Drug Recommendation toward Safe Polypharmacy

Wen-Hao Chiang, Li Shen, Li Shen, Xia Ning*

1. Preamble

2. The Proposed Method

1. Preamble

2. The Proposed Method

Drug-drug Interaction

1. DDIs: Drug-drug interactions

The pharmacological effects of a drug are altered by actions of another drug Leading to unpredictable consequences.

2. ADRs: Adverse drug reactions

Undesired or harmful reactions due to drug administration.

A recent thread is dedicated **to understanding the interaction patterns among high-order DDIs**, and how such patterns can relate to induced ADRs.

Two problems

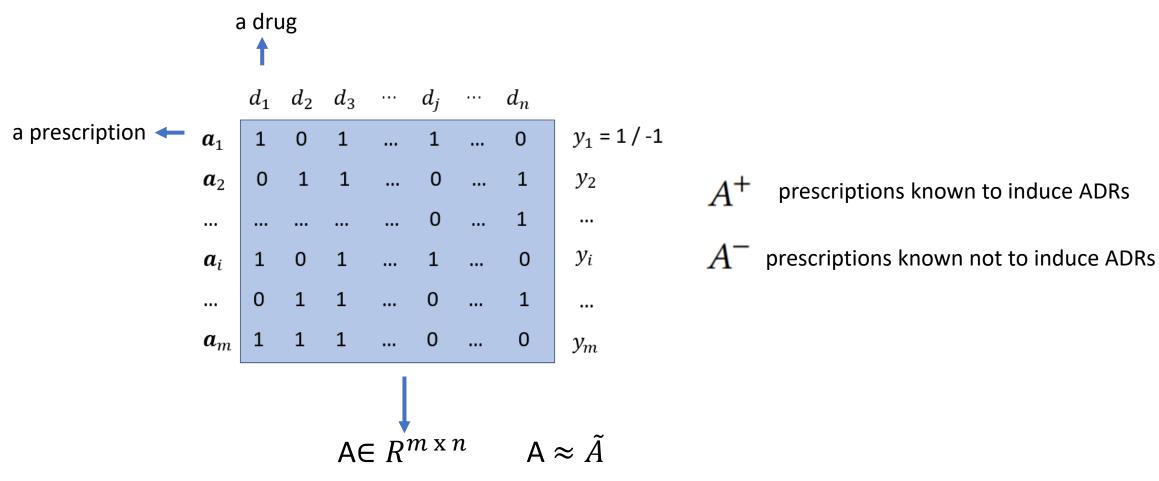
1. To-avoid Drug Recommendation:

Given the multiple drugs in a prescription that have been taken simultaneously, recommend a short of ranked list of drugs that should be avoided taking together with the prescription in order to avoid a particular ADR.

2. Safe Drug Recommendation

Given the multiple drugs in a prescription that have been taken simultaneously, **recommend a short ranked list of safe drugs** that, if token together with the prescription, are not likely to induce a particular ADR.

Notations



Two Methods

1. SLIM (Sparse Linear Method)

Motivation: A matrix decomposition method that decomposes to obtain a sparse matrix of W that can be used for feature selection and compression of the data.

$$\tilde{A} = AW$$

2. LogR (Logistic Regression)

Motivation: binary classification

SLIM:

SLIM learns a sparse coefficient matrix for the items in the system solely from user purchase/rating profiles by solving a regularized optimization problem.

$$\tilde{a}_{ij} = a_i w_j^\mathsf{T}$$
 $\tilde{A} = \mathsf{AW}$

$$\tilde{A} = AW$$

 \tilde{a}_{ij} is the estimated score of d_i in $a_i \cup \{d_i\}$.

 w_i^T is a spare column vector of aggregation coefficients

the sparse coefficient matrix

$$W = [\boldsymbol{w}_1^\mathsf{T}, \boldsymbol{w}_2^\mathsf{T}, \cdots, \boldsymbol{w}_n^\mathsf{T}] \qquad \qquad \min_{W} \quad \text{SLIM}(A; W, \alpha, \lambda) = \frac{1}{2} ||A - AW||_F^2 + \frac{\alpha}{2} ||W||_F^2 + \lambda ||W||_{\ell_1}$$
 subject to $W \ge 0$, $\operatorname{diag}(W) = 0$,

$$|\tilde{A} = AW|$$

Control the sparsity of W

Control the value of W

LogR

the probability of a prescription ai inducing the ADR is modeled as follows

$$p(y_i|\boldsymbol{a}_i;\boldsymbol{x},c) = (1 + \exp(-y_i(\boldsymbol{a}_i\boldsymbol{x}^\mathsf{T} + c)))^{-1}$$

