

## معرفی

شبکه های بی‌زین شامل  $n$  نود  $V = \{v_1, \dots, v_n\}$  که هر گره متغیر تصادفی است و یک توزیع احتمالی شرطی  $P(v_i | \text{par}(v_i))$  برای هر متغیر تصادفی  $v_i$  است که در آن  $\text{par}(v_i)$  احتمال مجموعه والدین  $v_i$  در  $G$  است. یال های  $G$  از یک‌گره به دیگری نشان‌دهنده وابستگی‌های مستقیم هستند و ساختار کامل  $G$  تمام وابستگی‌های شرطی بین متغیرها را کد می‌کند.

روش MINOBSx براساس روش MINOBS که یک الگوریتم جستجوی محلی مبتنی بر جستجوی محلی برای یادگیری ساختار BN که توسط Lee و vanbeek (2017) ساخته شده است، استوار است و قادر به رسیدگی به انواع زیر از محدودیت‌های موجود در شبکه های بی‌زین است:

- (1) وجود یک یال: ادعا میکند که یک یال جهتدار وجود دارد. همچنین کاربر می‌تواند مشخص کند که یک یال بدون جهت  $x-y$  در صورتی وجود دارد که جهت شناخته نشود.
- (2) نبود یال: ادعا می‌کند که یال  $x-y$  وجود ندارد.
- (3) قید و شرط: ادعا می‌کند  $x$  قبل از  $y$  در برخی مرتب کردن گره‌ها در شبکه BN دیده می‌شود.
- (4) محدودیت اجدادی (مثبت): ادعا می‌کند که مسیری جهتدار بین  $x$  و  $y$  را وجود دارد.

به چنین محدودیت‌هایی با عنوان محدودیت‌های جانبی اشاره می‌شود.

در مقابل روش CaMML که کشف تصادفی از طریق mml است و توسط Korb و Nicholson (۲۰۱۰) ارائه شده از جستجوی تصادفی براساس زنجیره مارکوف مونت کارلو استفاده می‌کند. این روش قادر به کنترل تمام محدودیت‌های (۱) - (۴) به جز محدودیت نبود یال است. روش CaMML برای مقایسه انتخاب می‌شود چون به طور گسترده در زمینه‌های کاربردی استفاده می‌شود.

در اینجا نتایج روش MINOBSx و روش CaMML در آزمایشات یکسان مقایسه می‌شود.

## داده ها

در ادامه نتایج روش MINOBSx با روش CaMML را بر روی چندین داده ی (Dataset) مختلف با شبکه های کوچک و متوسط انجام داده و نتایج هریک را به طور جداگانه بررسی شده اند.

داده های مورد آزمایش همگی از مخزن شبکه Bayesian انتخاب شده، که از <http://www.bnlearn.com/bnrepository> برگرفته و جمع آوری شده اند.

در مجموع 11 داده شامل 5 داده با شبکه کوچک و 6 داده با شبکه ی متوسط در هر دو روش مورد مقایسه قرار گرفته اند. مشخصات هر داده در جدول نتایج ذکر شده است.

## نمونه گیری

هریک از داده ها در مقیاس مشاهدات  $N$  به طور تصادفی و توسط روش زیر نمونه گیری شده:

- i. انتخاب  $P\%$  از یال های جهتدار با احتمال  $\frac{1}{2}$  و گزارش  $x \rightarrow y$  یا  $x-y$
- ii. انتخاب  $P\%$  از جفت های  $(x, y)$  که بین آنها یالی نیست و گزارش  $x \rightarrow y$
- iii. انتخاب  $P\%$  از جفت های  $(x, y)$  که یک مسیر بین آنها یالی هست و گزارش  $x \rightarrow y$
- iv. انتخاب یک مرتب سازی توپولوژیک و  $P\%$  تمام  $(n_2)$  محدودیت مرتب سازی و گزارش  $x < y$

لذا در آزمایشات از نمونه های یکسان برای هر دو روش در درصد های مختلف استفاده شده که در جدول نتایج مشخص است. در مراحل کار هر یک از داده ها به طور جداگانه نمونه گیری شده اند.

به طور تصادفی ۶ مجموعه داده کوچک و ۶ مجموعه داده بزرگ از توزیع های احتمالی مشترک شبکه های واقعی این داده ها نمونه برداری شده است.

این آزمایشات یک بار تنها تحت محدودیت های اجدادی و بار دیگر تحت محدودیت های مختلف اعمال شد.

