

Supplementary Information for Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model for Microbiome Data Analysis

Xiaojing Luo¹ and Qinqin Hu^{1*}

^{1*}School of Mathematics and Statistics, Shandong University, No.180 Wenhua West Road, Weihai, 264209, Shandong, People's Republic of China.

*Corresponding author(s). E-mail(s): qqhu@sdu.edu.cn;

Appendix A Supplementary Materials

Table A1 Model Performance Comparison between the Proposed Methods with Increasing Source Sample Size n_S under Different Sparsity for Continuous Outcomes

n_S	Model	Deviance	R^2	MSE	MAE
6 Moderate Effects					
–	Targetonly	435.540	0.545	4.859	1.750
100	BTL-CGLMM ($\alpha = 1$)	253.554	0.928	0.778	0.695
	BTL-CGLMM ($\alpha = 0.3$)	271.746	0.907	0.991	0.773
	Adaptive BTL-CGLMM	239.733	0.926	0.773	0.678
	Adaptive BTL-CGLM	266.901	0.908	0.958	0.765
300	BTL-CGLMM ($\alpha = 1$)	345.338	0.826	1.877	1.089
	BTL-CGLMM ($\alpha = 0.3$)	332.506	0.846	1.654	1.022
	Adaptive BTL-CGLMM	245.268	0.919	0.841	0.703
	Adaptive BTL-CGLM	268.928	0.902	1.011	0.776
500	BTL-CGLMM ($\alpha = 1$)	361.723	0.796	2.212	1.181
	BTL-CGLMM ($\alpha = 0.3$)	342.342	0.831	1.823	1.073
	Adaptive BTL-CGLMM	247.621	0.916	0.868	0.712
	Adaptive BTL-CGLM	269.650	0.902	1.016	0.780
12 Moderate Effects					
–	Targetonly	446.805	0.562	5.649	1.870
100	BTL-CGLMM ($\alpha = 1$)	242.746	0.947	0.683	0.650
	BTL-CGLMM ($\alpha = 0.3$)	268.860	0.913	1.045	0.771
	Adaptive BTL-CGLMM	254.131	0.898	1.192	0.785
	Adaptive BTL-CGLM	283.223	0.856	1.659	0.914
300	BTL-CGLMM ($\alpha = 1$)	338.039	0.868	1.740	1.051
	BTL-CGLMM ($\alpha = 0.3$)	318.573	0.890	1.435	0.955
	Adaptive BTL-CGLMM	250.799	0.911	1.074	0.760
	Adaptive BTL-CGLM	284.798	0.859	1.643	0.919
500	BTL-CGLMM ($\alpha = 1$)	356.528	0.841	2.091	1.154
	BTL-CGLMM ($\alpha = 0.3$)	330.118	0.878	1.609	1.010
	Adaptive BTL-CGLMM	251.396	0.914	1.049	0.756
	Adaptive BTL-CGLM	284.617	0.861	1.614	0.911

n_S denotes the source sample size and only applies to methods that incorporate source data. Adaptive BTL-CGLMM: the Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study; Adaptive BTL-CGLM: a simplified version of Adaptive BTL-CGLMM that ignores the dependencies among samples; BTL-CGLMM ($\alpha = 1$) and BTL-CGLMM ($\alpha = 0.3$): the Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study with a fixed parameter α ; Targetonly: directly applies the Bayesian Compositional Generalized Linear Mixed Model (BCGLMM) to the target dataset. Deviance, R^2 , MSE, and MAE are evaluation metrics for the continuous outcome.