Where x^{T} and c are the parameters. To learn the parameters, LogR solves the following optimization problem,

$$\min_{\mathbf{x},c} \quad \text{LogR}(\mathbf{y}|A;\mathbf{x},c,\beta,\gamma)$$

$$= \sum_{i=1}^{m} \log\{1 + \exp[-y_i(\mathbf{a}_i\mathbf{x}^{\mathsf{T}} + c)]\} + \frac{\beta}{2} \|\mathbf{x}\|_2^2 + \gamma \|\mathbf{x}\|_1$$
where $\mathbf{y} = [y_1; y_2; \cdots, y_m], \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|, \text{ and } \|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} x_i^2.$

1. Preamble

2. The Proposed Method

SlimLogR

SlimLogR learns the SLIM and LogR components jointly through solving the following optimization problem

$$\min_{W^+,W^-, \ \boldsymbol{x}, \ c} \quad \text{SLIM}(A^+;W^+,\alpha,\lambda) + \text{SLIM}(A^-;W^-,\alpha,\lambda)$$

$$\omega\{ \text{LogR}(\boldsymbol{y}^+|\tilde{A}^+ \circ M^+;\boldsymbol{x},c,\beta,\gamma) + \quad M = 1$$

$$\text{LogR}(\boldsymbol{y}^-|\tilde{A}^- \circ M^-;\boldsymbol{x},c,\beta,\gamma) \} \quad M = \mathbb{I}(A)$$
 SlimLogR_{ex} subject to
$$\tilde{A}^+ = A^+W^+, \tilde{A}^- = A^-W^-,$$

$$W^+ \geq 0, W^- \geq 0, \operatorname{diag}(W^+) = 0, \operatorname{diag}(W^-) = 0$$

where \mathbb{I} is an indicator function ((I)(x) = 0 if x = 0, 1 otherwise)

Training SlimLogR

$$\begin{split} \min_{W^+,W^-,\ \boldsymbol{x},\ c} & \quad \text{SLIM}(A^+;W^+,\alpha,\lambda) + \text{SLIM}(A^-;W^-,\alpha,\lambda) \\ & \quad \omega\{\text{LogR}(\boldsymbol{y}^+|\tilde{A}^+\circ M^+;\boldsymbol{x},c,\beta,\gamma) + \\ & \quad \text{LogR}(\boldsymbol{y}^-|\tilde{A}^-\circ M^-;\boldsymbol{x},c,\beta,\gamma)\} \\ & \text{subject to} & \quad \tilde{A}^+ = A^+W^+,\tilde{A}^- = A^-W^-, \\ & \quad W^+ \geq 0, W^- \geq 0, \text{diag}(W^+) = 0, \text{diag}(W^-) = 0 \end{split}$$

$$\begin{split} \min_{W^+,W^-,Z^+,Z^-,\ \pmb{x},\ c} & L(W^+,W^-,Z^+,Z^-,\ \pmb{x},\ c,\ \pmb{u}^+,\ \pmb{u}^-,\rho^+,\rho^-) = \\ & \mathrm{SLIM}(A^+;W^+,\alpha,\lambda) + \mathrm{SLIM}(A^-;W^-,\alpha,\lambda) \\ & \omega \{ \mathrm{LogR}(\pmb{y}^+|\tilde{B}^+\circ M^+;\pmb{x},c,\beta,\gamma) + \\ & \mathrm{LogR}(\pmb{y}^-|\tilde{B}^-\circ M^-;\pmb{x},c,\beta,\gamma) \} \\ & \pmb{u}^{+\mathsf{T}}\pmb{v}^+ + \frac{\rho^+}{2} \|\pmb{v}^+\|_2^2 + \pmb{u}^{-\mathsf{T}}\pmb{v}^- + \frac{\rho^-}{2} \|\pmb{v}^-\|_2^2, \\ \mathrm{subject\ to} & \tilde{B}^+ = A^+Z^+, \tilde{B}^- = A^-Z^-, \\ & \pmb{v}^+ = \mathrm{vec}(W^+) - \mathrm{vec}(Z^+), \pmb{v}^- = \mathrm{vec}(W^-) - \mathrm{vec}(Z^-), \\ & W^+ = Z^+, W^- = Z^-, W^+ \geq 0, W^- \geq 0, \\ & \mathrm{diag}(W^+) = 0, \mathrm{diag}(W^-) = 0, \end{split}$$

