Word2Vec-BERT-bmu:Classification of RISC-V Architecture Software Package Build Failures

Shitian Ma, Hui Li*, Jiaxin Zhu, Xiaohui He, Shuyang Zhang, Junfeng Zeng

The 28th International Conference on Engineering of Complex Computer Systems (ICECCS 2024)

联系方式：李慧 15210848139 lihui2012@otcaix.iscas.ac.cn

Background

◆ A large number of packages have to be rebuilt for RISC-V architecture.

◆ OBS, hosts **8,749** projects, with **137,659** packages, in **9,970** repositories and is used by over **4,000** confirmed developers.

Add the missing dependencies in spec files.

Challenge

◆ Building failures often cause software to **fail** to be successfully ported to RISC-V.

◆ A large number of building failures cannot be **accurately fixed** in a short time.

◆ The cause of package building failures is **complex** and diverse.

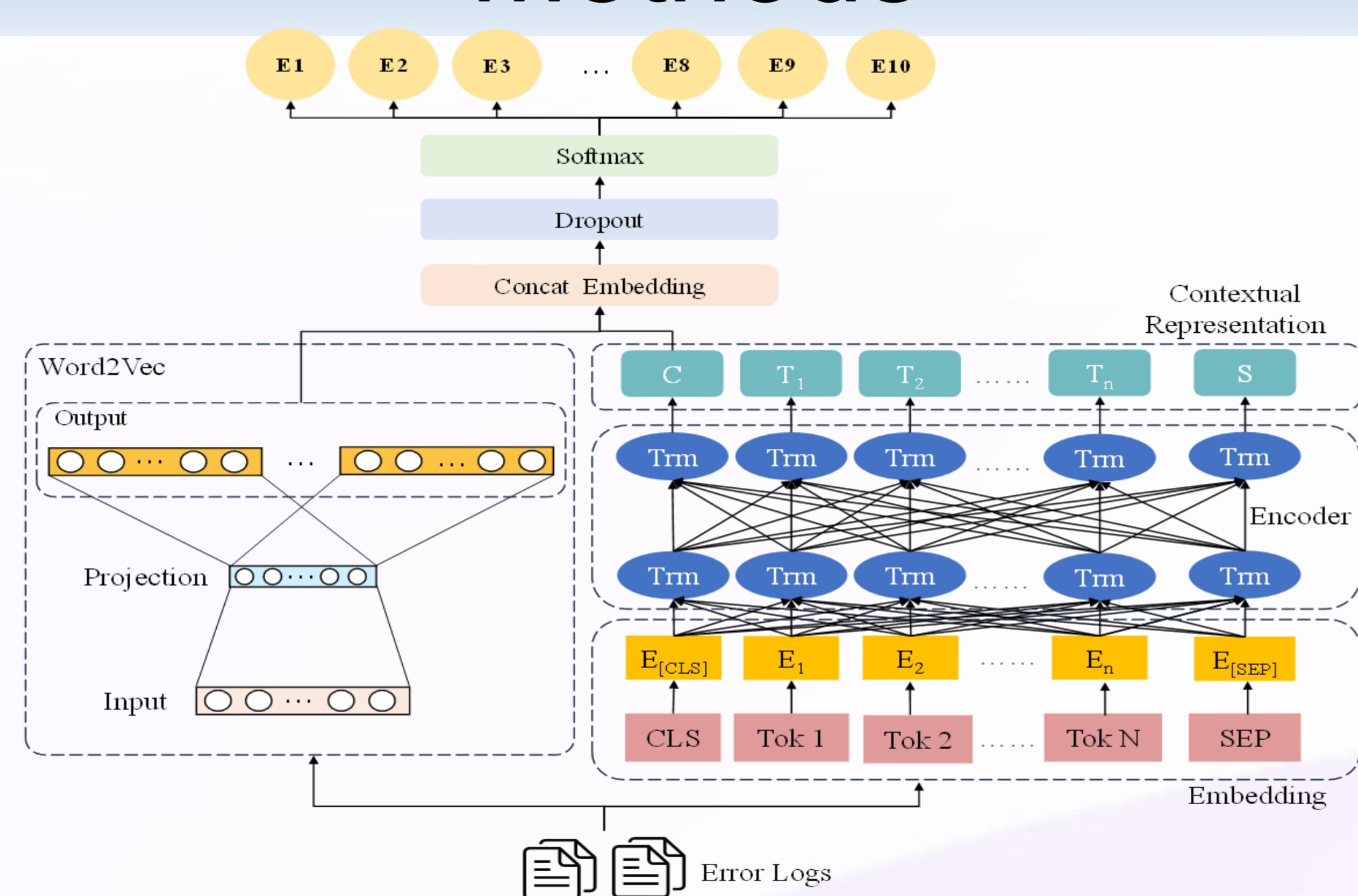◆ Knowledge ability among developers is **different**.

Motivation

RQ1: What causes RISC-V package building failures?

RQ2: How to effectively classify RISC-V package building failures using deep learning method ?

Methods

Results

Failure patterns		Examples	Numbers
Build configuration problem (E1)	Loss of dependency	dom4j	485
	Dependent versions do not match	dnf	
Test case problem (E2)	Test case execution failed	apache-commons-pool2	164
Incompatibility problem(E3)	The compiler is incompatible	ghc-pandoc-lua-marshal	113
	Architecture incompatibility	boxfort	
	Function update	monero-gui	
	File conflicts	dconf	
Source code problem(E4)	Variables/functions, etc., undefined	coturn	89
	Function misuse	janus-gateway	
	Code spelling/grammar failures	matrix-quaternion	
Files/directories are missing problem(E5)	The build process could not find the relevant file or directory	kubernetes1.21	72
Memory problem(E6)	Memory overflow	amsynth	35
Time-out problem (E7)	The code execution time exceeds the threshold. Procedure	atlas	27
Plug-in problem(E8)	Plugin cannot compile	hadoop	15
Network problem(E9)	Unable to download dependencies	hive	20
Other(E10)	None of the above categories can be classified	file-roller	37
		total	1057

Three unknown packages has been fixed.

ID	Package Name	Failure Patterns	Fix Strategy
1	Boxfort	Incompatibility problem(E3)	Add support for RISC-V architecture
2	Nomacs	Build configuration problem(E1)	Add missing dependencies in spec files
3	Libjxl	Source code problem(E4)	Import relevant definitions/headers

Conclusion

◆ **Ten** types packages building failures are summarized.

◆ A high-quality RISC-V building failure log dataset are available.

https://github.com/mstsky115/Build_error_logs_data

◆ Word2Vec-BERT-bmu, a effective deep learning method is provided to classify the packages building failures.

https://log-analyze.rvpt.top