Scopus

Documents

Arkok, B., Zeki, A.M.

Classification of Quranic Topics Using SMOTE Technique
(2021) *International Conference of Modern Trends in ICT Industry: Towards the Excellence in the ICT Industries, MTICTI 2021*, .

DOI: 10.1109/MTICTI53925.2021.9664774

International Islamic University Malaysia, Kulliyyah of Information and Communication Technology, Malaysia

Abstract
This paper aims to classify the Quranic topics that differ in their number of verses by applying the SMOTE technique. SMOTE is used to rebalance samples of minority classes in these Quranic topics. Moreover, SMOTE is combined with many classifiers to choose the best technique for the Quranic classification. Also, the k-values of SMOTE were studied to select the best values for the Quranic datasets and improve the performance of imbalanced classification. The SMOTE was implemented with many classifiers to choose the best one. The results showed that the Voted Perceptron classifier was the best technique when implemented with the SMOTE method to classify the Quranic topics. Also, it is concluded that the best range of K numbers in SMOTE method is [1, 10], to obtain the higher performance of Quranic classification. © 2021 IEEE.

Author Keywords
Imbalanced Classification; Quran; Re-sampling techniques; SMOTE; Topics

Index Keywords
Imbalanced classification, k-Value, K-values, Performance, Quran, Re-sampling technique, Resampling, Sampling technique, SMOTE, Topic; Classification (of information)

References

- Haixiang, G.
  Learning from class-imbalanced data: Review of methods and applications
  (2017) *Expert Systems with Applications*, 73, pp. 220-239.
- Lopez, V.
  An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics
  (2013) *Information Sciences*, 250, pp. 113-141.
- Chawla, N.V.
  SMOTE: Synthetic minority over-sampling technique
  (2002) *Journal of Artificial Intelligence Research*, 16, pp. 321-357.
- Tahir, M.A.
  A multiple expert approach to the class imbalance problem using inverse random under sampling
  (2009) *International Workshop on Multiple Classifier Systems*,
  Springer
- Chawla, N.V., Japkowicz, N., Kotcz, A.
  Special issue on learning from imbalanced data sets
  (2004) *Acm Sigkdd Explorations Newsletter*, 6 (1), pp. 1-6.

- Ali, H.
  A review on data pre-processing methods for class imbalance problem
  (2019) *International Journal of Engineering &technology*, 8 (3), pp. 390-397.

- He, H., Garcia, E.A.
  Learning from imbalanced data
  (2009) *Ieee Transactions on Knowledge and Data Engineering*, 21 (9), pp. 1263-1284.

- Blaszczynski, J., Stefanowski, J.
  Neighbourhood sampling in bagging for imbalanced data
  (2015) *Neurocomputing*, 150, pp. 529-542.

- Han, H., Wang, W.-Y., Mao, B.-H.
  Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning
  (2005) *International Conference on Intelligent Computing*,
  Springer

- Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.
  Safelevel-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem
  (2009) *Pacific-Asia Conference on Knowledge Discovery and Data Mining*,
  Springer

- He, H.
  ADASYN: Adaptive synthetic sampling approach for imbalanced learning
  (2008) *2008 Ieee International Joint Conference on Neural Networks IEEE World Congress on Computational Intelligence*,

- Tang, B., He, H.
  KernelADASYN: Kernel based adaptive synthetic data generation for imbalanced learning
  (2015) *2015 Ieee Congress on Evolutionary Computation (CEC*,

- Dong, Y., Wang, X.
  A new over-sampling approach: Random-SMOTE for learning from imbalanced data sets
  (2011) *International Conference on Knowledge Science Engineering and Management*,
  Springer

- Hu, S.
  MSMOTE: Improving classification performance when training data is imbalanced
  (2009) *2009 Second International Workshop on Computer Science and Engineering*,

- Zhang, H., Li, M.
  RWO-Sampling: A random walk over-sampling approach to imbalanced data classification
  (2014) *Information Fusion*, 20, pp. 99-116.

- Sukarna Barua, M.
  MWMOTE-Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning
  (2012) *Ieee Transactions on Knowledge and Data Engineering*, 26, p. 2.

- Babu, S., Ananthanarayanan, N.
  EMOTE: Enhanced minority oversampling technique
  (2017) *Journal of Intelligent & Fuzzy Systems*, 33 (1), pp. 67-78.

- Zhang, T., Yang, X.
  (2018) *G-SMOTE: A GMM-based Synthetic Minority Oversampling Technique for Imbalanced Learning*,
  arXiv preprint arXiv 1810 10363

- Piri, S., Delen, D., Liu, T.
  A synthetic informative minority oversampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets
  (2018) *Decision Support Systems*, 106, pp. 15-29.

- Zhu, T.
  Minority oversampling for imbalanced ordinal regression
  Knowledge-Based Systems, 2019 (166), pp. 140-155.

- Koziarski, M., Krawczyk, B., Wozniak, M.
  Radial-Based oversampling for noisy imbalanced data classification
  Neurocomputing, 2019 (343), pp. 19-33.

- Al-Khair, A.A., Kabbani, M.A.
  Quran teacher intonation
  (2003) *The Tunisian Company for Distribution*,

- Khedher, M.Z.
  Multiword corpus of the Holy Quran
  (2017) *International Journal on Islamic Applications in Computer Science and Technology*, 5 (1), pp. 1-11.

- Khorsheed, M.S., Al-Thubaity, A.O.
  Comparative evaluation of text classification techniques using a large diverse Arabic dataset
  (2013) *Language Resources and Evaluation*, 47 (2), pp. 513-538.

- Skryjomski, P., Krawczyk, B.
  Influence of minority class instance types on SMOTE imbalanced data oversampling
  (2017) *First International Workshop on Learning with Imbalanced Domains: Theory and Applications*,

- Blagus, R., Lusa, L.
  SMOTE for high-dimensional classimbalanced data
  (2013) *Bmc Bioinformatics*, 14 (1), p. 106.

- Al-Azani, S., El-Alfy, E.-S.M.
  Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text
  (2017) *ANT/SEIT*,

- Patel, H., Thakur, G.S.
  Classification of imbalanced data using a modified fuzzy-neighbor weighted approach
  (2017) *International Journal of Intelligent Engineering and Systems*, 10 (1), pp. 56-64.

- Tang, Y.
  SVMs modeling for highly imbalanced classification
  (2009) *Ieee Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics*, 39 (1), pp. 281-288.

Publisher: Institute of Electrical and Electronics Engineers Inc.

Conference name: 2021 International Conference of Modern Trends in Information and Communication Technology Industry, MTICTI 2021
Conference date: 4 December 2021 through 6 December 2021
Conference code: 176252

ISBN: 9781665414449
Language of Original Document: English
Abbreviated Source Title: Int. Conf. Modern Trends ICT Ind.: Towards Excell. ICT Ind., MTICTI
2-s2.0-85124960388
Document Type: Conference Paper
Publication Stage: Final
Source: Scopus