Potential blend words detected	True blends matched	Precision
10448	128	0.012251

Table 2: Detection Stats

Knowledge Gained

1. The preprocessing steps helps in reducing the potentially useless candidates which in turn results in higher precision.
2. The Jaro algorithm is a measure of characters in common, with consideration for transpositions.[3] It alone isn't much effective in detecting two words fused to form blends.But can be used to get similarity in parts of the blends.
3.The combination of preprocessing steps, the word length difference between candidate and dictionary and along with Jaro similarity results in increased precision for detecting blends in Twitter data[4].

Conclusion

Defining a true blend is complicated by the difficulty of determining which parts of a new word are "recoverable" i.e. the roots which can be distinguished [2].
But this report establishes that the process of detecting true blends in data can be made effective with increased precision by utilizing preprocessing and Jaro similarity with difference in length technique.

Reference

[1]https://medium.com/@appaloosastore/string-similarity-algorithms-compared-3f7b4d12f0ff

[2] https://en.wikipedia.org/wiki/Blend_word

[3]https://stackoverflow.com/questions/25540581/difference-between-jaro-winkler-and-levenshtein-distance

[4] Jacob Eisenstein, Brendan O'Connor, Noah A. Smith,and Eric P. Xing. 2010. A latent variable model for geographic lexical variation. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010), pages 1277–1287

[5] Deri, A. and Knight, K. (2015) How to Make a Frenemy: Multitape FSTs for Portmanteau Generation. In Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, pages 206–210

[6] Das, K. and Ghosh, S. (2017) Neuramanteau: A Neural Network Ensemble Model for Lexical Blends. In Proceedings of the 8th International Joint Conference on Natural Language Processing, pages 576–583

[7] Cook, P. and Stevenson, S. (2010) Automatically Identifying the Source Words