

BDT test in Xi analysis

Liu Liang

liangzy@mail.ustc.edu.cn

1958

Motivation

- High background level in the signal region of neutron anti-neutron invariant mass.
- It might have an opportunity to be improved by applying an MVA method.



Motivation

	Neutron Channel	Anti-neutron channel
2009	2461 ± 56	2946 ± 61
2012	12862 ± 127	13217 ± 127
2018	49162 <u>+</u> 250	55638 <u>+</u> 264
2019	46409 ± 246	51507 ± 252

	Neutron channel	Anti-neutron channel
2009	405±40	377±41
2012	2020±92	1754±86
2018	7991±182	7589±184
2019	7213±304	7246±178

The background level is about ~15%

Brief introduction of BDT

- Decision tree: a binary tree structured classifier (regressor).
- The phase space is split this way into many regions that are eventually classified as signal or background, depending on the majority of training events that end up in the final leaf node.



Brief introduction of BDT

- The boosting of a decision tree extends this concept from one tree to several trees which form a forest.
- The boosted event classifi- cation yBoost(x) is then given by

$$y_{\text{Boost}}(\mathbf{x}) = rac{1}{N_{\text{collection}}} \cdot \sum_{i}^{N_{\text{collection}}} \ln(lpha_i) \cdot h_i(\mathbf{x}) \ ,$$

for signal and background as h(x) = +1 and -1, respectively. The boost weight is derived from the misclassification rate, err, of the previous tree

$$\alpha = \frac{1 - \operatorname{err}}{\operatorname{err}}$$

AdaBoost method, Gradient Boosting method.

Variables of photon shower

double	energy()	Shower energy after correction.
double	dE()	Error for shower energy.
double	eSeed()	Seed"s energy.
double	e3x3()	Energy sum of 3x3 crystal around seed.
double	e5x5()	Energy sum of 5x5 crystal around seed.
double	secondMoment()	Second moment for shower shape.
double	latMoment()	Lateral moment for shower shape.
double	A20Moment()	Zernike-20 moment for shower shape.
double	A42Moment()	Zernike-42 moment for shower shape.
	· · · · · · · · · · · · · · · · · · ·	r *.

Energy, numHits, secondMoment, eSeed, e3x3, e5x5, latMoment, A20Moment, A42Moment

Variables of photon shower



Variables of photon shower



Correlations

Correlation Matrix (signal)

							Line	ar cor	relatio	n coef	licient	s in %		100
A42Moment	-12	-20	12	-23	-22	6	-33	-13		7	34	100		100
A42Moment	8	16	39	4	5	48	-34	-45	1	13	100	34		80
A42Moment	2	36	64	-10	-1	65	-14	-73	1	100	13	7	_	60
A42Moment		2	54			5		44	100	1	1			40
A20Moment	-8	-41	-45	2	-4	-74	18	100	44	-73	-45	-13	_	20
utron eSeed	86	42	-16	90	89	-7	100	18		-14	-34	-33		0
tron latmom	20	61	69	10	17	100	-7	-74	5	65		6		U
neutron e5x5	97	57	4	99	100	17	89	-4		-1	5	-22	_	-20
neutron e3x3	96	52	-1	100	99	10	90	2		-10	4	-23		-40
ron secmom	7	37	100	-1	4	69	-16	-45	54	64	39	12		-60
neutron hits	57	100	37	52	57	61	42	-41	2	36	16	-20		-80
utron energy	100	57	7	96	97	20	86	-8		2	8	-12		10
	ani	i-nan	ani	i-nai	ani	i-nan	ani I-noi	ti-nei	ani	an	ani I-noi	i-nant	-00.	-10
	neutron durneutron durneutron durneutron autoneutron durneutron durneutron durneutron durneutron durneutron durneutron 3 A 42 Money and a 42													

Correlation Matrix (background)









Liu Liang (USTC)

September 15, 2022



Correlation Matrix (background) Linear correlation coefficients in % 100 100 8 neutron var3 80 60 neutron var2 100 40 20 8 A42Moment 100 0 ron secmom 100 -20 -40 neutron hits 100 -60 -80 utron energy 100 -100neutron energy neutron secmom neutron A42Moment neutron hits neutron var2 neutron var3

Liu Liang (USTC)







Class	ifier	(#signal,	#backgr.)	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
	BDTG: BDT:	(1000, 1000.	1000) 1000)	-0.4409 -0.0824	23.1972 23.082	943.664 934.3693	711.202	0.9437 0.9344	0.7112



Correlation Matrix (background)





Background rejection versus Signal efficiency Background rejection versus Signal efficiency Background rejection Background rejection 1 0.9 0.9 0.8 0.8 0.7 0.7 0.6 0.6 0.5 0.5 0.4 0.4 MVA Method: MVA Method: BDTG BDTG 0.3 0.3 BDT BDT 0.2 0.2 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.1 1 n Signal efficiency Signal efficiency

Neutron channel

Anti-neutron channel

Backup



Neutron channel

BDTG

BDT

Rank	:	Variable	:	Variable Importance
1	:	shower_var2	:	2.149e-01
2	:	shower_energy	:	1.981e-01
3	:	shower_A42Moment	:	1.753e-01
4	:	shower_hits	:	1.649e-01
5	:	shower_secmom	:	1.535e-01
6	:	shower_var3	:	9.323e-02
Ranki	ing	g result (top vari	iał	ole is best ranked)
Ranki Rank	ing :	g result (top vari Variable	i ał 	ole is best ranked) Variable Importance
Ranki Rank 1	ing :	g result (top vari Variable shower_energy	i ał : :	ole is best ranked) Variable Importance 2.088e-01
Ranki Rank 1 2	ing : :	g result (top vari Variable shower_energy shower_var2	i ał : : :	ole is best ranked) Variable Importance 2.088e-01 1.974e-01
Rank Rank 1 2 3	ing : : :	g result (top vari Variable shower_energy shower_var2 shower_secmom	i ał : : : :	ole is best ranked) Variable Importance 2.088e-01 1.974e-01 1.921e-01
Rank Rank 1 2 3 4	ing : : :	g result (top vari Variable shower_energy shower_var2 shower_secmom shower_A42Moment	i ał : : : : :	ole is best ranked) Variable Importance 2.088e-01 1.974e-01 1.921e-01 1.821e-01
Rank Rank 1 2 3 4 5	ing : : : :	g result (top vari Variable shower_energy shower_var2 shower_secmom shower_A42Moment shower_hits	i ał : : : : :	Dle is best ranked) Variable Importance 2.088e-01 1.974e-01 1.921e-01 1.821e-01 1.571e-01

BDTG	
------	--

BDT

:	Ranking input variables (method specific) Ranking result (top variable is best ranked)								
:	Rank : Variable	: Variable Importance							
· · · · · · · · · · · · · · · · · · ·	<pre>1 : shower_var2 2 : shower_energy 3 : shower_A42Moment 4 : shower_hits 5 : shower_secmom 6 : shower_var3 Ranking result (top var)</pre>	: 2.149e-01 : 1.944e-01 : 1.840e-01 : 1.642e-01 : 1.502e-01 : 9.235e-02 riable is best ranked)							
:	Rank : Variable	: Variable Importance							
• • • • • • • • • • •	<pre>1 : shower_energy 2 : shower_var2 3 : shower_A42Moment 4 : shower_secmom 5 : shower_hits 6 : shower_var3</pre>	: 2.058e-01 : 1.978e-01 : 1.861e-01 : 1.835e-01 : 1.668e-01 : 6.002e-02							