Table A2 Model Performance Comparison between the Proposed Methods with Increasing Source Sample Size n_S under Different Sparsity for Binary Outcomes

n_S	Model	Deviance	MSE	MAE	Misclassification Rate
6 Moderate Effects					
–	Targetonly	118.102	0.202	0.411	0.310
100	BTL-CGLMM ($\alpha = 1$)	45.776	0.051	0.196	0.011
	BTL-CGLMM ($\alpha = 0.3$)	71.136	0.096	0.291	0.017
	Adaptive BTL-CGLMM	57.023	0.069	0.240	0.006
	Adaptive BTL-CGLM	87.501	0.133	0.343	0.064
300	BTL-CGLMM ($\alpha = 1$)	56.763	0.074	0.232	0.050
	BTL-CGLMM ($\alpha = 0.3$)	82.070	0.120	0.327	0.044
	Adaptive BTL-CGLMM	53.466	0.062	0.227	0.007
	Adaptive BTL-CGLM	89.397	0.137	0.349	0.067
500	BTL-CGLMM ($\alpha = 1$)	61.245	0.085	0.244	0.074
	BTL-CGLMM ($\alpha = 0.3$)	77.684	0.111	0.311	0.054
	Adaptive BTL-CGLMM	50.246	0.057	0.215	0.005
	Adaptive BTL-CGLM	89.241	0.137	0.349	0.066
12 Moderate Effects					
–	Targetonly	117.968	0.203	0.414	0.315
100	BTL-CGLMM ($\alpha = 1$)	42.147	0.045	0.182	0.005
	BTL-CGLMM ($\alpha = 0.3$)	70.040	0.093	0.289	0.011
	Adaptive BTL-CGLMM	54.817	0.065	0.233	0.004
	Adaptive BTL-CGLM	87.017	0.134	0.339	0.102
300	BTL-CGLMM ($\alpha = 1$)	46.814	0.056	0.196	0.027
	BTL-CGLMM ($\alpha = 0.3$)	77.623	0.109	0.314	0.026
	Adaptive BTL-CGLMM	50.225	0.056	0.216	0.002
	Adaptive BTL-CGLM	89.392	0.139	0.347	0.110
500	BTL-CGLMM ($\alpha = 1$)	54.683	0.072	0.221	0.057
	BTL-CGLMM ($\alpha = 0.3$)	77.467	0.110	0.312	0.032
	Adaptive BTL-CGLMM	47.619	0.052	0.206	0.002
	Adaptive BTL-CGLM	88.461	0.136	0.345	0.100

n_S denotes the source sample size and only applies to methods that incorporate source data. Adaptive BTL-CGLMM: the Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study; Adaptive BTL-CGLM: a simplified version of Adaptive BTL-CGLMM that ignores the dependencies among samples; BTL-CGLMM ($\alpha = 1$) and BTL-CGLMM ($\alpha = 0.3$): the Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study with a fixed parameter α ; Targetonly: directly applies the Bayesian Compositional Generalized Linear Mixed Model (BCGLMM) to the target dataset. Deviance, MSE (Mean Squared Error), MAE (Mean Absolute Error), and Misclassification Rate are evaluation metrics for the binary outcome.

Table A3 Model Performance Comparison between the Proposed Methods with Different Sources under Heterogeneous Source Dataset for Continuous Outcomes

ξ	Model	Deviance	R^2	MSE	MAE
–	Targetonly	435.540	0.545	4.859	1.750
0.05	BTL-CGLMM ($\alpha = 1$)	331.189	0.849	1.628	1.016
	BTL-CGLMM ($\alpha = 0.3$)	319.358	0.865	1.450	0.960
	Adaptive BTL-CGLMM	255.638	0.908	0.956	0.746
	Adaptive BTL-CGLM	268.234	0.907	0.976	0.771
0.5	BTL-CGLMM ($\alpha = 1$)	333.639	0.845	1.668	1.029
	BTL-CGLMM ($\alpha = 0.3$)	320.714	0.863	1.471	0.968
	Adaptive BTL-CGLMM	252.832	0.910	0.934	0.737
	Adaptive BTL-CGLM	268.733	0.905	0.984	0.774
1	BTL-CGLMM ($\alpha = 1$)	345.859	0.825	1.882	1.096
	BTL-CGLMM ($\alpha = 0.3$)	326.871	0.854	1.565	0.999
	Adaptive BTL-CGLMM	252.571	0.912	0.917	0.733
	Adaptive BTL-CGLM	268.963	0.905	0.983	0.774
2	BTL-CGLMM ($\alpha = 1$)	381.970	0.744	2.738	1.321
	BTL-CGLMM ($\alpha = 0.3$)	350.704	0.816	1.985	1.129
	Adaptive BTL-CGLMM	256.577	0.907	0.968	0.749
	Adaptive BTL-CGLM	270.373	0.902	1.018	0.783