## برنامه ها و ابزارها

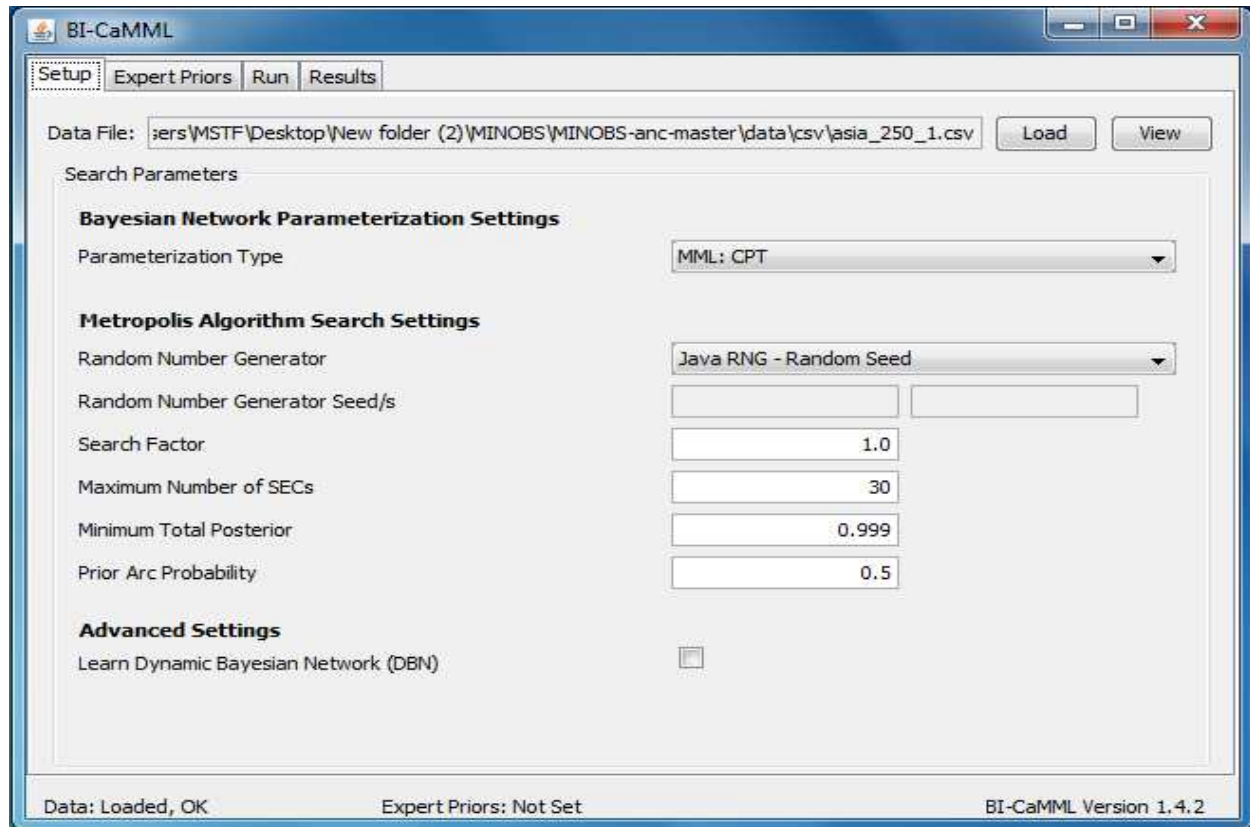
روش MINOBSx برای اجرا نیاز به امتیاز مجموعه والدین با استفاده از فرمول کوکران و روش GOBNILP دارد که از طریق <https://www.cs.york.ac.uk/aig/sw/gobnilp/> قابل دستیابی است. ما این مجموعه امتیازها را با استفاده از ابزار GOBNILP تا 3 والد محاسبه شده است. این ابزار تحت دستور gobnilp data\_1\_3.scores مجموعه امتیازها را تا 3 والد محاسبه می کند.

زمان اجرای GOBNILP در زمان اجرای MINOBSx و نتایج آن محاسبه شده است.

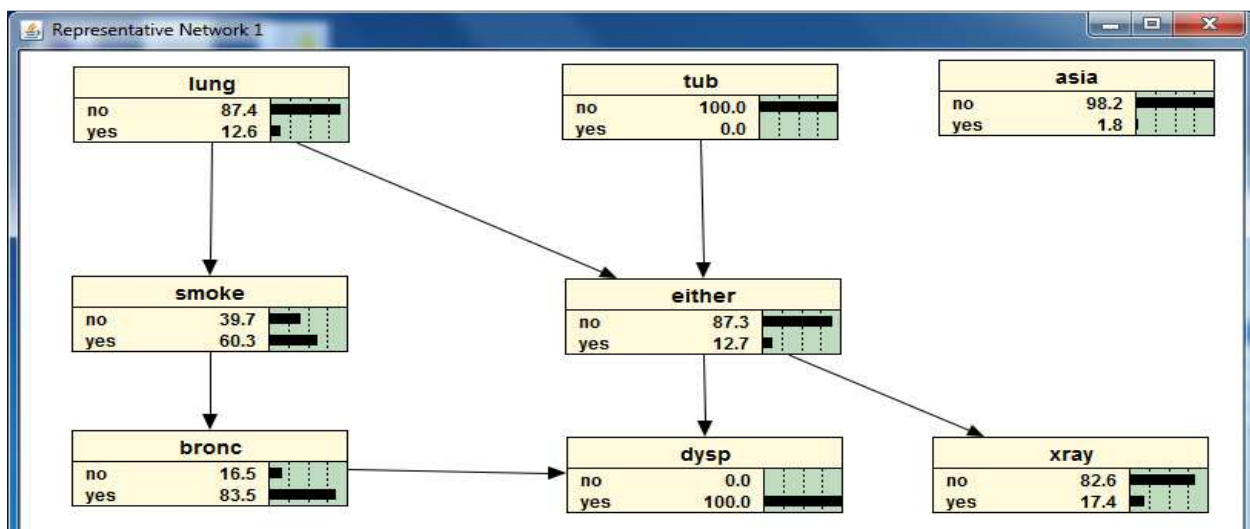
همچنین روش MINOBSx بر روی C++ با استفاده از کامپایلر gcc و g++ که در <https://gcc.gnu.org> اراعه شده کار میکند و برای استفاده از آن در ویندوز نیاز به نصب MinGW از طریق [www.mingw.org](http://www.mingw.org) است. اگرچه اجرای آن در ویندوز نیازمند کتابخانه های محلی دیگری نظیر boost که از طریق <https://www.boost.org/> قابل دسترسی است، میباشد ولی ابزار MINOBSx تحت Mac OSx انجام شده و از طریق <https://github.com/acliuw/MINOBS-dp> در دسترس است.