Training SlimLogR

$$\begin{split} \min_{W^+,W^-,Z^+,Z^-,\ \pmb{x},\ c} & L(W^+,W^-,Z^+,Z^-,\ \pmb{x},\ c,\ \pmb{u}^+,\ \pmb{u}^-,\rho^+,\rho^-) = \\ & \mathrm{SLIM}(A^+;W^+,\alpha,\lambda) + \mathrm{SLIM}(A^-;W^-,\alpha,\lambda) \\ & \omega \{ \mathrm{LogR}(\pmb{y}^+|\tilde{B}^+\circ M^+;\pmb{x},c,\beta,\gamma) + \\ & \mathrm{LogR}(\pmb{y}^-|\tilde{B}^-\circ M^-;\pmb{x},c,\beta,\gamma) \} \\ & \pmb{u}^{+\mathsf{T}} \pmb{v}^+ + \frac{\rho^+}{2} \| \pmb{v}^+ \|_2^2 + \| \pmb{u}^{-\mathsf{T}} \pmb{v}^- + \frac{\rho^-}{2} \| \pmb{v}^- \|_2^2, \\ \mathrm{subject\ to} & \tilde{B}^+ = A^+ Z^+, \tilde{B}^- = A^- Z^-, \\ & \pmb{v}^+ = \mathrm{vec}(W^+) - \mathrm{vec}(Z^+), \pmb{v}^- = \mathrm{vec}(W^-) - \mathrm{vec}(Z^-), \\ & W^+ = Z^+, W^- = Z^-, W^+ \geq 0, W^- \geq 0, \\ & \mathrm{diag}(W^+) = 0, \mathrm{diag}(W^-) = 0, \end{split}$$

Algorithm 1 Learning SlimLogR

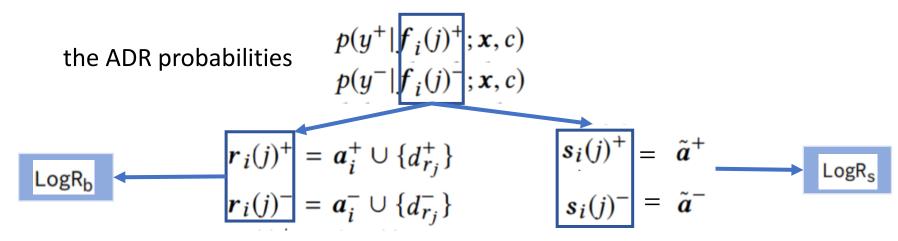
```
1: function SlimLogR(A, \omega, \alpha, \lambda, \beta, \gamma)
                                       \rho^+ = 10, \, \rho^- = 10, \, \boldsymbol{u}_{(0)}^+ = 0, \, \boldsymbol{u}_{(0)}^- = 0, \, k = 0
                                 Z_{(0)}^+ = W_{(0)}^+, Z_{(0)}^- = W_{(0)}^-
                                       learn W_{(0)}^+ and W_{(0)}^- from SLIM (Section 6.3)
                                        learn x_{(0)} and c_{(0)} from LogR (Section 6.2)
                                        while not converge do
                                                       \{W_{(k+1)}^+,\,W_{(k+1)}^-\} := \underset{W^+,\,W^-}{\operatorname{argmin}} \; L(W_{(k)}^+,\,W_{(k)}^-,\,Z_{(k)}^+,\,Z_{(k)}^-,\\ x_{(k)},\;c_{(k)},\;u_{(k)}^+,\;u_{(k)}^-)
                                                       \{Z_{(k+1)}^+, Z_{(k+1)}^-\} \coloneqq \underset{Z^+, Z^-}{\operatorname{argmin}} L(W_{(k+1)}^+, W_{(k+1)}^-, Z_{(k)}^+, Z_{(k)}^-, x_{(k)}^-, x_{(k)
     8:
                                                        \{ \boldsymbol{x}_{(k+1)}, \ c_{(k+1)} \} \coloneqq \underset{\boldsymbol{x}, \ c}{\operatorname{argmin}} \ L(W_{(k+1)}^+, W_{(k+1)}^-, Z_{(k+1)}^+, Z_{(k+1)}^-, \\ \boldsymbol{x}_{(k)}, \ c_{(k)}, \ \boldsymbol{u}_{(k)}^+, \ \boldsymbol{u}_{(k)}^-) 
                                                           u_{(k+1)}^+ = u_{(k)}^+ + \rho^+ (\text{vec}(W_{(k+1)}^+ - Z_{(k+1)}^+))
 10:
                                                          \mathbf{u}_{(k+1)}^- = \mathbf{u}_{(k)}^- + \rho^+ (\text{vec}(W_{(k+1)}^- - Z_{(k+1)}^-))
                                                            k = k + 1
12:
                                        end while
13:
                                        return W_{(k+1)}^+, W_{(k+1)}^-, Z_{(k+1)}^+, Z_{(k+1)}^-, \boldsymbol{x}_{(k+1)} and c_{(k+1)}
14:
15: end function
```

