mathematics

MDPI

Article

A Novel Low-Query-Budget Active Learner with Pseudo-Labels for Imbalanced Data

Alaa Tharwat * and Wolfram Schenck

Center for Applied Data Science Gütersloh (CfADS), FH Bielefeld-University of Applied Sciences, 33619 Bielefeld, Germany; wolfram.schenck@fh-bielefeld.de
* Correspondence: alaa.othman@fh-bielefeld.de

Abstract

In this report, the standard deviation (*Std*) of the results of our experiments are presented.

Experiments with synthetic dataset

Table S1. Comparison between the proposed models (LQBALI and LQBALII) and the random, LLR, AOD, CB, and LHCE models using synthetic data in terms of the standard deviation of the accuracy, sensitivity, and specificity results (in the form of $Std(Acc)/Std(Sen)/Std(Spec)$).

Fn.	IR	Random	LLR	AOD	CB	LHCE	LQBALI	LQBALII
F_1		2.7/ 5.9/3.3	2.1/ 4.5/1.5	0.9/ 1.9/0.5	2.5/ 3.7/5.1	2.1/ 4.0/3.1	0.8/ 1.5/1.7	0.7/ 1.7/2.0
F_2	1	1.2/ 5.9/5.6	1.3/ 4.8/5.6	0.7/ 1.3/0.0	1.4/ 6.0/6.0	1.3/ 7.2/7.1	1.3/ 4.6/5.3	1.6/ 2.5/2.7
F_3		1.8/ 4.8/6.6	1.9/ 5.1/7.8	0.0/ 0.0/0.0	1.6/ 5.7/7.9	1.6/ 5.6/7.5	1.2/ 4.3/4.8	1.5/ 4.3/2.5
F_1		2.0/ 6.4/2.4	2.0/ 7.4/0.9	0.7/ 2.3/0.8	5.2/ 4.7/8.5	2.2/ 7.2/2.3	0.8/ 2.4/0.5	0.6/ 1.5/0.7
F_2	2.33	2.4/ 5.5/6.0	2.0/ 4.9/6.0	0.8/ 1.1/0.0	3.5/ 7.5/6.6	2.1/ 4.9/5.7	1.7/ 3.6/4.0	1.2/ 1.7/4.0
F_3		2.0/ 6.5/8.2	1.8/ 5.4/7.8	0.0/ 0.0/0.0	2.2/ 5.9/7.9	2.1/ 7.1/8.7	1.9/ 5.2/7.1	2.0/ 3.5/3.0
F_1		1.8/ 7.7/2.3	1.2/ 8.0/1.0	0.7/ 2.7/0.9	7.3/ 4.2/9.7	1.6/ 7.2/1.9	6.9/ 3.5/9.1	6.9/ 2.2/8.9
F_2	4	1.8/ 2.9/3.7	3.1/ 5.0/5.9	1.3/ 1.5/0.2	5.2/ 8.2/7.2	2.7/ 4.7/6.5	6.3/ 10.0/9.0	6.1/ 9.2/6.4
F_3		2.9/ 7.5/8.4	2.9/ 5.8/7.3	0.0/ 0.0/0.0	3.1/ 5.6/7.9	3.0/ 7.4/8.4	2.6/ 4.6/6.4	2.6/ 3.5/3.0

Experiments with Real imbalanced datasets

Lower datasets

Table S2. Comparison between the proposed models (LQBALI and LQBALII) and the random, LLR, CB, and LHCE models in terms of standard deviation of the accuracy, sensitivity, and specificity results (in the form of $Std(Acc)/Std(Sen)/Std(Spec)$) with imbalanced datasets with IR< 9.

Ds.	Random	LLR	CB	LHCE	LQBALI	LQBALII
LD1	1.1/4.1/3.8	0.2/0.5/0.4	1.0/4.6/3.8	1.3/8.6/8.5	0.4/2.4/1.6	0.4/1.7/1.3
LD2	1.5/5.9/4.0	0.4/0.5/0.5	1.3/6.8/4.1	0.1/0.0/0.0	1.0/2.8/1.9	0.9/2.7/1.4
LD3	1.0/4.7/1.0	0.2/0.7/0.2	1.1/5.2/0.6	2.6/4.8/4.8	0.3/1.9/0.5	0.3/1.6/0.3
LD4	0.9/5.4/1.5	0.1/0.3/0.1	0.8/4.7/1.0	0.9/6.7/1.3	0.4/1.6/0.8	0.4/1.9/0.8
LD5	0.8/5.2/1.0	0.2/1.1/0.2	0.7/7.0/0.9	0.4/4.2/0.3	0.6/3.9/0.8	0.6/3.4/0.7
LD6	0.6/6.5/1.1	0.1/0.4/0.1	0.4/6.7/0.9	0.1/1.4/0.1	0.2/5.9/0.6	0.2/4.0/0.5

check for updates

Citation: Tharwat, A.; Schenck, W. A Novel Low-Query-Budget Active Learner with Pseudo-Labels for Imbalanced Data. *Mathematics* **2022**, *10*, 1068. https://doi.org/10.3390/math10071068

Academic Editor: Catalin Stoean

Received: 31 January 2022
Accepted: 23 March 2022
Published: 26 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Higher datasets

Table S3. Comparison between the proposed models (LQBALI and LQBALII) and the random, LLR, CB, and LHCE models in terms of the standard deviation of the accuracy, sensitivity, and specificity results (in the form of $Std(Acc)/Std(Sen)/Std(Spec)$) with imbalanced datasets with IR$\geq$ 9.

