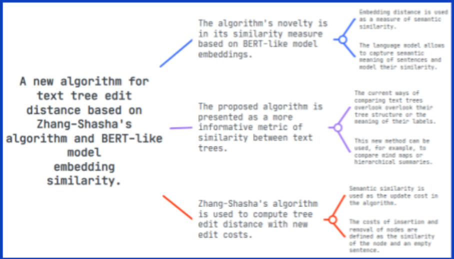
APEIE 25
Actual Problems of Electronic Instrument Engineering

IEEE

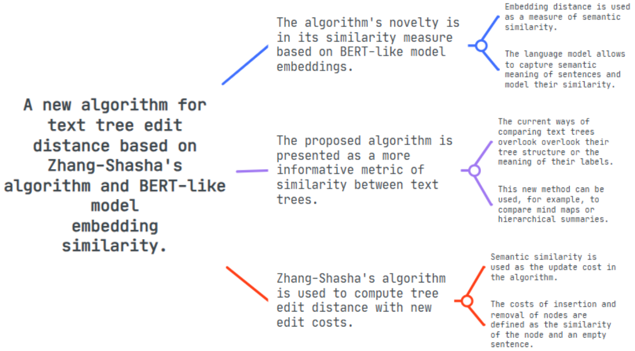
НГТУ НЭТИ

Text Tree Edit Distance: Comparing Text Hierarchies Using Language Models



Fedor Sobolevsky, MIPT, graduate student
Konstantin Vorontsov, MIPT, professor, Ph.D.

Research Motivation



An example of a text tree — a hierarchical summary of this study in the form of a *mind map*

Problem: How to compare hierarchical summaries, considering both their structure and semantics?

Hierarchical Summarization Problem Statement

Let $\mathcal{S}$ be the *set of texts* over a given vocabulary.
Text tree — a tree $T = (V, E)$, where $E \subset V^2$ and for each $v \in V$ a text $s(v) \in \mathcal{S}$ is defined.
$\mathcal{T}$ — the considered *set of text trees*.

Task: Find a mapping $f : D \mapsto T$ that constructs a hierarchical summary $T \in \mathcal{T}$ from a document D, minimally different from a reference summary T^* of D constructed by an expert:

$$\rho(f(D), T^*) \longrightarrow \min_f .$$

Question: How do we choose the metric $\rho : \mathcal{T}^2 \to \mathbb{R}_+$?

Proposed Metric — *TTED*

TTED *(text tree edit distance)* — tree edit distance[1], where the cost of edit operations is:
a) *replacement* of vertex v with v': $r(s(v), s(v'))$;
b) *addition/removal* of vertex v: $r(s(v), \lambda)$;
where λ — empty string.

Semantic distance r can be measured as the distance between *embeddings* (vector representations) of texts, obtained using a language model $\mathrm{LM} : \mathcal{S} \to \mathbb{R}^n$:

$$\forall s, s' \in \mathcal{S} \quad r(s, s') = \rho_n(\mathrm{LM}(s), \mathrm{LM}(s')),$$

where ρ_n — a metric in $\mathbb{R}^n$.

[1] *Zhang Kaizhong, Statman Richard, Shasha Dennis.* On the Editing Distance Between Unordered Labeled Trees (1992)

Baseline Text Tree Comparison Method

In the study of *Zhang et al., 2024*[2] the similarity of text trees $T = (V, E)$ and $T' = (V', E')$ is defined as

$$\mathrm{Sim}(T, T') = \max_{P \subset E \times E'} \sum_{(e, e') \in P} \sum_{i=0,1} \mathrm{ROUGE}(e_i, e'_i).$$

where P — a one-to-one mapping of edges of T to edges of T' (the optimal one is found by a greedy algorithm), $\mathrm{ROUGE}(v, v')$ — the averaged ROUGE-1, ROUGE-2, and ROUGE-L similarity score of $s(v)$ and $s(v')$.

