

The paper “Label-Guided Relation Prototype Generation for Continual Relation Extraction” by Shuang Liu et al. is devoted to a new method for Continual Relation Extraction.

Proposed method extracts semantic information from the label namespace, resulting in adequate descriptions of relation semantics by relation prototypes. As follows from the abstract, the method allows one to relatively well overcome the problem of catastrophic forgetting. According to the authors, the advantage of the method is also the fact that extracting additional information from data is not limited by memory capacity. The stated general description of the algorithm in the Abstract sounds very interesting.

The logic of presentation suggests the approximately following structure for such an article: first (1) a description of existing methods, then (2) the theory used, (3) the developed algorithm, (4) experiment, (5) analysis of the results of applying the method, (6) comparison of the results of the method with the results of the methods discussed at the beginning, (7) conclusion. This plan makes it possible to explain the material step by step. Formally, the authors adhered to almost this scheme. Let's look at all the parts.

The first step ((1) a description of existing methods) in the paper actually presents in parts Introduction and Related Work (“Work” is written in the singular for some reason).

Then the second and third logical steps ((2) the theory and (3) the developed algorithm) are mixed in one part Methodology. As a result, the description of the developed algorithm is simply lost in the description of the theory. It is proposed to follow the plan and first describe the theoretical part, and then the algorithm of the method (in particular describe all the symbols in Figure 1, which, as I understand it, depicts the algorithm).

The next three logical steps ((4) experiment, (5) analysis of the results of applying the method and (6) comparison of the results of the method with the results of the methods discussed at the beginning) are presented in one part Experiment with references to Table I, Table II, Figure 2, Figure 3 and Figure 4. The result of mixing several topics in this part is compounded by the level of evidence. The authors make a strong statement that “method has achieved competitive performance compared to the state-of-the-art baseline models in CRE”, but in Table I does not even average the data across all Tasks, which will make it possible to evaluate of the effectiveness of the developed method. The Ablation Study section describes the results of the experiments without explaining which part of the algorithm the abbreviations used apply to (LKI on lines 232, 233 and MSD on lines 236, 239). The Memory Size Analysis section describes the memory size in unnamed units (line 260). In the section Effectiveness of Multi-Loss Distillation the results are explained based on two scatterplots (on Figure 4) without a clear description. And, unfortunately, there is no analysis of how the method allows to overcome the problem of catastrophic forgetting, as indicated in the Abstract. It is proposed to rewrite the part according to the plan, increase the level of evidence taking into account comments and make corrections to the Supplement.

Minor comments:

1) In formulas:

- in the formula (1) it is unclear what $W_1[h^{11}; h^{21}]$ is, it does not explain what d is in R^d (line 125).
- in the formula (2) there is W_2 , but after the formula W_1 is described (line 128) and h is not described.
- after the formula (3) the description of $P(r|x_i; \theta_k)$ is unclear (line 130).
- in the formula (5) the undescribed operator $Att.$ and d_1 are used, linear transformation of H_{x_r} isn't described.
- after the formula (7) $z_{x_i}, z_{r_i}, z_r, \tau_1$ are not described.
- in the formula (9) the variable S_{im} should probably have other indices - S_{ij} .
- in the formulas (10) and (11) it is also necessary to check the indices.
- after the formulas (12) and (13) the variables α, β, λ are not described.
- in the formula (15) L_{rd} is used, but in the formula (8) L_{nd} is used. Are these different variables?

2) Why make Continual Learning section in the Related Work part if there is only one?

3) Unclear sense of sentences on lines 130, 131, 151, 169.

4) The text contains more than 40 obvious spelling errors, even the title of the article contains an error.

5) Captions on Figures and Tables are unclear. The scales in the scatterplots in Figure 2 are different, as are those in Figure 4.

Overall, the article presents a potentially interesting method. However, the structure of the paper and some presented results are somewhat confusing and require either more explanation or corrections.