Junpeng LUO, Jingxuan ZHANG, Zhiqiu HUANG, Yong XU, Chenxing SUN, 2022. Toward an accurate method renaming approach via structural and lexical analyses. *Frontiers of Information Technology & Electronic Engineering*, 23(5):732-748. https://doi.org/10.1631/FITEE.2100470

Toward an accurate method renaming approach via structural and lexical analyses

Key words: Method renaming; Code refactor; Deep learning; Convolutional neural networks

Corresponding author: Jingxuan ZHANG
E-mail: jxzhang@nuaa.edu.cn
ORCID: https://orcid.org/0000-0002-8437-6640

Challenges and motivation

Challenges

1. Enforcing naming convention rules is inefficient. There is a lack of thinking about the composition of method name structure, and the programming context is not fully used.

2. Short names and prefixes are not specific enough and may result in ambiguous results.

3. It is not easy to find similar or even the same methods in large corpora, because there may be a lot of noises.

Motivation

1. Method names should consist of verbs or gerund. Verbs represent the functional behavior of methods, and nouns represent the functional objects of methods.

2. Similar methods enjoy similar method bodies.

3. Much information can be used directly in the programming context .

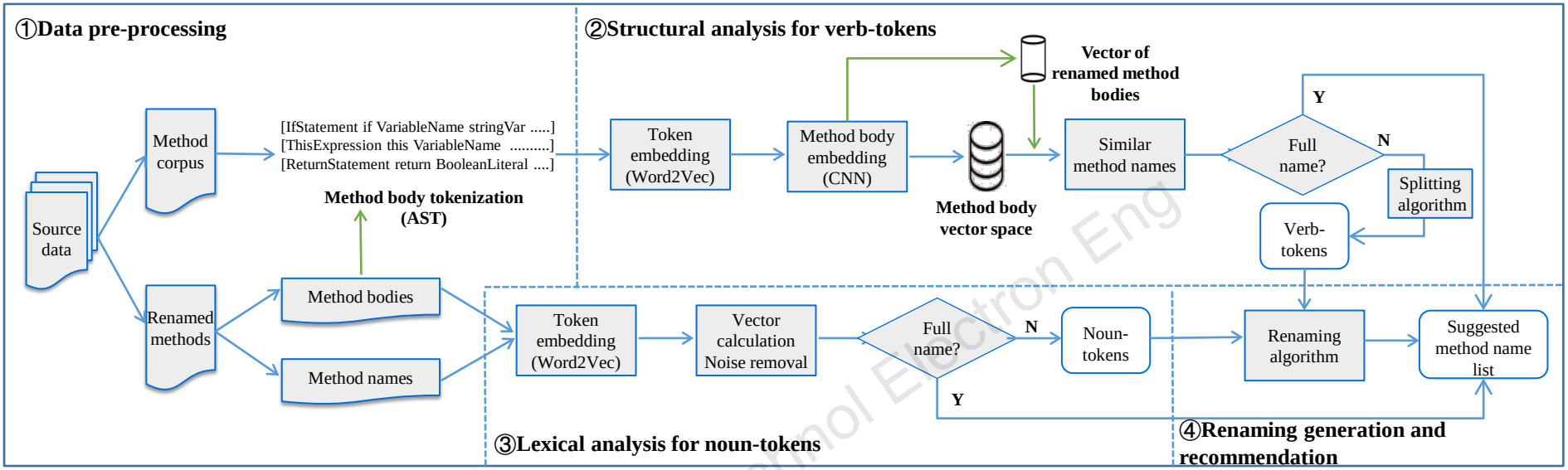
Main idea

1. We propose a novel method renaming approach that leverages structural analysis and lexical analysis to suggest high-quality method names.

2. We fully leverage method name structure and lexical analysis with programming context information to discover verb-tokens and noun-tokens to generate high-quality method names.