For consistency, the distance measure used is

$$\rho(T, T') = \sqrt{\mathrm{Sim}(T, T) + \mathrm{Sim}(T', T') - \mathrm{Sim}(T, T') - \mathrm{Sim}(T', T)}.$$

[2] *Zhang Zhuowei, Hu Mengting, Bai Yinhao, and Zhang Zhen.* Coreference Graph Guidance for Mind-Map Generation (2024)

Aspects of Text Tree Difference

Let for $T \in \mathcal{T}$ the following sets of trees be defined:

1. $P(T)$ — trees differing from T only in paraphrasing;
2. $S(T)$ — trees differing from T only in structure;
3. $M(T)$ — trees differing from T only in semantics (in meaning/content).

Idea: for an adequate metric ρ on $\mathcal{T}$ it should hold that

$$\langle \rho(T, T') \rangle_{T' \in P(T)} \ll \langle \rho(T, T'') \rangle_{T'' \in S(T)},$$

$$\langle \rho(T, T') \rangle_{T' \in P(T)} \ll \langle \rho(T, T''') \rangle_{T''' \in M(T)}.$$

Metric Quality Criteria

Consider a sample $\mathcal{D} = \{T, T'_1, \ldots, T'_p, T''_1, \ldots, T''_s, T'''_1, \ldots, T'''_m\}$, where $T \in \mathcal{T}$, $T'_i \in P(T)$, $T''_j \in S(T)$, $T'''_k \in M(T)$.
Quality coefficients of metric ρ on sample $\mathcal{D}$:

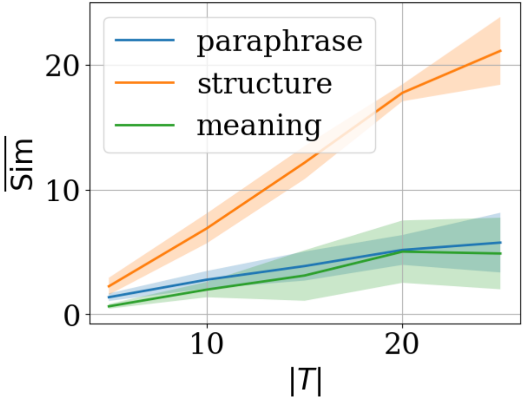
$$R_S^{\mathcal{D}}(\rho) = \frac{1}{sp}\sum_{i=1}^{p}\sum_{j=1}^{s}\frac{\rho(T, T'_i)}{\rho(T, T''_j)}, \quad R_M^{\mathcal{D}}(\rho) = \frac{1}{mp}\sum_{i=1}^{p}\sum_{k=1}^{m}\frac{\rho(T, T'_i)}{\rho(T, T'''_k)}.$$

$R_S^{\mathcal{D}}(\rho)$ — sensitivity of metric ρ to **paraphrasing** relative to **structure**;
$R_M^{\mathcal{D}}(\rho)$ — sensitivity to **paraphrasing** relative to **semantics**.
Optimization problem:

$$R_S^{\mathcal{D}}(\rho) \longrightarrow \min_{\rho}, \quad R_M^{\mathcal{D}}(\rho) \longrightarrow \min_{\rho}.$$

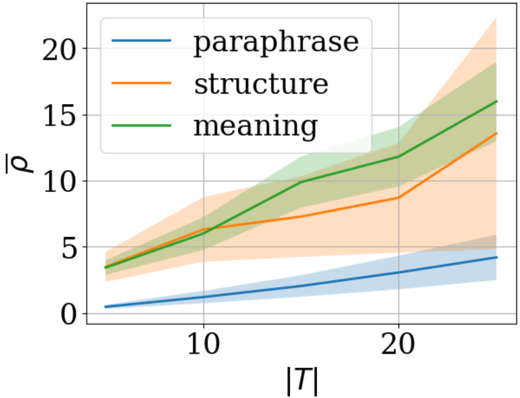
Method Testing — Results

a) Average values of the baseline similarity coefficient



$\mathrm{Sim}(\cdot,\cdot)$ reflects differences in semantics and paraphrasing similarly, noticeably less so for structure.

b) Average values of the TTED distance



TTED reflects differences in paraphrasing noticeably less than in structure and semantics, reflected similarly.

