CC2Vec: Distributed Representations of Code Changes

王德泽

Thong Hoang (James)
Associate Professor
Singapore Management University

David Lo
Associate Professor
Singapore Management University

Julia Lawall
Senior Research Scientist
Sorbonne University/Inria/LIP6, France

```
1   commit 45d787b8a946313b73e8a8fc5d501c9aea3d8847
2   Author: Johannes Berg <johannes.berg@intel.com>
3   Date:  Fri Sep 17 00:38:25 2010 +0200
4
5     wext: fix potential private ioctl memory content leak
6
7     commit df6d02300f7c2fbd0fbe626d819c8e5237d72c62 upstream.
8
9     When a driver doesn't fill the entire buffer, old
10    heap contents may remain, ...
11
12    Reported-by: Jeff Mahoney <jeffm@suse.com>
13    Signed-off-by: Johannes Berg <johannes.berg@intel.com>
14    Signed-off-by: John W. Linville <linville@tuxdriver.com>
15    Signed-off-by: Greg Kroah-Hartman <gregkh@suse.de>
16
17  diff --git a/net/wireless/wext.c b/net/wireless/wext.c
18  index d98ffb7..6890b7e 100644
19  --- a/net/wireless/wext.c
20  +++ b/net/wireless/wext.c
21  @@ -947,7 +947,7 @@ static int ioctl_private_iw_point(...
22          } else if (!iwp->pointer)
23                  return -EFAULT;
24
25  -       extra = kmalloc(extra_size, GFP_KERNEL);
26  +       extra = kzalloc(extra_size, GFP_KERNEL);
27          if (!extra)
28                  return -ENOMEM;
29
```

Log Message Written in natural language

Code Changes Remove or add across one or multiple files

Figure 3.   A bug fixing patch, applied to stable kernel Linux 2.6.27
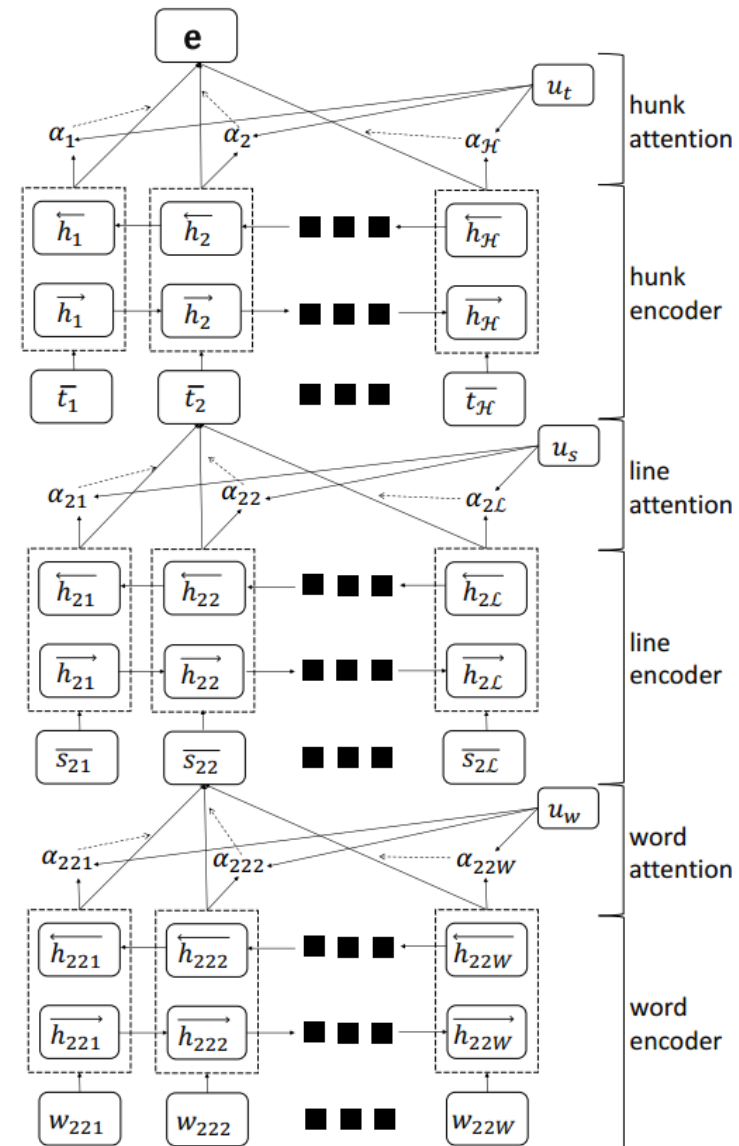
Overall Framework