Applying SlimLogR

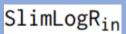
• Step 1: use the learned SLIM component to recommend a list of potential to-avoid/safe drug candidates $\{d_{r_i}\}$;

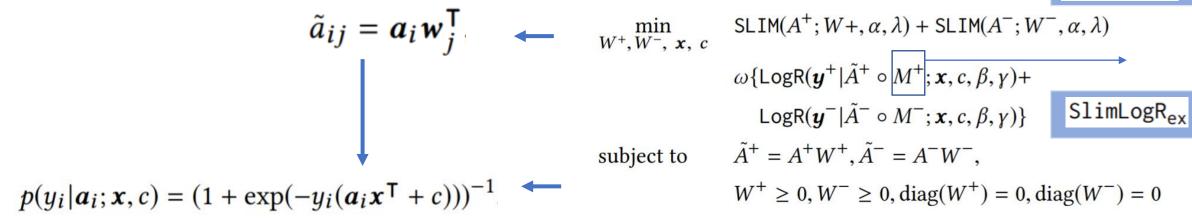
the scores of being potential to-avoid drugs: $\tilde{a}^+ = aW^+$ the scores of being potential safe drugs: $\tilde{a}^- = aW^-$

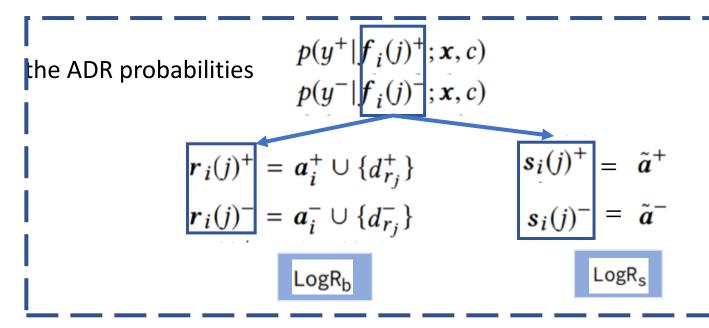
• Step 2: for each recommended candidate d_{r_j} , combine it with a (i.e., $a \cup \{d_{r_j}\}$) and use the learned LogR component to predict the ADR probability of the new prescription.



Overview







A_{pool}	mdl	prd	rect	prect	acct
	ClimborD	LogRs	0.2361	0.2361	0.2958
	SlimLogR _{in}	LogRb	0.2361	0.2361	0.2958
	SlimLogR _{ex}	LogRs	0.2108	0.2108	0.2795
1		- LogR _b	0.2108	0.2108	0.2795
AFAERS	Rand	-	0.0060	0.0118	0.0238
	LogR	-	0.2082	0.2081	0.2430
	SLIM	-	0.2127	0.2126	0.2811
	SLIM+LogR	-	0.2127	0.2126	0.2813