ξ denotes the discrepancy between the source and the target effect, and only applies to methods that incorporate source data.

Adaptive BTL-CGLMM: the Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study; Adaptive BTL-CGLM: a simplified version of Adaptive BTL-CGLMM that ignores the dependencies among samples; BTL-CGLMM ($\alpha = 1$) and BTL-CGLMM ($\alpha = 0.3$): the Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study with a fixed parameter α ; Targetonly: directly applies the Bayesian Compositional Generalized Linear Mixed Model (BCGLMM) to the target dataset. Deviance, R^2 (R-squared), MSE (Mean Squared Error), and MAE (Mean Absolute Error) are evaluation metrics for the continuous outcome.

Table A4 Model Performance Comparison between the Proposed Methods with Different Sources under Heterogeneous Source Dataset for Binary Outcomes

ξ	Model	Deviance	MSE	MAE	Misclassification Rate
–	Targetonly	118.102	0.202	0.411	0.310
0.05	BTL-CGLMM ($\alpha = 1$)	49.301	0.059	0.206	0.027
	BTL-CGLMM ($\alpha = 0.3$)	73.517	0.102	0.298	0.034
	Adaptive BTL-CGLMM	54.752	0.065	0.231	0.007
	Adaptive BTL-CGLM	87.706	0.134	0.344	0.061
0.5	BTL-CGLMM ($\alpha = 1$)	48.768	0.058	0.205	0.026
	BTL-CGLMM ($\alpha = 0.3$)	73.436	0.102	0.297	0.030
	Adaptive BTL-CGLMM	54.717	0.065	0.231	0.007
	Adaptive BTL-CGLM	87.723	0.134	0.343	0.065
1	BTL-CGLMM ($\alpha = 1$)	47.671	0.057	0.200	0.025
	BTL-CGLMM ($\alpha = 0.3$)	73.200	0.101	0.297	0.032
	Adaptive BTL-CGLMM	54.740	0.065	0.231	0.007
	Adaptive BTL-CGLM	88.607	0.136	0.346	0.071
2	BTL-CGLMM ($\alpha = 1$)	46.899	0.056	0.197	0.028
	BTL-CGLMM ($\alpha = 0.3$)	76.004	0.107	0.307	0.034
	Adaptive BTL-CGLMM	54.481	0.065	0.230	0.008
	Adaptive BTL-CGLM	88.282	0.135	0.346	0.065

ξ denotes the discrepancy between the source and the target effect, and only applies to methods that incorporate source data.

Adaptive BTL-CGLMM: the Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study; Adaptive BTL-CGLM: a simplified version of Adaptive BTL-CGLMM that ignores the dependencies among samples; BTL-CGLMM ($\alpha = 1$) and BTL-CGLMM ($\alpha = 0.3$): the Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study with a fixed parameter α ; Targetonly: directly applies the Bayesian Compositional Generalized Linear Mixed Model (BCGLMM) to the target dataset. Deviance, MSE (Mean Squared Error), MAE (Mean Absolute Error), and Misclassification Rate are evaluation metrics for the binary outcome.