A prediction function $\mathbf{f} : P \rightarrow Y$
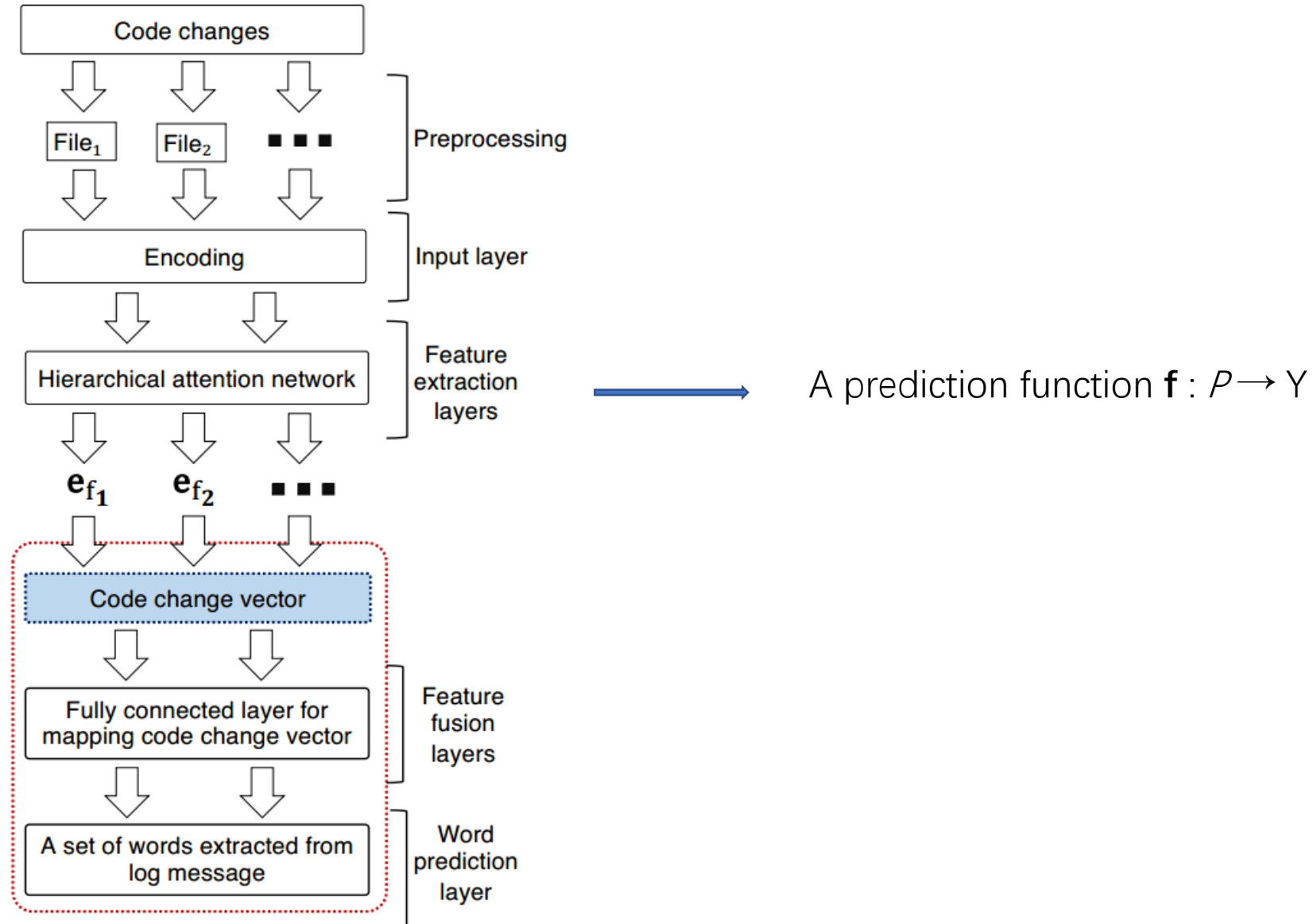
Hierarchical Attention network