3. We have conducted a series of experiments to validate the effectiveness of our approach.

Method



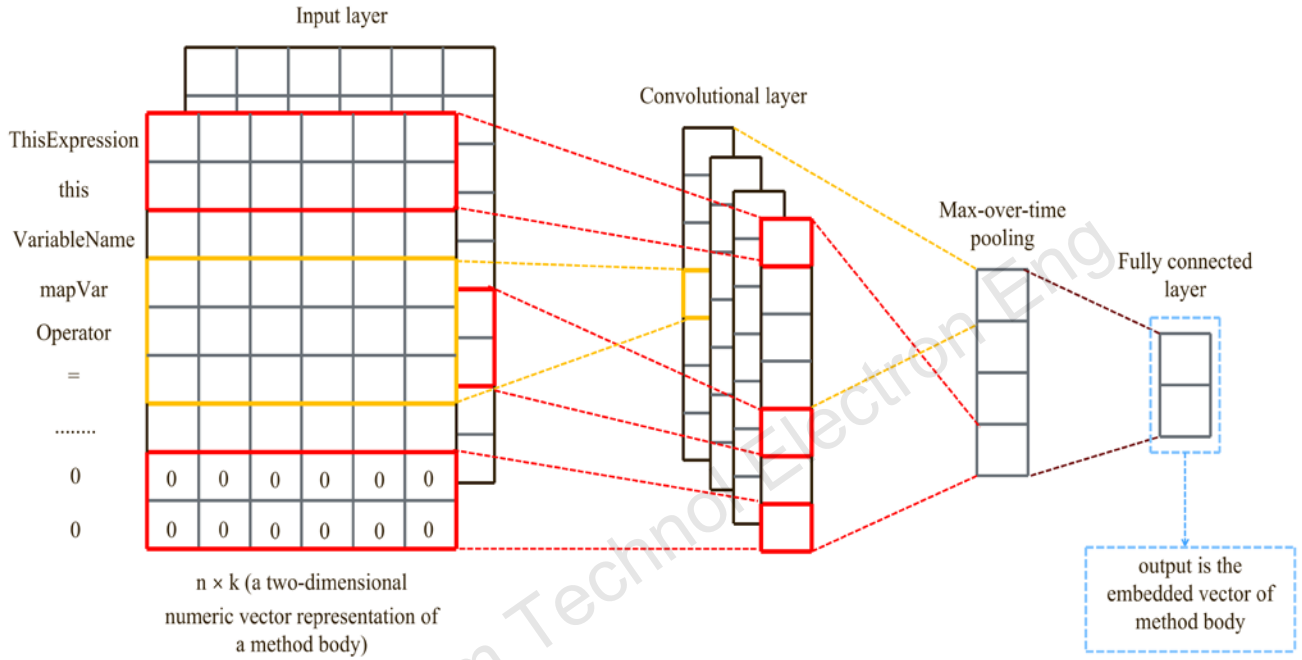
1. Data preprocessing: downloading, analyzing, and preprocessing experimental data

2. Structural analysis for verb-tokens: obtaining the verb-tokens between the renamed method and the method corpus by machine learning