ابزار روش CaMML نیز تحت جاوا کار میکند و برای اجرای آن نیاز به نصب نسخه ی Java 6 و بالاتر است که از طریق <https://www.java.com/en/download/> می توان به جدیدترین نسخه جاوا دسترسی داشت.

برای اجرای CaMML ابتدا نیاز به اجرای فایل `camml.bat` در سیستم عامل های ویندوزی و یا فایل `camml.sh` در سیستم عامل های لینوکسی و Mac Os است. سپس با استفاده از فایل `camml_gui.bat` و `camml_gui.sh` می توان وارد رابط کاربری CaMML شد.



تصویر 1: محیط CaMML



تصویر 2: مصور سازی شبکه تولید شده توسط CaMML

## تنظیمات

در تمامی آزمایش های صورت گرفته ابزار CaMML تحت تنظیمات پیش فرض خود اجرا شد. هنگام مقایسه روش های MINOBSx و CaMML محدودیت های نبود یال  $x \rightarrow y$  حذف شوند چرا که CaMML این محدودیت ها را قبول نمی کرد.

برای هر مجموعه داده از آزمایش ها،  $P\%$  ثابت را انتخاب شده و ۵ مجموعه از محدودیت را در آن درصد از شبکه واقعی نمونه گیری شده است.

هریک از نمونه ها با ابزار روش MINOBSx و CaMML مورد آزمایش قرار گرفته و نتایج به صورت جدول در آمده است.

## نتایج

نتایج نشان می دهد که تمامی محدودیت ها مربوط به CaMML با سطح اعتماد بالایی در اکثر شبکه های کوچک مشخص می شوند، در حالیکه CaMML در محدودیت های اجدادی با کاهش عملکرد روبرو است. همچنین در شبکه های متوسط نیز کارایی در بسیاری از موارد نامناسب بوده و با افزایش تعداد متغیرها کارایی CaMML به شدت کاهش می یابد.

همانگونه که از جدول نتایج 2 پیداست، در انجام این آزمایشات زمان بیشتری به CaMML اختصاص داده میشود. با این وجود این روش در شبکه های توسط حتی گاهی به راه حل های مناسبی دست نمی یابد.

در مقابل روش MINOBSx در تمامی موارد شامل شبکه های بزرگ و کوچک راه حل های بهتر و قابل اطمینان بیشتری را نسبت به روش CaMML ارائه می دهد. این روش حتی هنگامی که تعداد متغیر ها افزایش چشمگیری پیدا می کنند راه حل های با اطمینان بسیار بالا را در زمان بسیار کمتری ارائه می دهد.

