

Supplemental Material for the Paper: Optimizing an Approximation of ROUGE – a Problem-Reduction Approach to Extractive Multi-Document Summarization

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

1 Submodularity of $\tilde{\rho}$

Let $S = \{s_i | i \leq m\}$ be a set of sentences and $\tilde{\rho}(S)$ its ROUGE-N approximation:

$$\tilde{\rho}(S) = \sum_{i=1}^n \rho(s_i) - \sum_{a,b \in S, a \neq b} \tilde{\epsilon}(a \cap b) \quad (1)$$

To prove the submodularity of $\tilde{\rho}$, we prove that $\tilde{\rho}$ follows the diminishing returns property: $\forall S \subseteq T$ and a sentence a : $\tilde{\rho}(S \cup a) - \tilde{\rho}(S) \geq \tilde{\rho}(T \cup a) - \tilde{\rho}(T)$

Proof: Let $S = \{s_i | i \leq m\}$ and $T = \{t_i | i \leq l\}$ be two sets of sentences such that $S \subseteq T$. We study the following difference:

$$\tilde{\rho}(S \cup a) - \tilde{\rho}(S) - (\tilde{\rho}(T \cup a) - \tilde{\rho}(T)) \quad (2)$$

We recall the formula for any set X and any sentence a :

$$\tilde{\rho}(X \cup a) = \tilde{\rho}(X) + \tilde{\rho}(a) - \tilde{\epsilon}(X \cap a) \quad (3)$$

When applied to equation (2) we obtain:

$$\tilde{\epsilon}(T \cap a) - \tilde{\epsilon}(S \cap a) \quad (4)$$

We recall the definition of $\tilde{\epsilon}(X)$ for any set X :

$$\tilde{\epsilon}(X) = \frac{1}{R_N} \sum_{g \in X} \mathbb{1}[freq(g) \geq \alpha] \quad (5)$$

$S \subseteq T$ implies $S \cap a \subseteq T \cap a$. Therefore we can split $\tilde{\epsilon}(T \cap a)$ and the expression (4) becomes:

$$\begin{aligned} & \frac{1}{R_N} \sum_{g \in S \cap a} \mathbb{1}[freq(g) \geq \alpha] \\ & + \frac{1}{R_N} \sum_{g \in T \cap a \setminus S \cap a} \mathbb{1}[freq(g) \geq \alpha] \\ & - \frac{1}{R_N} \sum_{g \in S \cap a} \mathbb{1}[freq(g) \geq \alpha] \quad (6) \end{aligned}$$

Which can be simplified to:

$$\frac{1}{R_N} \sum_{g \in T \cap a \setminus S \cap a} \mathbb{1}[freq(g) \geq \alpha] \quad (7)$$

This is a sum of positive terms and it is therefore positive. This proves that expression (2) is ≥ 0 which is equivalent to:

$$\tilde{\rho}(S \cup a) - \tilde{\rho}(S) \geq \tilde{\rho}(T \cup a) - \tilde{\rho}(T) \quad (8)$$

Which concludes the proof of the submodularity of $\tilde{\rho}$.

2 System Output Samples

Topic d081a (DUC-2002): More than 4,000 workers at a coal mine in the southern city of Jastrzebie went on strike today to demand legalization of Solidarity and higher wages, a spokeswoman said. Andrzej Andrzejczak, head of the Morcinek mine strike committee, said: “We support all of the demands of the July Manifesto mine workers.” Romanians appear solidly opposed to the strike, angered by the miners’ demands for an average wage of the equivalent of Dollars 290 a month. Strikes were also reported in the Pechora Coal Basin in the Russian republic, and Chervonograd in the western Ukraine, official media said. By evening, 14 mines in the region, home of the country’s third-largest coal fields, were reported at a standstill. Some of the miners at the Karl Marx, “Chernovonny Zhovten”, “Chervonny Profinter”, and “Yunkom” pits joined the strike. The strike in Jastrzebie, about 275 miles southwest of Warsaw, comes on the eighth anniversary of the founding of the outlawed trade federation. Most of the country’s 1.2 million coal miners are still working, but output in some key regions has been halved, saddling Gorbachev with a substantial labor and fuel problem even as he struggles to revive the Soviet economy.

Topic d30010t (DUC-2003): The Islamic militant group Hamas, which has been trying to sabotage the peace accord, claimed responsibility for the 9:45 a.m. (0745 GMT) attack in the Mahane Yehuda market, police said. The blast killed two assailants, wounded 21 Israelis and prompted Israel to suspend implementation of the peace accord with the Palestinians. Palestinian officials condemned the attack, but accused Prime Minister Benjamin Netanyahu of political blackmail for holding up the agreement that was to give the Palestinians another 13 percent of the West Bank. In response, Israel's Cabinet suspended implementation of the new peace accord with the Palestinians.

Topic d30042t (DUC-2003): LONDON (AP) Libyan leader Moammar Gadhafi said the suspects in the Lockerbie bombing are "very happy" to be tried in the Netherlands, and he hoped the trial would lead to a better relationship with the United States. African countries voted in June to ignore the U.N. flight ban that was imposed in 1992 to try and force Libya to hand over for trial two suspects wanted in the 1988 bombing of an American airliner over Lockerbie, Scotland. Libya has been under U.N. sanctions since 1992 to force it to hand over the two defendants in the Lockerbie case.