Table A5 Model Performance Comparison between the Proposed Methods with Increasing Source Sample Size n_S under Heterogeneous Source Dataset for Continuous Outcomes

n_S	Model	Deviance	R^2	MSE	MAE
–	Targetonly	435.540	0.545	4.859	1.750
100	BTL-CGLMM ($\alpha = 1$)	302.536	0.885	1.240	0.885
	BTL-CGLMM ($\alpha = 0.3$)	300.734	0.886	1.220	0.880
	Adaptive BTL-CGLMM	248.511	0.916	0.873	0.717
	Adaptive BTL-CGLM	268.628	0.906	0.984	0.773
300	BTL-CGLMM ($\alpha = 1$)	349.201	0.819	1.946	1.112
	BTL-CGLMM ($\alpha = 0.3$)	332.570	0.846	1.654	1.027
	Adaptive BTL-CGLMM	255.020	0.909	0.949	0.743
	Adaptive BTL-CGLM	269.800	0.901	1.027	0.782
500	BTL-CGLMM ($\alpha = 1$)	365.798	0.788	2.294	1.206
	BTL-CGLMM ($\alpha = 0.3$)	343.711	0.829	1.843	1.084
	Adaptive BTL-CGLMM	262.845	0.900	1.037	0.774
	Adaptive BTL-CGLM	268.736	0.905	0.993	0.774

n_S denotes the source sample size and only applies to methods that incorporate source data. Adaptive BTL-CGLMM: the Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study; Adaptive BTL-CGLM: a simplified version of Adaptive BTL-CGLMM that ignores the dependencies among samples; BTL-CGLMM ($\alpha = 1$) and BTL-CGLMM ($\alpha = 0.3$): the Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study with a fixed parameter α ; Targetonly: directly applies the Bayesian Compositional Generalized Linear Mixed Model (BCGLMM) to the target dataset. Deviance, R^2 (R-squared), MSE (Mean Squared Error), and MAE (Mean Absolute Error) are evaluation metrics for the continuous outcome.

Table A6 Model Performance Comparison between the Proposed Methods with Increasing Source Sample Size n_S under Heterogeneous Source Dataset for Binary Outcomes

n_S	Model	Deviance	MSE	MAE	Misclassification Rate
–	Targetonly	118.102	0.202	0.411	0.310
100	BTL-CGLMM ($\alpha = 1$)	43.578	0.047	0.188	0.007
	BTL-CGLMM ($\alpha = 0.3$)	75.368	0.105	0.305	0.026
	Adaptive BTL-CGLMM	57.227	0.069	0.241	0.007
	Adaptive BTL-CGLM	88.723	0.136	0.347	0.071
300	BTL-CGLMM ($\alpha = 1$)	54.388	0.070	0.222	0.047
	BTL-CGLMM ($\alpha = 0.3$)	68.015	0.091	0.278	0.032
	Adaptive BTL-CGLMM	52.223	0.061	0.222	0.007
	Adaptive BTL-CGLM	87.728	0.134	0.344	0.062
500	BTL-CGLMM ($\alpha = 1$)	59.032	0.081	0.235	0.075
	BTL-CGLMM ($\alpha = 0.3$)	64.342	0.086	0.263	0.046
	Adaptive BTL-CGLMM	50.836	0.059	0.217	0.007
	Adaptive BTL-CGLM	88.426	0.135	0.346	0.066

n_S denotes the source sample size and only applies to methods that incorporate source data. Adaptive BTL-CGLMM: the Adaptive Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study; Adaptive BTL-CGLM: a simplified version of Adaptive BTL-CGLMM that ignores the dependencies among samples; BTL-CGLMM ($\alpha = 1$) and BTL-CGLMM ($\alpha = 0.3$): the Bayesian Transfer Learning under Compositional Generalized Linear Mixed Model proposed in this study with a fixed parameter α ; Targetonly: directly applies the Bayesian Compositional Generalized Linear Mixed Model (BCGLMM) to the target dataset. Deviance, MSE (Mean Squared Error), MAE (Mean Absolute Error), and Misclassification Rate are evaluation metrics for the binary outcome.