

## EXTRACTING DATA FROM THE HASH MEMORY USING AN ERROR KEY

M.H. Nguyen, Yu.O. German

*Educational Institution “Belarusian State University of Informatics and  
Radioelectronics”, Minsk, Republic of Belarus*

**Abstract.** This paper addresses the problem of retrieving data from hash memory using erroneous keys. Propose a character-level Convolutional Neural Network (CNN) model that corrects misspelled input words before hash table lookup. Approach generates synthetic errors (deletion, insertion, transposition) with up to two modifications from a base vocabulary of 10 words. The model encodes input strings as character sequences, applies embedding and 1D convolution layers, and classifies them into the correct word classes. Experimental results show high accuracy in correcting common typing errors, demonstrating the feasibility of using neural networks as a preprocessing layer for fault-tolerant hash memory systems.

**Keywords:** character-level; CNN; accuracy; hash memory; fault-tolerant; deletion; insertion; transposition; tensorflow; epoch.

### Introduction

Hash memory enables fast data retrieval but requires exact key matching, making it vulnerable when keys contain spelling errors. This paper proposes a character-level Convolutional Neural Network (CNN) that corrects erroneous keys before hash table access, using synthetic errors (deletion, insertion, transposition) for training. Experimental results on a 10-word vocabulary demonstrate that our approach achieves high accuracy in correcting keys with up to two errors, offering a fault-tolerant solution for hash-based memory systems.

## Main Part

In hash-based memory systems, even a single character error in the lookup key leads to either a cache miss or retrieval of incorrect data. Traditional error correction methods (e.g., Hamming codes) work at the bit level but cannot handle semantic errors in human-generated keys.

Existing approaches include Edit distance (Levenshtein, 1966), Phonetic algorithms (Soundex, Metaphone), Neural approaches (Kim et al., 2016), and BERT-based correction (Devlin et al., 2019). Unlike these methods that require dictionary lookup or large language models, our CNN-based classifier offers a balance between speed and accuracy for fixed-vocabulary hash key correction.

Simulate three common error types:

- deletion: remove a character (e.g., apple → aple);
- insertion: add a random character (e.g., apple → appple);
- transposition: swap adjacent characters (e.g., apple → papple).

CNN architecture consists of:

- embedding layer: 32-dimensional character embeddings;
- conv1D: 64 filters, kernel size 3, ReLU activation;
- global Max Pooling: Extract most salient features;
- dense layers: 64 neurons (ReLU) followed by 10-class softmax.

Experiments were conducted on a CPU with 16GB RAM. The model was implemented using TensorFlow 2.x. Character-level CNN model has evaluated on 200 randomly generated test samples (20 per word). The figure 1 shows the training progress over 5 epochs.

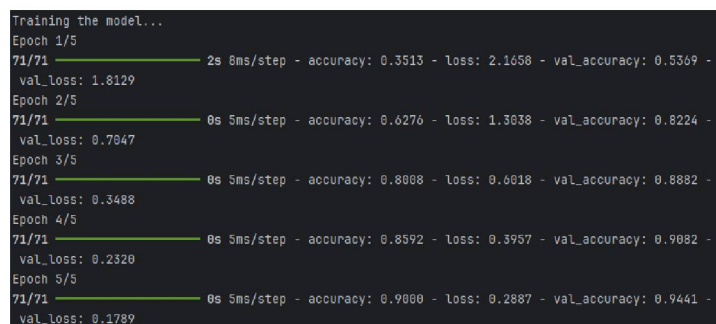


Figure capture: Training and validation accuracy over 5 epochs

The model achieved: accuracy by word: apple – 90.00% (18/20); banana – 100.00% (20/20); orange – 100.00% (20/20); grape – 95.00% (19/20); peach – 100.00% (20/20); cherry – 100.00% (20/20); lemon – 90.00% (18/20); mango – 100.00% (20/20); berry – 100.00% (20/20); melon – 75.00% (15/20).

Summary of findings: overall test accuracy: 95.0%; validation accuracy: 94.4%; training accuracy: 90.0%; words with 100% accuracy: 8/10; lowest accuracy: 75% (melon).

## Conclusion

This paper has proposed a character-level CNN for correcting erroneous keys in hash memory systems, achieving 95.0% accuracy on synthetic errors with up to two modifications. The model successfully handles deletion, insertion, and transposition errors while maintaining sub-millisecond inference time, making it suitable for real-time preprocessing.

## References

1. Devlin J., Chang M. W., Lee K., Toutanova K. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *In Proceedings of NAACL-HLT*, 4171–4186.
2. Kim Y., Jernite Y., Sontag D., Rush A. M. (2016) Character-aware neural language models. *In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2741–2749.

## Information about the authors

**Nguyen M.H.**, Master’s student, Educational Institution “Belarusian State University of Informatics and Radioelectronics”, minhhiiep452001@gmail.com.

**German Yu.O.**, Cand. Sci. (Tech.), Associate Professor, Associate Professor of the Department of Information Technologies of Automated Systems, Educational Institution “Belarusian State University of Informatics and Radioelectronics”, jgerman@bsuir.by.