1. Preamble

2. The Proposed Method

Experiments

Datasets:

datasat	atata	A^-			A^+
dataset	stats	A_{-}^{-}	A_0^-	A_0^+	A_{+}^{+}
	#{a}	621,449	1,264	8,986	27,387
	$\#\{d\}$	1,209	417	881	1,201
A_{FAERS}	avgOrd	6.100	2.351	3.588	7.096
	avgFrq	1.761	225.317	13.730	1.402
	avgOR	-	0.546	16.343	-
	#{a}	2,200	1,264	2,464	1,000
	$\#\{d\}$	562	417	692	679
A_*	avgOrd	2.678	2.351	3.809	7.615
	avgFrq	42.082	225.317	20.565	5.520
	avg0R	-	0.546	31.998	-

Experiments

Table 3: Comparison based on the Best $\overline{rec_t}$ (N = 5)

A_{pool}	mdl	prd	rect	prect	acct
*	ClimbooD	LogRs	0.2361	0.2361	0.2958
	SlimLogR _{in}	LogRb	0.2361	0.2361	0.2958
	61 (1 1)	LogRs	0.2108	0.2108	0.2795
4	SlimLogR _{ex}	LogRb	0.2108	0.2108	0.2795
AFAERS	Rand	-	0.0060	0.0118	0.0238
	LogR	-	0.2082	0.2081	0.2430
	SLIM	-	0.2127	0.2126	0.2811
	SLIM+LogR	-	0.2127	0.2126	0.2813
	$SlimLogR_{in}$	LogRs	0.2618	0.2615	0.2815
		LogRb	0.2618	0.2615	0.2815
	63 / - L B	LogRs	0.2479	0.2479	0.2748
Λ	SlimLogR _{ex}	LogRb	0.2479	0.2479	0.2748
A_*	Rand	-	0.0029	0.0059	0.0034
	LogR	-	0.1780	0.1781	0.1032
	SLIM	-	0.2469	0.2467	0.2754
	SLIM+LogR	-	0.2469	0.2467	0.2754

The column "mdl" corresponds to models. The column "prd" corresponds to prediction methods. The best $\overline{rec_t}$ is underlined. The best overall performance is **bold**.

Table 4: Comparison based on the Best $\overline{prec_t}$ (N = 5)

A_{pool}	mdl	prd	rect	prect	acct
	Climi and	LogRs	0.2361	0.2361	0.2958
	SlimLogR _{in}	LogRb	0.2361	0.2361	0.2958
	ClimboaD	LogRs	0.2108	0.2108	0.2795
A	SlimLogR _{ex}	LogRb	0.2108	0.2108	0.2795
AFAERS	Rand	-	0.0060	0.0118	0.0238
	LogR	-	0.2082	0.2081	0.2430
	SLIM		0.2127	0.2128	0.2809
	SLIM+LogR	-	0.2127	0.2128	0.2811
	SlimLogR _{in}	LogRs	0.2613	0.2615	0.2821
		LogRb	0.2613	0.2615	0.2821
	SlimLogR _{ex}	LogRs	0.2477	0.2479	0.2759
A		LogRb	0.2477	0.2479	0.2759
A_*	Rand	-	0.0029	0.0059	0.0034
	LogR	-	0.1780	0.1781	0.1032
	SLIM	-	0.2465	0.2467	0.2754
	SLIM+LogR		0.2465	0.2467	0.2754

The column "mdl" corresponds to models. The column "prd" corresponds to prediction methods. The best prect is underlined. The best overall performance is **bold**.

Table 5: Comparison based on the Best $\overline{acc_t}$ (N = 5)

Apool	mdl	prd	rect	prect	acct
- 0	ClimboaD	LogRs	0.2355	0.2355	0.2969
	SlimLogRin	LogRb	0.2355	0.2355	0.2969
	Climi och	LogRs	0.2099	0.2099	0.2797
	SlimLogR _{ex}	LogRb	0.2108	0.2108	0.2797
AFAERS	Rand	-	0.0060	0.0118	0.0238
	LogR	-	0.2036	0.2039	0.246
	SLIM	-	0.2019	0.2020	0.283
	SLIM+LogR	-	0.2019	0.2020	0.282
	$SlimLogR_{in}$	LogRs	0.2589	0.2592	0.2832
		LogRb	0.2589	0.2592	0.2832
	63 i I D	LogRs	0.2477	0.2479	0.276
A	SlimLogR _{ex}	LogRb	0.2477	0.2479	0.276
A_*	Rand	-	0.0029	0.0059	0.0034
	LogR	-	0.1780	0.1781	0.1032
	SLIM	-	0.2434	0.2432	0.2776
	SLIM+LogR	-	0.2434	0.2432	0.2776

The column "mdl" corresponds to models. The column "prd" corresponds to prediction methods. The best acct is underlined. The best overall performance is **bold**.