instance	N	%	MINOBSx			CaMML		
			% feasible	% sat	t	% feasible	% sat	t
<b>cancer</b> 5 variables 10 parameters	250	10 / 5	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	4.0 / 3.9
		25 / 10	100 / 100	100 / 100	0.1 / 0.0	100 / 96.7	100 / 98.9	4.1 / 4.0
		50 / 15	100 / 100	100 / 100	0.1 / 0.1	96.7 / 100	99.2 / 100	4.2 / 4.3
		100 / 20	100 / 100	100 / 100	0.1 / 0.0	100 / 100	100 / 100	4.2 / 4.2
	1000	10 / 5	100 / 100	100 / 100	0.1 / 0.1	96.7 / 100	96.7 / 100	3.8 / 3.6
		25 / 10	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	3.8 / 3.8
		50 / 15	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	4.2 / 4.0
		100 / 20	100 / 100	100 / 100	0.0 / 0.0	100 / 100	100 / 100	4.1 / 3.8
<b>earthquake</b> 5 variables 10 parameters	250	10 / 5	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	4.2 / 4.3
		25 / 10	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	4.3 / 4.5
		50 / 15	100 / 100	100 / 100	0.1 / 0.0	100 / 100	100 / 100	4.4 / 4.1
		100 / 20	100 / 100	100 / 100	0.1 / 0.0	100 / 100	100 / 100	4.5 / 4.1
	1000	10 / 5	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	3.8 / 3.9
		25 / 10	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	3.8 / 3.9
		50 / 15	100 / 100	100 / 100	0.1 / 0.0	100 / 100	100 / 100	4.0 / 3.8
		100 / 20	100 / 100	100 / 100	0.0 / 0.0	100 / 100	100 / 100	4.3 / 4.0
<b>survey</b> 6 variables 21 parameters	250	10 / 5	100 / 100	100 / 100	0.2 / 0.1	100 / 100	100 / 100	4.7 / 4.6
		25 / 10	100 / 100	100 / 100	0.3 / 0.1	100 / 100	100 / 100	5.0 / 4.6
		50 / 15	100 / 100	100 / 100	0.2 / 0.1	93.3 / 96.7	99.0 / 99.4	4.9 / 4.4
		100 / 20	100 / 100	100 / 100	0.1 / 0.1	100 / 100	100 / 100	4.9 / 4.4
	1000	10 / 5	100 / 100	100 / 100	0.1 / 0.1	96.7 / 96.7	98.3 / 98.9	4.3 / 4.2
		25 / 10	100 / 100	100 / 100	0.2 / 0.1	96.7 / 96.7	99.17 / 99.3	4.6 / 4.4
		50 / 15	100 / 100	100 / 100	0.1 / 0.1	93.3 / 93.3	99.0 / 98.9	4.5 / 4.2
		100 / 20	100 / 100	100 / 100	0.1 / 0.1	100 / 96.7	100 / 99.6	4.7 / 4.1
<b>asia</b> 8 variables 18 parameters	250	10 / 5	100 / 100	100 / 100	1.1 / 0.5	100 / 100	100 / 100	5.8 / 5.4
		25 / 10	100 / 100	100 / 100	1.3 / 0.5	100 / 100	100 / 100	6.3 / 5.5
		50 / 15	100 / 100	100 / 100	0.9 / 0.3	100 / 100	100 / 100	6.8 / 5.9
		100 / 20	100 / 100	100 / 100	0.5 / 0.2	100 / 100	100 / 100	8.3 / 5.7
	1000	10 / 5	100 / 100	100 / 100	0.9 / 0.4	100 / 100	100 / 100	5.5 / 5.0
		25 / 10	100 / 100	100 / 100	1.1 / 0.4	100 / 100	100 / 100	6.0 / 5.3
		50 / 15	100 / 100	100 / 100	0.7 / 0.3	100 / 96.7	100 / 99.7	6.3 / 5.6
		100 / 20	100 / 100	100 / 100	0.4 / 0.2	100 / 100	100 / 100	7.3 / 5.6
<b>sachs</b> 11 variables 178 parameters	250	10 / 5	100 / 100	100 / 100	2.0 / 1.0	100 / 100	100 / 100	7.0 / 6.2
		25 / 10	100 / 100	100 / 100	3.5 / 0.8	100 / 93.3	100 / 99.4	7.4 / 6.1
		50 / 15	100 / 100	100 / 100	4.3 / 0.8	100 / 100	100 / 100	8.4 / 6.4
		100 / 20	100 / 100	100 / 100	2.3 / 0.7	16.7 / 100	96.2 / 100	10.3 / 6.1
	1000	10 / 5	100 / 100	100 / 100	1.6 / 1.0	80.0 / 83.3	93.3 / 97.2	6.2 / 5.9
		25 / 10	100 / 100	100 / 100	2.5 / 0.8	30.0 / 60.0	87.8 / 94.8	7.2 / 6.0
		50 / 15	100 / 100	100 / 100	2.9 / 0.6	23.3 / 50.0	90.0 / 96.9	8.1 / 6.2
		100 / 20	100 / 100	100 / 100	2.2 / 0.6	0 / 70.0	95.5 / 98.5	9.9 / 6.3

#### جدول نتایج 1: نتایج در شبکه های کوچک

- نتایج عملکرد برای محدودیت های اجدادی (شماره اول در هر جفت) و با محدودیت های مختلف (شماره دوم در هر جفت)
- %: ثابت درصد مورد استفاده برای محدود کردن نمونه ها، درصد مواردی است که راه حل آن است رضایت همه محدودیت های تحمیل شده
- % feasible: درصد قابل اجرا، درصد مواردی که راه حل آن توسط رضایت همه محدودیت ها تحمیل شده
- % sat: درصد محدودیت های برآورده شده توسط محدودیت های تحمیل شده
- t: زمان در حال اجرا و مورد نیاز برنامه
- N: تعداد مشاهدات
- سلول های برجسته نمایانگر برآورده نشدن تمامی محدودیت ها



