\begin{array}{ccccc} \sigma & : & \mathcal{D} & \rightarrow & \mathcal{S} \\ & & D & \mapsto & S_{D,\sigma} \end{array} \quad (4)$$

Theorem The theorem states that for any summarizer σ there exists at least one tuple (θ, O) which is equivalent to σ:

Theorem 1 $\forall \sigma, \exists (\theta, O)$ *such that:*
$\forall D \in \mathcal{D}, \sigma(D) = O(\theta, D)$

Proof We can construct a function θ_σ from σ which reconstructs the exact same summaries as σ when optimized by O.

Suppose that $\sigma(D) = S_{D,\sigma}$. We define θ_σ to be the following function:

$$\theta_\sigma(S) = \begin{cases} 1, if S = S_{D,\sigma} \\ 0, otherwise \end{cases} \quad (5)$$

It is clear that $\forall D \in \mathcal{D} : \sigma(D = O(\theta_\sigma, D))$, because the optimal summaries according to θ_σ are the summaries produced by σ.

Going further At this point, the theorem is proved. While for every summarizer σ there exists at least one tuple (θ, O), in practice there exist multiple tuples, and the one proposed by the proof would not be useful to rank models of summary quality. We can formulate an algorithm which constructs θ from σ and which yields an ordering of candidate summaries.

Let $\sigma_{D \setminus \{s_1, \ldots, s_n\}}$ be the summarizer σ which still uses D as initial document collection, but which is not allowed to output sentences from $\{s_1, \ldots, s_n\}$ in the final summary.

For a given summary S to score, let $R_{\sigma,S}$ be the smallest set of sentences $\{s_1, \ldots, s_n\}$ that one has to remove from D such that $\sigma_{D \setminus R}$ outputs S. Then the definition of θ_σ follows:

$$\theta_\sigma(S) = \frac{1}{R_{\sigma,S} + 1} \quad (6)$$

Therefore, if S is the summary outputed by σ without modifying anything, then $\theta_\sigma(S) = 1$ is the highest possible score. The scores are decreasing for summaries which need more sentences to be removed. Indeed, these summaries have low scores according to σ and should also have low scores according to θ_σ.