3. Lexical analysis for noun-tokens: obtaining the noun-tokens that make up the method name in the programming context

4. Renaming generation and recommendation: generating high-quality method names and recommending them to developers

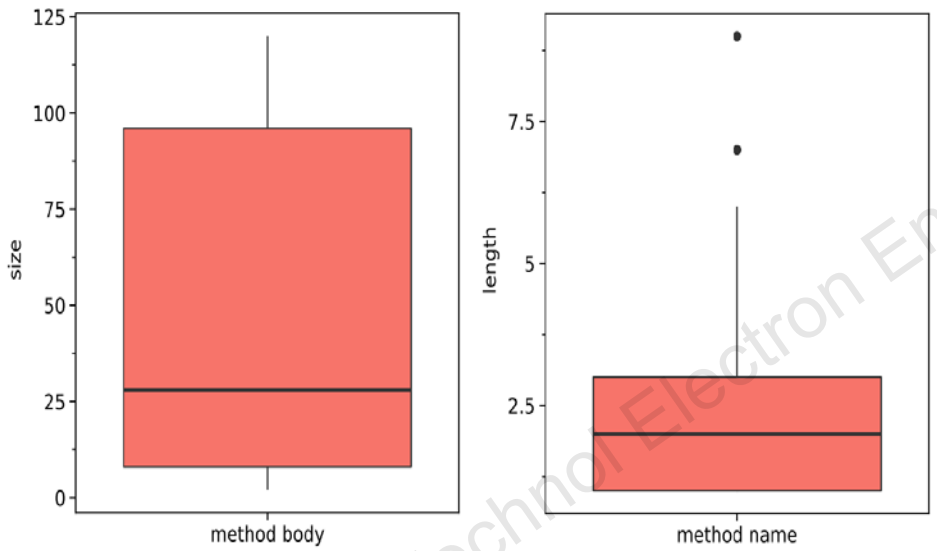
Method



Architecture of the convolutional neural network (CNN)

Our approach uses the CNN architecture as the embedding function. The input is two-dimensional numerical vectors of method bodies.The convolutional layer and max-over-time pooling layer are used to capture the local features of methods and reduce the dimension of the input data. The network layers from the pooling layer to the fully connected layer are fully connected, and can combine all local features captured by the convolutional and pooling layers. We choose the output of the fully connected layer as vector representations of method bodies that synthesize all local features captured by the previous layers.

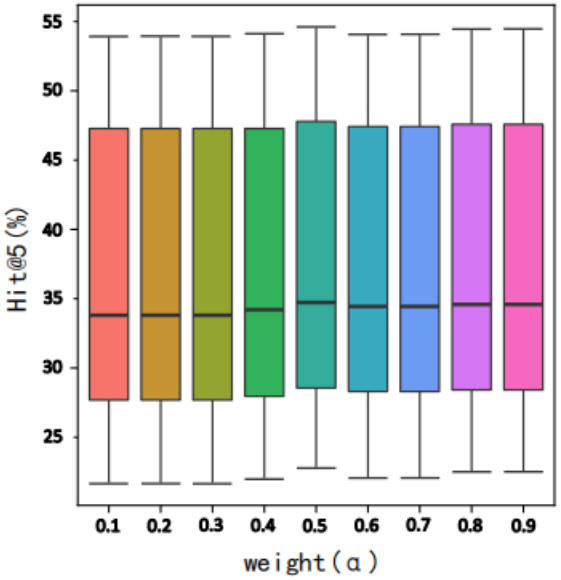
Method



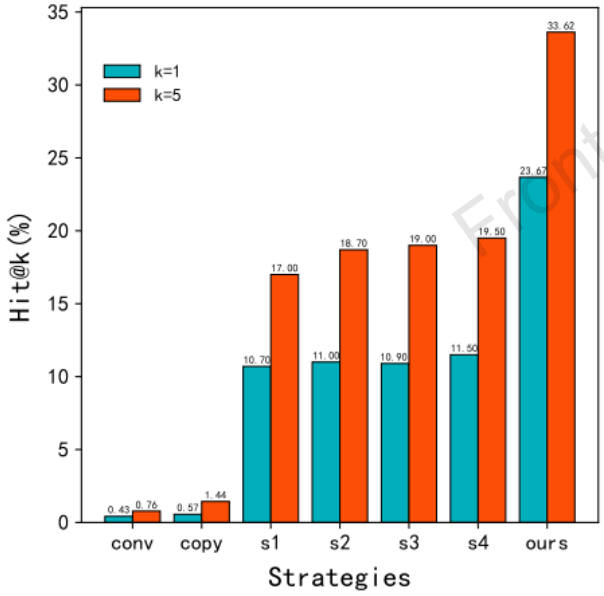
The size distribution of method bodies and the length of method names