instance	N	%	MINOBSx			CaMML		
			% feasible	% sat	t	% feasible	% sat	t
child 20 variables 230 parameters	500	10 / 5	100 / 100	100 / 100	11.2 / 7.2	100 / 100	100 / 100	66.8 / 58.4
		25 / 10	100 / 100	100 / 100	29.9 / 7.3	100 / 100	100 / 100	86.0 / 62.4
		50 / 15	100 / 100	100 / 100	37.8 / 4.1	100 / 100	100 / 100	120.6 / 64.6
		100 / 20	100 / 100	100 / 100	26.1 / 5.1	66.7 / 100	99.5 / 100	197.6 / 64.6
	2000	10 / 5	100 / 100	100 / 100	9.4 / 6.7	100 / 100	100 / 100	61.0 / 55.1
		25 / 10	100 / 100	100 / 100	19.1 / 6.0	100 / 100	100 / 100	83.7 / 59.7
		50 / 15	100 / 100	100 / 100	25.7 / 3.0	100 / 100	100 / 100	115.6 / 60.6
		100 / 20	100 / 100	100 / 100	20.4 / 3.2	100 / 100	100 / 100	180.2 / 60.1
insurance 27 variables 984 parameters	500	10 / 5	100 / 100	100 / 100	180.5 / 104.9	50.0 / 70.0	95.9 / 99.0	439.5 / 325.3
		25 / 10	100 / 100	100 / 100	318.9 / 56.5	56.7 / 50.0	98.0 / 98.9	723.9 / 385.0
		50 / 15	100 / 100	100 / 100	328.7 / 52.8	30.0 / 33.3	98.5 / 98.1	1165.6 / 485.2
		100 / 20	100 / 100	100 / 100	292.5 / 37.4	0 / 53.3	98.1 / 99.6	2052.3 / 571.7
	2000	10 / 5	100 / 100	100 / 100	124.0 / 88.3	0 / 53.3	78.8 / 98.1	438.5 / 309.8
		25 / 10	100 / 100	100 / 100	236.3 / 49.6	16.7 / 23.3	92.4 / 96.8	748.6 / 393.2
		50 / 15	100 / 100	100 / 100	251.5 / 48.2	3.3 / 10.0	94.4 / 96.6	1175.2 / 487.9
		100 / 20	100 / 100	100 / 100	233.1 / 33.4	0 / 10.0	95.6 / 98.8	1956.8 / 571.1
water 32 variables 10083 parameters	1000	10 / 5	100 / 100	100 / 100	436.5 / 236.0	86.7 / 100	99.3 / 100	2740.5 / 2138.4
		25 / 10	100 / 100	100 / 100	987.1 / 265.2	80.0 / 43.3	99.5 / 99.0	4934.6 / 2782.8
		50 / 15	100 / 100	100 / 100	1083.2 / 166.0	46.7 / 30.0	99.2 / 98.9	8499.3 / 3239.0
		100 / 20	100 / 100	100 / 100	1015.1 / 87.9	16.7 / 30.0	99.3 / 99.1	15299.3 / 3967.5
	4000	10 / 5	100 / 100	100 / 100	358.9 / 232.5	30 / 63.3	92.2 / 98.9	2912.2 / 2153.3
		25 / 10	100 / 100	100 / 100	788.7 / 169.2	10.0 / 10.0	94.5 / 98.4	5081.3 / 2725.6
		50 / 15	100 / 100	100 / 100	901.6 / 139.2	6.7 / 3.3	95.3 / 97.7	8325.3 / 3377.0
		100 / 20	100 / 100	100 / 100	934.6 / 62.0	0 / 0	98.5 / 97.6	15341.1 / 4210.0
mildew 35 variables 540150 parameters	8000	10 / 5	100 / 100	100 / 100	193.8 / 754.7	0 / 0	26.9 / 75.6	2134.2 / 1442.0
		25 / 10	100 / 100	100 / 100	632.1 / 334.2	0 / 0	29.4 / 86.5	3906.5 / 1835.7
		50 / 15	100 / 100	100 / 100	943.3 / 321.0	0 / 0	40.3 / 88.2	7336.1 / 2215.0
		100 / 20	100 / 100	100 / 100	1085.4 / 309.7	0 / 0	54.4 / 89.8	12649.3 / 2572.5
	32000	10 / 5	100 / 100	100 / 100	181.6 / 353.0	0 / 0	19.4 / 77.8	2553.9 / 1664.6
		25 / 10	100 / 100	100 / 100	471.6 / 266.2	0 / 0	19.4 / 87.7	4763.4 / 1983.3
		50 / 15	100 / 100	100 / 100	702.2 / 210.2	0 / 0	24.0 / 88.0	8546.6 / 2462.6
		100 / 20	100 / 100	100 / 100	761.0 / 264.8	0 / 0	28.2 / 92.3	17148.3 / 2400.38
alarm 37 variables 509 parameters	1000	10 / 5	100 / 100	100 / 100	954.0 / 160.9	86.7 / 96.7	99.4 / 99.9	2423.0 / 1736.6
		25 / 10	100 / 100	100 / 100	1893.2 / 316.9	76.7 / 90.0	99.5 / 99.9	4070.2 / 2019.4
		50 / 15	100 / 100	100 / 100	1627.7 / 185.1	73.3 / 100	99.5 / 100	6753.0 / 2263.7
		100 / 20	100 / 100	100 / 100	1207.2 / 147.3	83.3 / 100	99.9 / 100	12377.2 / 2692.7
	4000	10 / 5	100 / 100	100 / 100	602.3 / 115.3	70.0 / 83.3	97.2 / 99.7	2314.6 / 1705.3
		25 / 10	100 / 100	100 / 100	1336.8 / 224.7	86.7 / 86.7	99.3 / 99.9	3807.8 / 1988.3
		50 / 15	100 / 100	100 / 100	1438.2 / 133.3	90.0 / 100	99.8 / 100	6268.1 / 2163.7
		100 / 20	100 / 100	100 / 100	1195.6 / 116.5	50.0 / 80.0	99.0 / 99.8	11169.0 / 2581.8
barley 48 variables 114005 parameters	2000	10 / 5	100 / 100	100 / 100	2321.4 / 5866.8	0 / 0	70.3 / 89.3	19824.8 / 11666.1
		25 / 10	100 / 100	100 / 100	4228.9 / 2941.3	0 / 0	74.4 / 93.6	37034.3 / 14864.9
		50 / 15	100 / 100	100 / 100	7163.0 / 3518.9	0 / 0	73.9 / 95.0	70433.0 / 20339.2
		100 / 20	100 / 100	100 / 100	7246.6 / 1806.3	0 / 0	80.9 / 96.3	114036.7 / 22366.9
	8000	10 / 5	100 / 100	100 / 100	4761.1 / 6032.7	0 / 0	44.0 / 82.8	18092.6 / 10759.6
		25 / 10	100 / 100	100 / 100	4063.8 / 3620.0	0 / 0	52.5 / 91.8	35308.3 / 14477.5
		50 / 15	100 / 100	100 / 100	5137.8 / 3022.8	0 / 0	54.0 / 95.1	64893.5 / 18142.0
		100 / 20	100 / 100	100 / 100	5675.6 / 1638.8	0 / 0	53.4 / 95.1	111338.4 / 21184.4

