

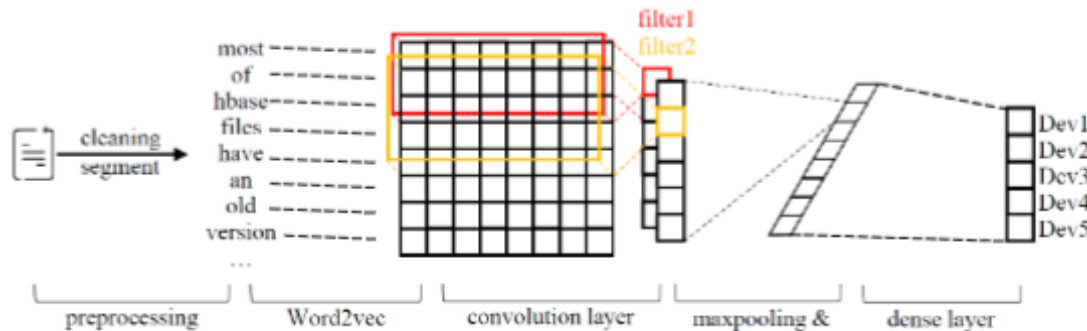
# Automatic Bug Assignments Without Texts: A Study

**Zexuan LI, Kaixin HUANG**

Frontiers of Computer Science, DOI: [10.1007/s11704-024-3299-6](https://doi.org/10.1007/s11704-024-3299-6)

# Problems & Ideas

- Problems of automatic bug assignment approaches:
  - Existing approaches focus on improving NLP techniques for more accurate assignments without pre-validation of their effectiveness.
  - Previous researches point out the limitation of texts under classical techniques.
- Ideas: Exploring the effectiveness of texts for assignments under advanced NLP techniques; Exploring the influential features for further improvements.



**Table 2** Accuracy of TextCNN/TF-IDF with only texts

Method	hbase	cassandra	lucene	pdfbox	shiro
TextCNN	0.206	0.169	0.394	0.500	0.658
SVM+Tf-Idf	0.103	0.120	0.233	0.535	0.711
NB+Tf-Idf	0.107	0.137	0.231	0.531	0.711

**Table 3** Effectiveness comparison of textual features and nominal features

Type	cassandra	hbase	lucene	pdfbox	shiro
Textual Features	0.2228	0.2071	0.5281	0.8497	0.6897
Nominal Features	0.2823	0.3993	0.4494	0.6764	0.6207

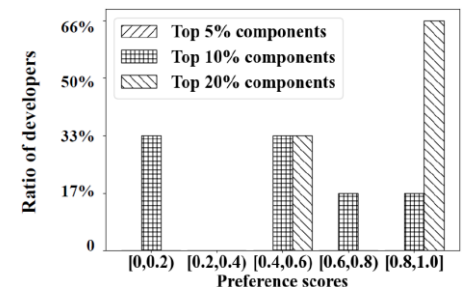
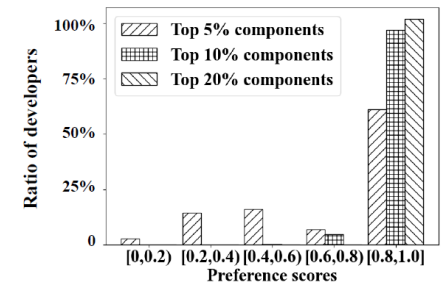
Advanced technique (TextCNN) conducted on automatic bug assignment. Left: TextCNN outperforms classical techniques; Right: TextCNN Disparity map produced by the proposed joint view synthesis and disparity refinement model and its error map (25.80%).

# Main Contributions

- Contributions:
  - Textual features are overestimated that they can hardly achieve great performance as expected without well-designed approaches;
  - Research indicates that the effectiveness of nominal features is underestimated. Nominal features can achieve competitive results without using texts;
  - Research provides the explanation from a statistical view that nominal features all direct to developers' expertise and interest.

**Table 4** Key features of projects

	cassandra	hbase	lucene	pdfbox	shiro
Selected Features	Reporter	Reporter	Reporter	Version	Version
	Version	Version	Version	Component	Platform
	Description	Priority	Component	Reporter	Component
	Component	Platform	Priority	Priority	Priority
	Priority	Component	Summary	Platform	Summary
	Platform	Description	Platform	Summary	Reporter
	Summary	Summary	Description	Description	Description



Preference scores distribution of developers on components. More than 80% developers focus on a very small number of components. It means that the range of candidate developers is limited if the component is given.