Metric Testing — Results

Results of tests for the baseline distance measure and TTED with different encoder models for text embedding generation on synthetic data

Method	$R_S^{\mathcal{D}}(\rho)$	$R_M^{\mathcal{D}}(\rho)$
Baseline	2.05±0.79	0.96±0.10
TTED with DistilRoBERTa	0.58±0.22	0.53±0.11
TTED with SPECTER	0.69±0.35	0.46±0.14
TTED with MPNet	**0.44±0.12**	0.48±0.11
TTED with fine-tuned MPNet	0.61±0.78	**0.45±0.12**

Significant differences compared to insignificant ones are reflected better by TTED than by the baseline method.

Testing TTED Modifications

Dependence of quality coefficients on the **metric for comparing embeddings** in TTED

$r(x, y)$	$\overline{\rho}_1$, $\|T\| = 10$	$\overline{\rho}_3$, $\|T\| = 10$	$R_M^{\mathcal{D}}(\rho)$
$\sqrt{1 - S_C(x, y)}$	3.89±0.71	8.41±0.80	**0.48±0.11**
$\|\|x - y\|\|_2$	5.50±1,00	11.89±1.14	**0.48±0.11**
$\|\|x - y\|\|_1$	119.70±21.60	259.05±25.12	**0.48±0.11**

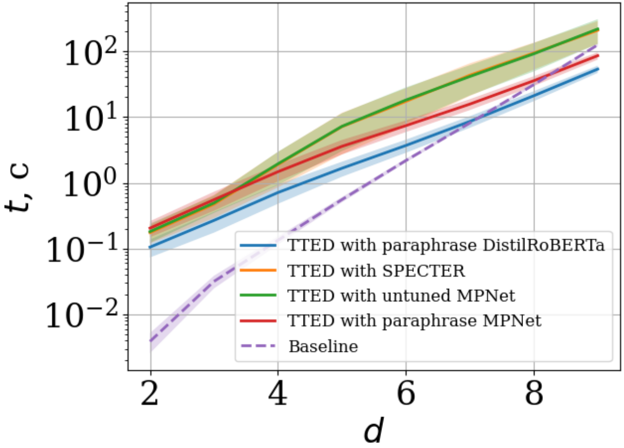
Dependence of quality coefficients on the use of **context** in TTED

Method	$R_S^{\mathcal{D}}(\rho)$	$R_M^{\mathcal{D}}(\rho)$
Without context	0.44±0.12	0.48±0.11
With context	**0.43±0.19**	**0.35±0.08**

The optimal and most interpretable TTED configuration is the one with the fine-tuned MPNet encoder, distance based on cosine similarity, and using texts from parent vertices as context.

Computation Time

Average computation times for different distances between full binary text trees of various depth d



Main Results

1. A new metric on the set of text trees — TTED — has been proposed.
2. It is shown that TTED reflects significant differences between text trees better than the previously used similarity coefficient.
3. An optimal configuration for TTED has been selected.
4. The proposed metric can be used to assess quality in tasks of hierarchical summarization, mind map construction, and other tasks of automatic generation of text hierarchies.

References

- *Zhang Z., Hu M., et al.* Coreference Graph Guidance for Mind-Map Generation // Proceedings of the AAAI Conference on Artificial Intelligence. — 2024. — Vol. 38. — P. 19623–19631.
- *Zhang K., Statman R., Shasha D.* On the editing distance between unordered labeled trees. // Information processing letters. 1992 May 25; 42(3): 133-9.
- *Vrbanec T., Meštrović A.* Comparison study of unsupervised paraphrase detection: Deep learning — The key for semantic similarity detection. // Expert systems. 2023 Nov; 40(9): e13386.