جدول نتایج 2: نتایج در شبکه های متوسط

همچنین کیفیت شبکه‌های یادگیری (learn) شده توسط روش MINOBSx را در برابر شبکه‌های واقعی در جدول های امتیاز مقایسه شده اند. کیفیت شبکه‌های یادگیری (learn) شده توسط روش CaMML را در برابر شبکه‌های واقعی مقایسه نمی شود چرا که این دو روش از توابع امتیازدهی مختلف استفاده می‌کنند، که می‌تواند تاثیر زیادی بر روی نتایج داشته باشد.

instance	N	%	Missing	Extra	Reversed	SID	Score (BDeu)
cancer 5 variables 10 parameters	250	0*	1.8	0.2	0.8	9.2	0%
		10 / 5	1.4 / 1.0	0.3 / 0.2	0.5 / 0.5	6.1 / 4.8	0.1% / 0.1%
		25 / 10	0.8 / 1.1	0.3 / 0.2	0.3 / 0.5	3.6 / 5.2	0.2% / 0.2%
		50 / 15	0.7 / 0.6	0.5 / 0.3	0.3 / 0.3	3.3 / 3.3	0.3% / 0.3%
		100 / 20	0.2 / 0.8	0.3 / 0.2	0.0 / 0.3	0.5 / 4.0	0.4% / 0.2%
	1000	0*	0.8	0.0	0.5	4.3	0%
		10 / 5	0.7 / 0.6	0.1 / 0.0	0.5 / 0.4	4 / 3.1	0.0% / 0.0%
		25 / 10	0.5 / 0.5	0.2 / 0.1	0.3 / 0.5	2.6 / 3.4	0.1% / 0.0%
		50 / 15	0.3 / 0.3	0.1 / 0.1	0.1 / 0.1	1.0 / 1.4	0.1% / 0.1%
		100 / 20	0.0 / 0.4	0.0 / 0.0	0.0 / 0.3	0 / 2.3	0.1% / 0.1%
earthquake 5 variables 10 parameters	250	0*	0.8	1.0	1.5	8.7	0%
		10 / 5	0.6 / 0.4	1.0 / 1.0	0.6 / 0.3	4.2 / 2.7	0.2% / 0.3%
		25 / 10	0.4 / 0.6	0.9 / 1.0	0.2 / 0.5	1.7 / 3.9	0.3% / 0.3%
		50 / 15	0.4 / 0.3	1.2 / 1.1	0.3 / 0.2	1.9 / 2.3	0.5% / 0.7%
		100 / 20	0.2 / 0.4	1.2 / 1.2	0.0 / 0.2	0.2 / 2.1	0.7% / 0.6%
	1000	0*	0.0	0.5	0.3	1.5	0%
		10 / 5	0.0 / 0.0	0.4 / 0.3	0.1 / 0.0	0.6 / 0	0.0% / 0.0%
		25 / 10	0.0 / 0.0	0.3 / 0.3	0.0 / 0.0	0 / 0.3	0.0% / 0.0%
		50 / 15	0.0 / 0.0	0.3 / 0.5	0.0 / 0.1	0 / 0.9	0.0% / 0.1%
		100 / 20	0.0 / 0.0	0.3 / 0.3	0.0 / 0.0	0 / 0	0.0% / 0.1%
survey 6 variables 21 parameters	250	0*	4.5	0.0	0.5	15.5	0%
		10 / 5	3.6 / 3.1	0.8 / 0.4	0.6 / 0.6	13.1 / 12.2	0.4% / 0.4%
		25 / 10	2.8 / 2.8	1.4 / 0.2	0.5 / 0.1	11.7 / 8.9	0.7% / 0.5%
		50 / 15	2.4 / 2.8	1.5 / 0.6	0.5 / 0.1	9.1 / 9.7	1.0% / 0.7%
		100 / 20	2.5 / 2.1	2.0 / 0.7	0.0 / 0.3	6.2 / 8.0	1.2% / 1.0%
	1000	0*	3.5	0.0	1.3	16.3	0%
		10 / 5	2.8 / 2.3	1.0 / 0.4	1.3 / 1.1	15.9 / 13.5	0.1% / 0.1%
		25 / 10	2.7 / 2.5	1.1 / 0.3	0.3 / 0.3	10.1 / 8.3	0.1% / 0.1%
		50 / 15	2.6 / 2.5	1.4 / 0.6	0.2 / 0.2	8.2 / 8.7	0.2% / 0.2%
		100 / 20	2.7 / 1.7	1.8 / 0.5	0.0 / 0.4	6.8 / 8.5	0.2% / 0.3%
asia 8 variables 18 parameters	250	0*	1.5	1.7	1.0	12.2	0%
		10 / 5	1.4 / 1.4	1.6 / 1.6	0.7 / 0.9	10.3 / 11.1	0.0% / 0.0%
		25 / 10	1.2 / 1.4	1.8 / 1.6	0.5 / 0.6	7.2 / 8.9	0.1% / 0.1%
		50 / 15	1.0 / 0.8	2.0 / 1.4	0.3 / 0.4	4.3 / 4.4	0.2% / 0.3%
		100 / 20	0.5 / 0.7	1.7 / 1.1	0.0 / 0.2	1.8 / 3.9	0.3% / 0.3%
	1000	0*	0.8	0.3	1	9.0	0%
		10 / 5	0.7 / 0.8	0.4 / 0.4	0.9 / 1.1	7.3 / 9.0	0.0% / 0.0%
		25 / 10	0.4 / 0.5	0.5 / 0.3	0.4 / 0.8	4.2 / 6.4	0.0% / 0.0%
		50 / 15	0.2 / 0.3	0.4 / 0.4	0.0 / 0.5	0.4 / 3.6	0.0% / 0.0%
		100 / 20	0.0 / 0.2	0.3 / 0.3	0.0 / 0.4	0 / 3.1	0.0% / 0.0%
sachs 11 variables 178 parameters	250	0*	6.5	1.0	5.3	48.0	0%
		10 / 5	6.0 / 5.0	0.8 / 0.5	3.5 / 3.4	36.2 / 34.7	0.2% / 0.4%
		25 / 10	5.5 / 4.5	0.6 / 0.3	2.6 / 1.7	33.4 / 24.8	0.3% / 0.6%
		50 / 15	4.9 / 3.9	0.3 / 0.2	1.4 / 1.5	24.5 / 22.0	0.5% / 0.9%
		100 / 20	3.2 / 3.5	0.0 / 0.0	0.0 / 0.8	15 / 16.2	0.9% / 0.8%
	1000	0*	1.5	0.0	6.0	30.5	0%
		10 / 5	1.3 / 1.1	0.0 / 0.0	4.1 / 3.7	22.2 / 20.7	0.0% / 0.0%
		25 / 10	1.1 / 0.5	0.0 / 0.0	3.4 / 1.8	20.7 / 13.8	0.0% / 0.1%
		50 / 15	0.9 / 0.8	0.0 / 0.0	2.6 / 2.2	15.9 / 13.6	0.1% / 0.1%
		100 / 20	0.5 / 0.5	0.0 / 0.0	0.0 / 1.4	1.5 / 10.4	0.1% / 0.1%