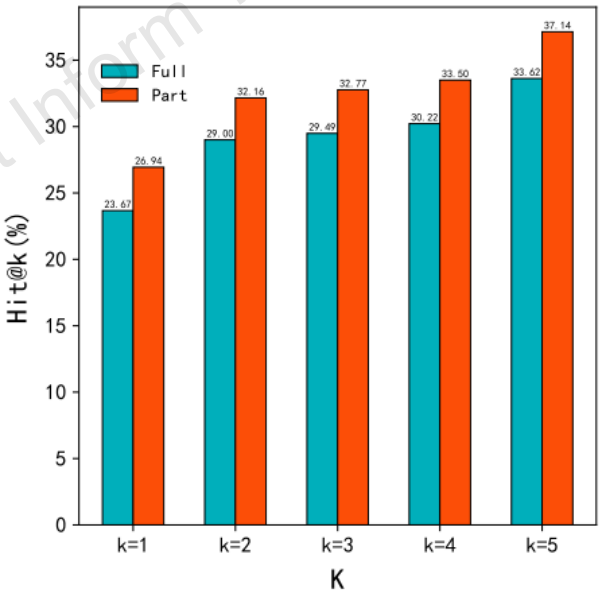
To avoid reduction in the intelligibility of methods and explosion of code tokens, we further limit the number of constitutive terms of method names (method name length) and the number of tokens in the method body (method body size). For each remaining method, we calculate the length of its name and the size of its body. The above figure shows the boxplot of the size distribution of method bodies and the method name length for all the remaining methods.

Major results



Our approach is insensitive to the weight *α*. When it is set to 0.5, our approach achieves the best results.





Our approach is superior to the baseline approaches in terms of the Hit Ratio.

Major results

Score	Portablity		Consistency	
	Number	Percentage	Number	Percentage
0	9	4.26%	0	0.00%
1	3	1.42%	4	1.89%
2	3	1.42%	3	1.42%
3	7	3.32%	4	1.89%
4	2	0.95%	2	0.95%
5	25	11.85%	7	3.32%
6	23	10.90%	22	10.43%
7	9	4.27%	20	9.48%
8	17	8.06%	14	6.04%
9	5	2.37%	23	10.90%
10	108	51.18%	112	53.08%

Most of the portability and consistency scores of the suggested method names are relatively high, which means that the suggested method names can benefit developers.

Conclusions

1. It is difficult to generate a specific and accurate method name for renaming methods and recommend it to developers.

2. We have proposed a novel approach to suggest effective names for methods using structural analysis and lexical analysis.

3. We have obtained verb-tokens and noun-tokens through structural and lexical analyses to generate high-quality method names and recommend them to developers. By analyzing the structure of method names and making full use of programming context, a novel identifier renaming method is proposed from another perspective.

4. The Hit@5 is increased by 14.12% compared with the state-of-the-art approaches.

Future work

1. Validate our approach in more corpora to generalize its performance.

2. Extend our approach to recommend new names for all the identifier categories, including methods, types, and fields.

3. Build an automatic tool that encapsulates our approach to help developers rename methods.

Junpeng LUO is currently pursuing his MS degree at Nanjing University of Aeronautics and Astronautics (NUAA), China. His research interests include mining software repositories and software data analytics.

Jingxuan ZhANG received his PhD degree in software engineering from Dalian University of Technology, China. He is an associate professor with College of Computer Science and Technology, NUAA. His research interests include mining software repositories and software data analytics.

Zhiqiu HUANG received his PhD degree in computer science from NUAA in 1999. He is a professor with College of Computer Science and Technology, NUAA. He has authored more than 80 papers in refereed journals and proceedings in the areas of software engineering and knowledge engineering. His research interests include software engineering, formal methods, and knowledge engineering.