Ds.	Random	LLR	CB	LHCE	LQBALI	LQBALII
HD1	0.7/6.5/0.7	0.4/4.2/0.1	0.5/6.4/0.6	0.6/6.3/0.5	0.2/1.4/0.3	0.1/1.0/0.2
HD2	1.0/2.2/1.3	0.7/0.9/0.8	0.9/1.7/1.1	0.5/1.0/0.7	1.6/3.3/2.1	1.4/3.9/1.9
HD3	0.5/7.1/0.4	0.3/4.0/0.1	0.4/5.1/0.5	0.4/6.8/0.4	0.3/4.9/0.1	0.1/1.8/0.1
HD4	0.3/6.9/0.3	0.0/0.0/0.0	0.3/5.7/0.1	0.3/7.6/0.2	0.3/2.2/0.3	0.3/1.7/0.3
HD5	0.4/4.8/0.6	0.0/0.0/0.0	0.5/3.7/0.6	0.4/4.3/0.5	0.1/5.4/0.2	0.1/6.7/0.2
HD6	0.1/6.1/0.2	0.0/0.0/0.0	0.0/3.3/0.1	0.1/5.4/0.2	0.1/6.7/0.1	0.1/7.5/0.2

Multi-class datasets

In Tables S4 and 5S, the standard deviations of the LLR algorithm results are zeros. This is because LLR selects the same points in all runs; therefore, the number of points selected for each class is constant. While the results of the other algorithms have some fluctuations because the results may change in each run.

Table S4. Comparison between the proposed model (LQBAL) and the random, LLR, CB, and LHCE models with multi-class imbalanced datasets, query budget 5%, and in terms of the standard deviation of the number of annotated points from each class.

Ds.	Random	LLR	CB	LHCE	LQBAL
MD1	0.7/0.8/0.7	0.0/0.0/0.0	0.6/0.8/0.6	0.7/0.7/0.7	0.4/0.4/0.4
MD2	0.5/0.6/0.8	0.0/0.0/0.0	0.5/0.5/0.6	0.6/0.6/0.8	0.3/0.1/0.3
MD3	0.8/1.5/1.6	0/0/0	0.6/1.6/1.6	0.9/1.2/1.3	0.8/0.8/0.7
MD4	0.7/0.7/0.4/0.4/0.4/0.6	0.0/0.0/0.0/0.0/0.0/0.0	0.8/0.7/0.4/0.4/0.4/0.5	0.7/0.8/0.3/0.3/0.2/0.6	0.2/0.4/0.2/0.2/0.2/0.2
MD5	1.0/0.7/0.5/0.7/0.3/0.6/0.2/0.1	0.0/0.0/0.0/0.0/0.0/0.0/0.0/0.0	1.1/0.7/0.5/0.8/0.2/0.6/0.1/0.1	1.1/0.7/0.5/1.0/0.2/0.5/0.0/0.2	0.2/0.5/0.2/0.3/0.0/0.4/0.0/0.0

Table S5. Comparison between the proposed model (LQBAL) and the random, LLR, CB, and LHCE models with multi-class imbalanced datasets, query budget 10%, and in terms of the standard deviation of the number of annotated points from each class.

Ds.	Random	LLR	CB	LHCE	LQBAL
MD1	0.9/1.0/0.9	0.0/0.0/0.0	0.9/1.0/1.0	0.7/0.7/0.7	0.6/0.7/0.6
MD2	0.7/0.8/1.0	0.0/0.0/0.0	0.8/0.7/1.0	0.6/0.6/0.8	0.2/0.2/0.3
MD3	1.0/1.8/2.0	0/0/0	0.8/3.0/3.1	0.9/1.2/1.3	0.9/0.9/1.1
MD4	1.0/1.1/0.7/0.4/0.5/0.7	0.0/0.0/0.0/0.0/0.0/0.0	1.0/1.0/0.5/0.5/0.5/0.8	0.7/0.8/0.3/0.3/0.2/0.6	0.5/0.6/0.5/0.4/0.3/0.3
MD5	1.6/1.0/0.6/1.0/0.4/0.8/0.2/0.2	0.0/0.0/0.0/0.0/0.0/0.0/0.0/0.0	1.4/1.1/0.8/1.2/0.3/0.9/0.2/0.3	1.1/0.7/0.5/1.0/0.2/0.5/0.0/0.2	0.7/0.5/0.3/0.6/0.1/0.5/0.1/0.0