#### جدول امتیاز 1: امتیازات در شبکه های کوچک

تعداد یال های گمشده، اضافی و معکوس (missing/extra/reverse) و فاصله مداخله سازهای (SID) بین شبکه‌های واقعیت زمینی و شبکه یادگیری شده در جدول های امتیاز گزارش شده اند. همچنین درصد اختلاف بین نمره BDeu شبکه آموخته شده و خط پایه (ردیف 0 از %) نمره BDeu از شبکه بهینه هنگامی که هیچ یک از محدودیت ها تحمیل نمی شوند، گزارش شده اند.



از این نتایج، به نظر می‌رسد که به ویژه با افزایش تعداد متغیرها، روش MINOBSx بسیار قویتر از روش CaMML است. حتی وقتی روش CaMML به طور قابل‌توجهی زمان بیشتری را نسبت به زمان روش MINOBSx صرف می‌کند، اما اغلب قادر به یافتن راحل‌های پذیرفتنی نیست. همچنین درصد اختلاف بین امتیازات BDeu در مساله همراه با محدودیت و مساله بدون محدودیت کوچک است و هرگز بیش از 4% نیست. از این رو در بدترین حالت، جواب‌های تولید شده توسط MINOBSx در محدوده خطای 4% قرار دارند.

instance	N	%	Missing	Extra	Reversed	SID	Score (BDeu)
child 20 variables 230 parameters	500	0*	5.3	1.0	3.0	115.7	0%
		10 / 5	4.8 / 4.8	1.1 / 0.8	1.9 / 1.3	91.4 / 82.6	0.1% / 0.2%
		25 / 10	4.7 / 4.3	1.9 / 1.6	1.6 / 1.8	93.9 / 79.4	0.2% / 0.3%
		50 / 15	3.9 / 4.2	1.4 / 1.2	1.7 / 0.7	76.1 / 69.9	0.3% / 0.4%
		100 / 20	2.2 / 3.6	2.2 / 0.8	0.0 / 0.3	35.8 / 57.6	0.9% / 0.4%
	2000	0*	1.7	0.2	3.5	79.2	0%
		10 / 5	1.5 / 1.3	0.2 / 0.2	0.5 / 0.1	26.0 / 20.8	0.0% / 0.0%
		25 / 10	0.7 / 1.0	0.3 / 0.3	0.3 / 0.5	12.1 / 18.0	0.0% / 0.0%
		50 / 15	0.7 / 1.0	0.3 / 0.3	0.5 / 0.0	12.2 / 16.2	0.0% / 0.0%
		100 / 20	0.2 / 1.0	0.3 / 0.1	0.0 / 0.0	2.5 / 16.6	0.1% / 0.0%
insurance 27 variables 984 parameters	500	0*	18.8	4.0	2.8	348.2	0%
		10 / 5	18.4 / 17.3	5.5 / 5.2	2.5 / 2.6	345.1 / 341.4	0.2% / 0.4%
		25 / 10	17.6 / 17.0	6.5 / 4.9	2.5 / 1.8	326.8 / 315.8	0.5% / 1.3%
		50 / 15	17.2 / 14.4	6.9 / 5.7	2.0 / 1.2	321.6 / 276.4	0.6% / 1.4%
		100 / 20	17.3 / 12.9	10.8 / 5.8	0.0 / 1.4	282.0 / 261.8	1.6% / 2.3%
	2000	0*	9.8	1.7	3.7	264.7	0%
		10 / 5	9.9 / 10.0	3.1 / 2.6	2.7 / 2.5	235.7 / 235.0	0.1% / 0.1%
		25 / 10	9.8 / 9.8	4.2 / 2.2	2.4 / 1.8	230.6 / 223.2	0.2% / 0.9%
		50 / 15	9.9 / 8.9	4.2 / 3.3	2.1 / 1.2	223.8 / 191.4	0.2% / 0.4%
		100 / 20	8.3 / 8.1	5.8 / 3.2	0.0 / 0.5	144.2 / 155.4	0.5% / 0.8%