Removed (added) code :
Three-dimensional matrix
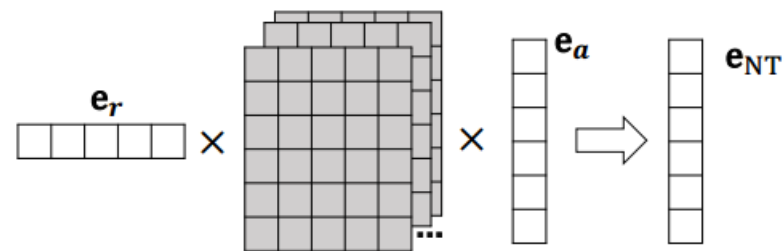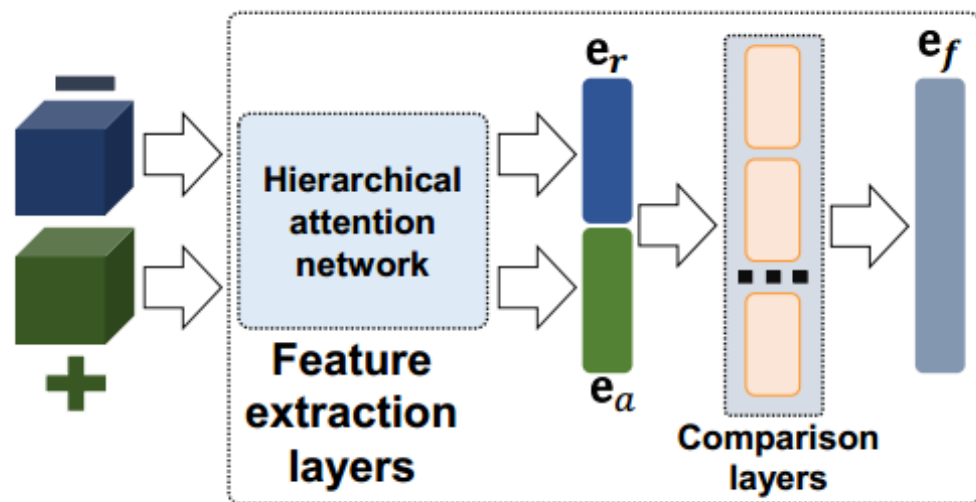$B \in R^{H \times L \times W}$

H: the number of hunks

L: the number of removed (added) code lines for each hunk
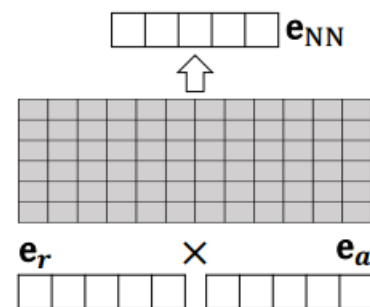
W: the number of words in each removed (added) code line in the affected file
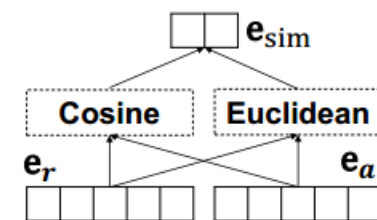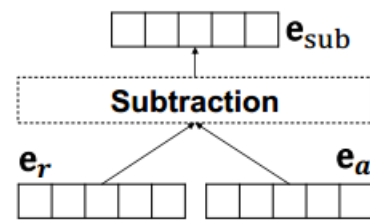
Feature Extraction Layers

