

Automatically Filtering Stable Bug Fixing Patches in Linux Kernel

ABSTRACT

ACM Reference format:

. 2017. Automatically Filtering Stable Bug Fixing Patches in Linux Kernel. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17)*, 2 pages.
<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

1 INTRODUCTION

2 RELATED WORK

3 APPROACH

3.1 Formalization

Let S be a set of stable bug fixing patches and P a set of all bug fixing patches in Linux kernel. We call $U \in (S \cap P)$ is a set of unknown bug fixing. In our problem, we try to provide a bug fixing patches ranking $\subset P^2$ for a developer, where each bug fixing patch has to meet the following properties [2]:

- *Totality*: $\forall i, j \in P : i \neq j \Rightarrow (i > j) \vee (j > i)$
- *Antisymmetry*: $\forall i, j \in P : (i > j) \wedge (j > i) \Rightarrow i = j$
- *Transitivity*: $\forall i, j, k \in P : (i > j) \vee (j > k) \Rightarrow i > k$

3.2 Bug Fixing Patch Ranking

The architecture of bug fixing patch ranking is represented in Fig. 1. Our model based on convolutional neural network [1] to map bug fixing patches to vectors, which can be used to compute their ranking score. In the following, we describe how the model computes the ranking score given by a pair of bug fixing patches. We briefly explain each component (i.e., hidden layer, logistic function, etc.) of our model.

3.3 Learning Relationship Between Text and Code in Single Patch

In Linux kernel, each patch contains both a textual commit message and commit code. The commit message, which can help a developer to speed up the reviewing process, is a short description of source code change. The commit code is the changes that are applied on the buggy file.

3.4 Hidden Layer

3.5 Feature Representation

3.6 Output Layer

3.7 Training

Our model is trained by minimizing the cross-entropy lost function:

$$\begin{aligned}\mathcal{L} &= -\log \prod_{i \in S, j \in U}^N p(y_{ij}|p_i, p_j) + \lambda \|\theta\|_2^2 \\ &= -\sum_{i \in S, j \in U}^N \log \sigma(\mathbf{a}_{ij}) + \lambda \|\theta\|_2^2\end{aligned}\quad (1)$$

where S and U are the stable and unknown bug fixing patches datasets respectively, $y_{ij} = 1$ since p_i has higher rank compared to p_j , N is the total number pairs of bug fixing patches in S and U , $\sigma(\cdot)$ represents the logistic function to identify the rank order of patches pair. \mathbf{a}_{ij} is the output of our ranking model and can be decomposed as:

$$\mathbf{a}_{ij} = \mathbf{a}_i - \mathbf{a}_j \quad (2)$$

to satisfy the properties mentioned in Sec. 3.1.

3.8 Regularization

REFERENCES

- [1] Yann LeCun, Yoshua Bengio, et al. 1995. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361, 10 (1995), 1995.
- [2] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI Press, 452–461.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2017 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

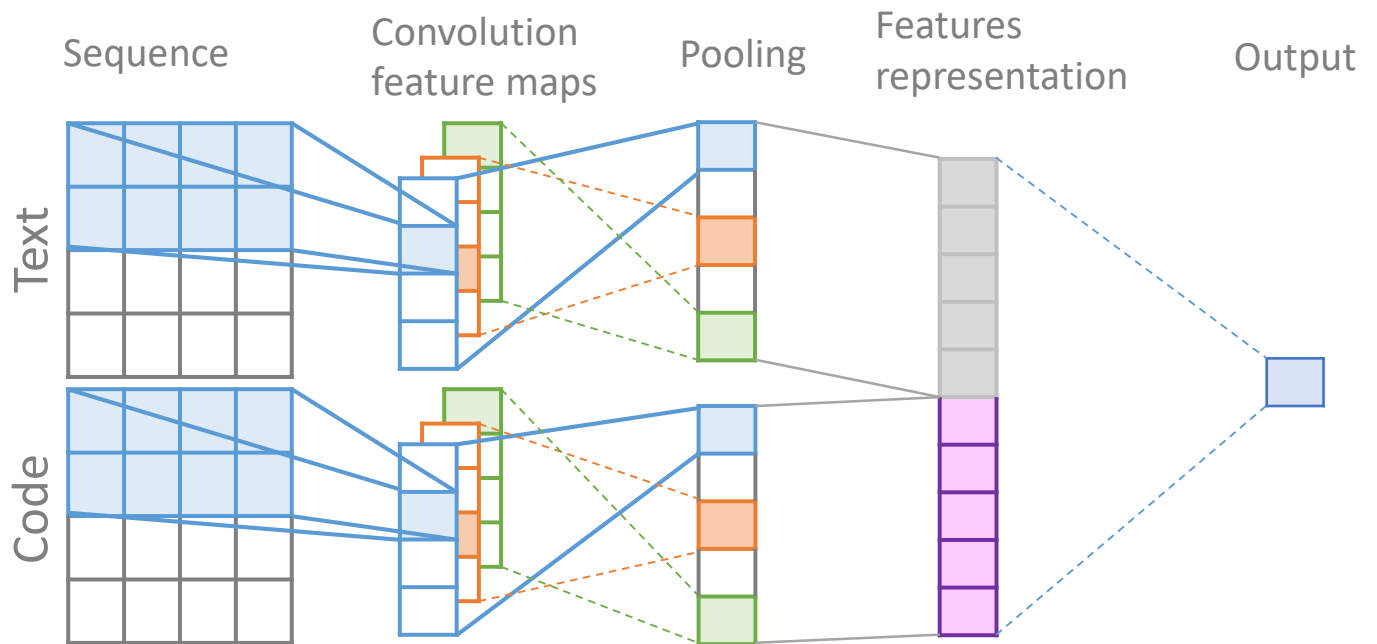


Figure 1: An architecture of bug fixing patches ranking model in Linux kernel.