Parameter Study

$$\begin{split} \min_{W^+,W^-,~\boldsymbol{x},~\boldsymbol{c}} & \quad \text{SLIM}(A^+;W^+,\alpha,\lambda) + \text{SLIM}(A^-;W^-,\alpha,\lambda) \\ & \quad \omega \{ \text{LogR}(\boldsymbol{y}^+|\tilde{A}^+ \circ M^+;\boldsymbol{x},\boldsymbol{c},\beta,\gamma) + \\ & \quad \text{LogR}(\boldsymbol{y}^-|\tilde{A}^- \circ M^-;\boldsymbol{x},\boldsymbol{c},\beta,\gamma) \} \\ & \quad \text{subject to} & \quad \tilde{A}^+ = A^+W^+,\tilde{A}^- = A^-W^-, \\ & \quad W^+ \geq 0, W^- \geq 0, \text{diag}(W^+) = 0, \text{diag}(W^-) = 0 \end{split}$$

Table 6: $\overline{\text{rec}_t}$ of $\overline{\text{SlimLogR}_{in}}$ + $\overline{\text{LogR}_s}$ (N=5) Table 7: $\overline{\text{prec}_t}$ of $\overline{\text{SlimLogR}_{in}}$ + $\overline{\text{LogR}_s}$ (N=5) Table 8: $\overline{\text{acc}_t}$ of $\overline{\text{SlimLogR}_{in}}$ + $\overline{\text{LogR}_s}$ (N=5)

$\omega \setminus \alpha$	100	50	20	10	5
20	0.2324	0.2344	0.2346	0.2349	0.2338
10	0.2315	0.2352	0.2349	0.2358	0.2344
5	0.2318	0.2355	0.2361	0.2361	0.2341
1	0.2315	0.2327	0.2321	0.2304	0.2290

 A_{pool} is A_{FAERS} . The best performance is **bold**.

100	50	20	10	5
0.2325	0.2343	0.2347	0.2349	0.2339
0.2315	0.2351	0.2351	0.2359	0.2345
0.2321	0.2355	0.2361	0.2359	0.2341
0.2315	0.2327	0.2321	0.2305	0.2290
	0.2325 0.2315 0.2321	0.2325 0.2343 0.2315 0.2351 0.2321 0.2355	0.2325 0.2343 0.2347 0.2315 0.2351 0.2351 0.2321 0.2355 0.2361	100 50 20 10 0.2325 0.2343 0.2347 0.2349 0.2315 0.2351 0.2351 0.2359 0.2321 0.2355 0.2361 0.2359 0.2315 0.2327 0.2321 0.2305

 A_{pool} is A_{FAERS} . The best performance is **bold**.

$\omega \setminus \alpha$	100	50	20	10	5
20	0.2951	0.2919	0.2890	0.2895	0.2874
10	0.2943	0.2967	0.2938	0.2915	0.2891
5	0.2955	0.2969	0.2958	0.2951	0.2924
1	0.2938	0.2953	0.2936	0.2919	0.2893

 A_{pool} is A_{FAERS} . The best performance is **bold**.

Top-N Performance

Table 9: Top-N Performance of SlimLogRin with LogRs

N	A_{FAERS}			A_*			
IV	rect	prect	acct	rect	prect	acct	
_	0.2358	0.2359	0.2915	0.2613	0.2615	0.2821	
5	0.2324	0.2325	0.2929	0.2589	0.2592	0.2832	
10	0.2497	0.2497	0.3076	0.3469	0.3467	0.3419	
10	0.2467	0.2467	0.3110	0.3434	0.3432	0.3439	
20	0.2885	0.2887	0.2762	0.4403	0.4407	0.4016	
20	0.2694	0.2695	0.3289	0.4265	0.4273	0.4135	

Column of "N" represents the number of recommended drugs. Columns of " $A_{\rm FAERS}$ " and " A_* " represent that " $A_{\rm FAERS}$ " and " A_* " are used as A_{pool} , respectively. The **bold** performance is the best under the corresponding metrics in each column.