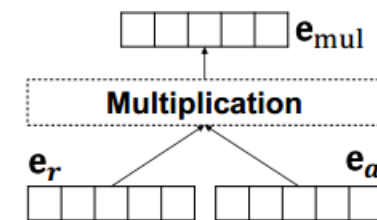(a) Neural Tensor Network (NT)

(b) Neural Network (NN)
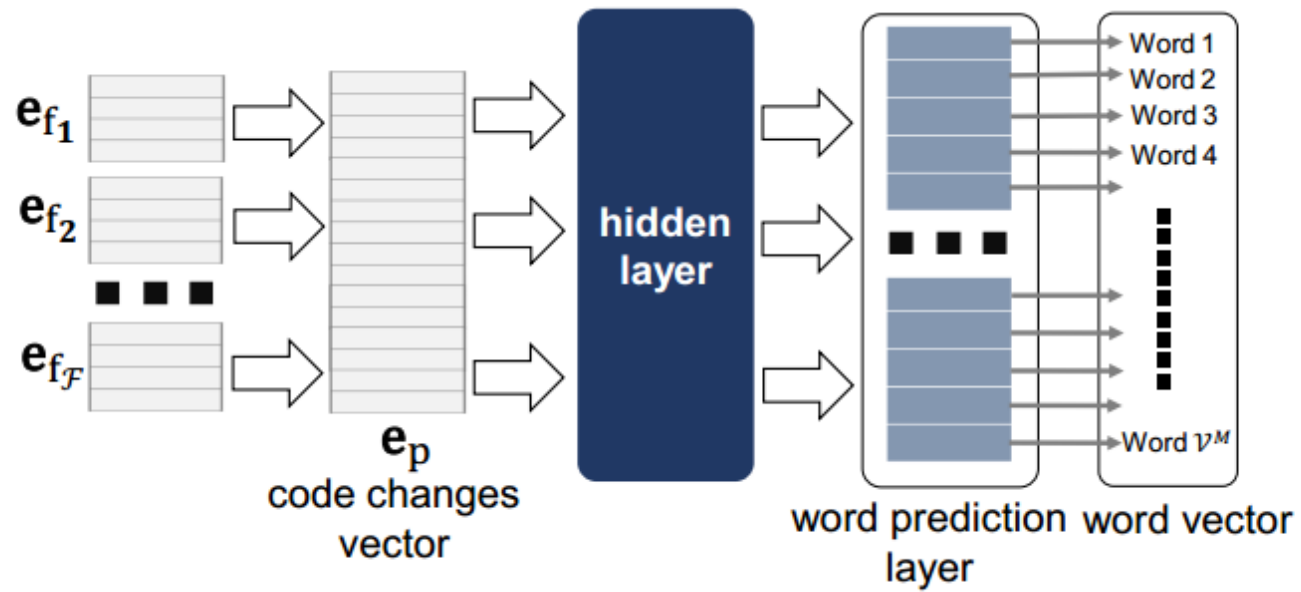
(c) Similarity (sim)

