Because the grid of the spatial data or clusters is expressed as MBR in two dimensions, four coordinates of two diagonal points are used. When each coordinate is 4 bytes, a total of 16 bytes of a space is required. Fig. 8 shows the space utilization before and after compression by increasing the storage unit of the hard disk from 512 to 4096 bytes, assuming each coordinate is 4 bytes.

Normal MBRs represent a 16-byte MBR, with 32 stored in a 512-byte space. However, when storing only one MBR in bit compression, it takes 2 bytes and 256 MBRs are stored. In this paper, the multi-compression technique saves more spatial data because three MBRs are stored on one axis. In other words, Multi-Bit Zipped, which represents multi-compression, stores up to 144 MBRs in 4096 storage spaces, which is 24 times more efficient than normal MBRs.

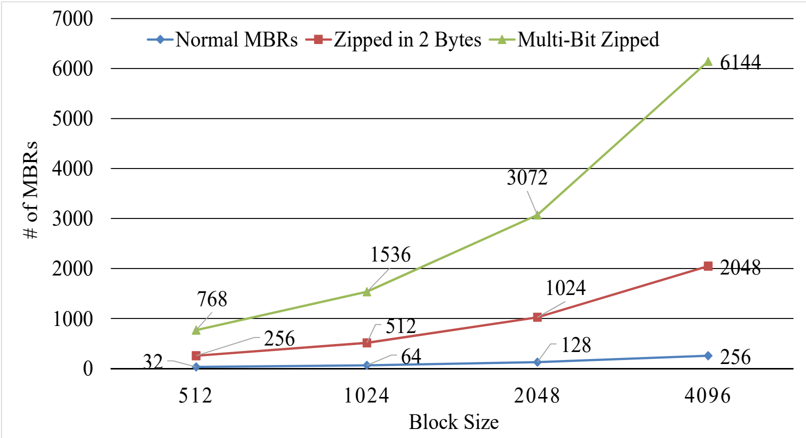

Fig. 8. Block utilization of multi-bit compression.

5. Conclusions

In this study, we proposed a multi-bit compression scheme in machine learning that can improve the storage efficiency and processing speed of clustered data bundles in geographic spatial data. Existing location-based services or clustering techniques have increased the processing efficiency by improving the algorithm, however the performance improvement through spatial data compression was not considered.

Therefore, we divided the search area into grids and extended each side of the MBR that surrounds the spatial data to grid lines. The location where the extended plane meets the grid line was represented by 1 and the location where it did not meet by 0, hence each axis was represented by 8 bits. In particular, the compression rate was increased by storing three MBRs in 8 bits. The experiments showed 24 times better performance in 4K blocks compared to the MBR before compression.

The proposed scheme is expected to increase the service satisfaction of users by providing faster services through coordinate compression in services using spatial data.

Acknowledgments

This paper was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (Ministry of Education) [NRF-2017R1D1A1B03035884].

References

[1] Elghazal, M. S.; Younis, S. A.; Musbah, M. S. (2017): Applying location based services for reducing mobile power consumption. 2017 International Conference on Engineering and Technology (ICET), Antalya, pp. 1–5.
[2] Ekong, P. S.; Cardona, C. J.; Bryssinckx, W.; Ikechukwu-Eneh, C.; Lombin, L. H.; Carpenter, T. E. (2017): Spatial clustering of pathology submissions during the initial introduction and spread of avian influenza H5N1 in poultry in Nigeria in 2006–2007. Veterinaria Italiana, 54(1), pp. 13–20.
[3] Gartner, G.; Huang, H. (2016): Progress in Location-Based Services 2016. Lecture Notes in Geoinformation and Cartography, Springer.
[4] Guttman, A. (1984): R-trees: a dynamic index structure for spatial searching. Proceedings of ACM SIGMOD International Conference on Management of Data, 14, pp. 47–57.
[5] Jannu, S.; Jana, P. K. (2016): A grid based clustering and routing algorithm for solving hot spot problem in wireless sensor networks. Wireless Networks, Springer, pp. 1901–1916.
[6] Petch, J. (1999): GIS, Organisations and People. CRC Press, 1ed.
[7] Spatial Data Generator, DaVisual Code1.0. Available on: http://isl.cs.unipi.gr.
[8] Wu, B.; Wilamowski, B. M. (2017): A Fast Density and Grid Based Clustering Method for Data With Arbitrary Shapes and Noise. IEEE Transactions on Industrial Informatics, 13(4), pp. 1620–1628.
[9] Zhang, J.; Feng, X.; Liu, Z. (2018): A Grid-Based Clustering Algorithm via Load Analysis for Industrial Internet of Things. IEEE Access, 6, pp. 13117–13128.
[10] Zhou, E.; Mao, S,; Li, M.; Sun, Z. (2016): PAM spatial clustering algorithm research based on CUDA. 2016 24th International Conference on Geoinformatics, Galway, pp. 1 –7.