

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.

 Loss of dependency

```

[321s] CMake Error at /usr/share/cmake/Modules/FindPackageHandleStandardArgs.cmake:230 (message):
[321s] Could NOT find BZip2 (missing: BZIP2_LIBRARIES BZIP2_INCLUDE_DIR)
[321s] Call Stack (most recent call first):
[321s] /usr/share/cmake/Modules/FindPackageHandleStandardArgs.cmake:600 (FPHSA_FAILURE_MESSAGE)
[321s] /usr/share/cmake/Modules/FindBZip2.cmake:66 (FIND_PACKAGE_HANDLE_STANDARD_ARGS)
[321s] /usr/share/cmake/Modules/CMakeFindDependencyMacro.cmake:47 (find_dependency)
[321s] /usr/lib64/cmake/QuaZip-QZ5-1.4/QuaZip-QZ5Config.cmake:52 (find_dependency)
[321s] cmake/Unix.cmake:106 (find_package)
[321s] CMakeLists.txt:122 (include)
[321s]
[321s] -- Configuring incomplete, errors occurred!
[321s] See also
[321s] /home/buildrpm/build/BUILD/nomacs-3.16.224/magelounge/build/CMakeFiles/CMakeOutput.log'.
[321s] error: Bad exit status from /var/tmp/rpm-tmp.EsuhaQ (1)build)
[321s]
[321s] RPM build errors:
[321s] Bad exit status from /var/tmp/rpm-tmp.EsuhaQ (1)build)
[321s] ## VM INTERACTION START ##
[322s] [ 314.93730] T1) sysrq: Power Off
[322s] [ 314.971659] T1) reboot: Power down
[322s] ## VM INTERACTION END ##
[322s] old-cirrus failed 'build nomacs.spec' at Tue Jan 24 06:31:58 UTC 2023.
[322s]
[2329s] -- creating baselibs
[2331s] -- running 01-delete-x86_64-subarches
[2331s] -- saving statistics
[2332s] -- saving built packages
[2332s] RPMS/noarch/nomacs-lang-3.16.224-1.1.noarch.rpm
[2332s] SRPMS/nomacs-3.16.224-1.1.src.rpm
[2332s] OTHER/ statistics
[2332s] OTHER/rpmlintlog
[2332s]
[2332s] build70 finished 'build nomacs.spec' at Sun May 14 12:57:38 UTC 2023.
[2332s]
[2332s] ## VM INTERACTION START ##
[2332s] [ 2327.125728] T1) sysrq: Power Off
[2332s] [ 2327.126920] T1) reboot: Power down
[2332s] ## VM INTERACTION END ##
[2332s] build: extracting built packages...
[2332s] RPMS/noarch/nomacs-lang-3.16.224-1.1.noarch.rpm
[2332s] RPMS/riscv64/nomacs-3.16.224-1.1.riscv64.rpm
[2332s] SRPMS/nomacs-3.16.224-1.1.src.rpm
[2332s] OTHER/ statistics
[2332s] OTHER/rpmlintlog
[2332s]
44 BuildRequires: pkgconfig(libraw) >= 0.17
45 BuildRequires: pkgconfig(libtiff-4)
46 BuildRequires: pkgconfig(zlib)
47 BuildRequires: libb2-devel
48 Recommends: %(name)-lang
49

```

Add the missing dependencies in spec files.

Motivation

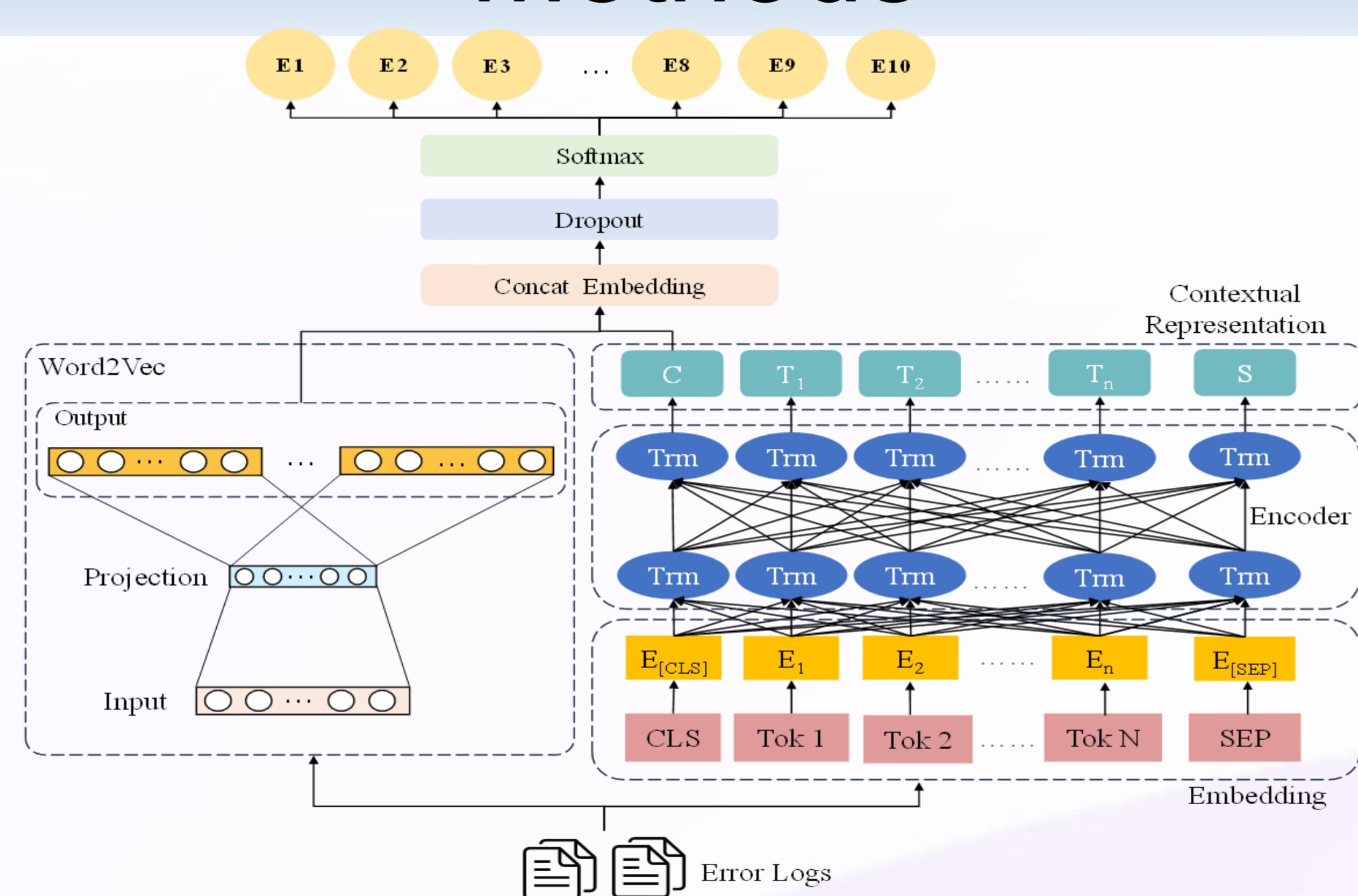
RQ1: What causes RISC-V package building failures?

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

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**.

Methods



Results

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