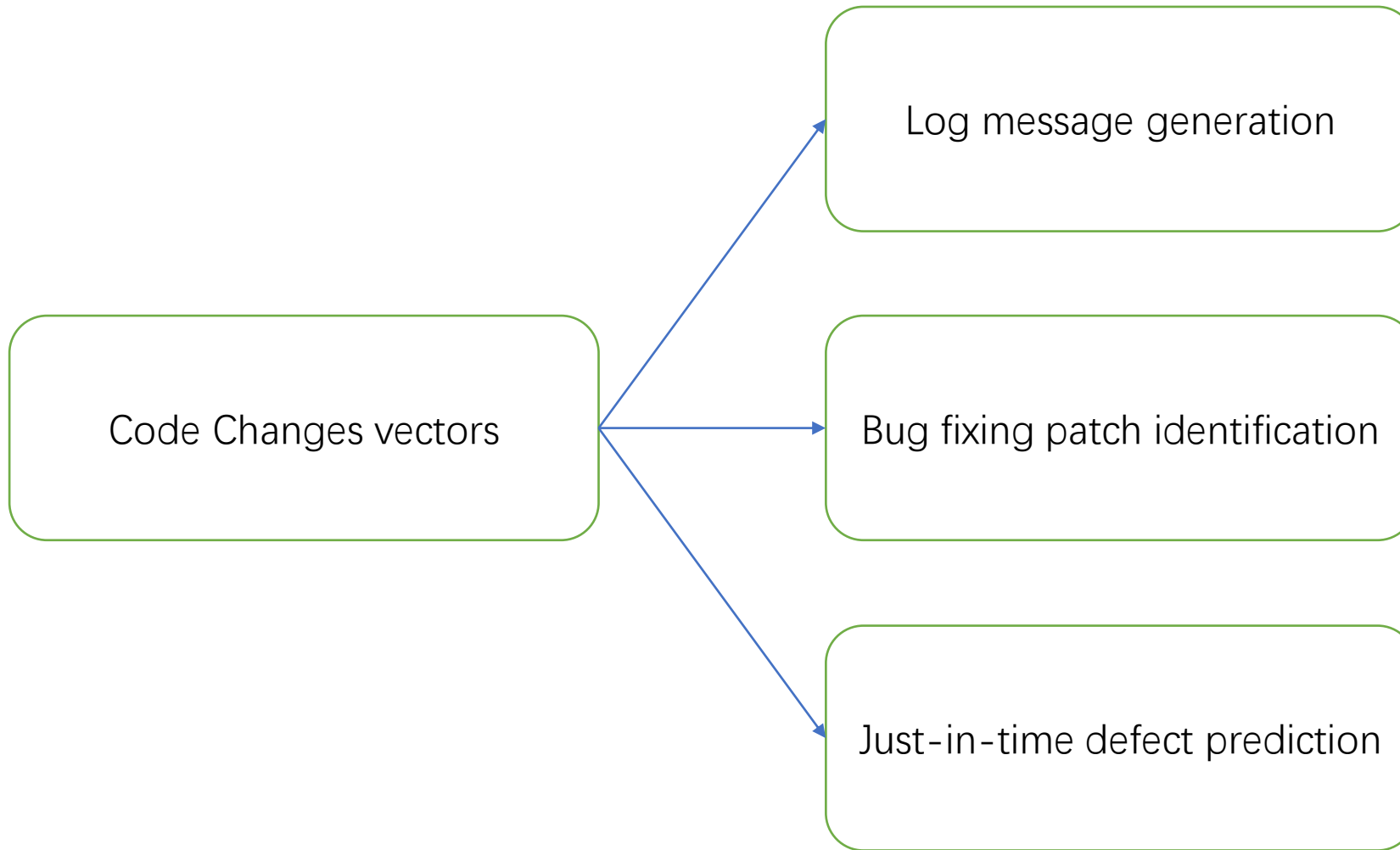(d) Element-wise subtraction

(e) Element-wise multiplication

Word Prediction layer

Experiments

想法

- 经验：
  - 经典思想应用新的领域（first to code change）
  - 实验部分显示更强的应用价值

- 问题：
  - HAN的代价（H、L、W的统一）
  - Feature Extraction比较added/removed向量的意义
  - 没有source code