water 32 variables 10083 parameters	1000	0*	38.3	21.5	5.7	439.5	0%
		10 / 5	38.1 / 35.6	19.3 / 16.6	5.6 / 4.8	434.4 / 426.1	0.1% / 0.2%
		25 / 10	37.3 / 33.1	19.8 / 16.5	4.3 / 3.7	446.5 / 415.1	0.2% / 0.9%
		50 / 15	35.9 / 30.8	18.5 / 15.4	2.9 / 2.3	420.3 / 381.0	0.4% / 0.8%
		100 / 20	36.3 / 28.9	20.3 / 14.6	0.0 / 1.7	359.8 / 365.1	0.6% / 0.9%
	4000	0*	33.2	18.0	5.2	434.3	0%
		10 / 5	31.8 / 30.8	14.8 / 14.2	5.0 / 4.7	428.6 / 421.4	0.0% / 0.1%
		25 / 10	30.8 / 27.9	15.5 / 14.4	3.4 / 3.9	417.9 / 396.4	0.1% / 0.5%
		50 / 15	30.0 / 26.1	15.6 / 13.1	2.5 / 1.7	390.2 / 348.1	0.2% / 0.3%
		100 / 20	30.5 / 25.1	19.3 / 12.7	0.0 / 1.6	343.7 / 347.9	0.3% / 0.3%
mildew 35 variables 540150 parameters	8000	0*	14.2	2.0	7.5	448.7	0%
		10 / 5	15.0 / 15.1	8.6 / 8.1	5.6 / 5.6	410.7 / 408.4	0.5% / 1.1%
		25 / 10	15.1 / 14.0	13.2 / 9.2	5.5 / 2.8	405.7 / 304.0	1.4% / 1.7%
		50 / 15	11.6 / 13.8	11.2 / 9.5	2.1 / 2.3	256.3 / 303.8	1.7% / 2.8%
		100 / 20	9.0 / 10.6	9.2 / 8.7	0.0 / 2.0	145 / 212.6	2.3% / 3.9%
	32000	0*	6.3	5.0	7.3	317.1	0%
		10 / 5	7.1 / 8.4	5.9 / 5.1	6.0 / 4.5	277.5 / 233.6	0.0% / 0.1%
		25 / 10	6.6 / 7.5	5.7 / 4.1	4.9 / 1.5	244.1 / 140.2	0.0% / 0.3%
		50 / 15	7.7 / 8.6	7.4 / 5.5	2.6 / 1.1	202.8 / 153.7	0.1% / 0.7%
		100 / 20	7.5 / 7.3	7.8 / 6.1	0.0 / 1.2	122.3 / 141.9	0.2% / 0.7%
alarm 37 variables 509 parameters	1000	0*	2.2	5.8	1.3	45.7	0%
		10 / 5	2.0 / 1.6	6.2 / 5.5	1.1 / 1.6	34.7 / 46.4	0.0% / 0.1%
		25 / 10	2.0 / 1.8	6.3 / 5.4	0.7 / 0.8	27.7 / 32.2	0.1% / 0.1%
		50 / 15	2.0 / 1.5	6.1 / 5.0	0.3 / 0.6	22.4 / 27.7	0.1% / 0.2%
		100 / 20	2.0 / 1.5	6.2 / 4.4	0.0 / 0.1	18.0 / 18.3	0.1% / 0.2%
	4000	0*	2.0	3.2	1.8	39.5	0%
		10 / 5	2.0 / 1.6	4.6 / 4.5	0.6 / 1.2	24.9 / 38.8	0.0% / 0.0%
		25 / 10	2.0 / 1.8	4.5 / 4.3	0.3 / 0.6	20.1 / 28.0	0.0% / 0.0%
		50 / 15	2.0 / 1.5	4.1 / 4.2	0.0 / 0.4	18.0 / 22.3	0.0% / 0.1%
		100 / 20	1.7 / 1.4	4.2 / 3.7	0.0 / 0.3	12.3 / 20.4	0.0% / 0.1%
barley 48 variables 114005 parameters	2000	0*	32.3	8.2	9.7	949.5	0%
		10 / 5	33.3 / 31.7	14.4 / 12.2	5.1 / 4.4	792.9 / 756.7	0.6% / 1.4%
		25 / 10	31.9 / 31.0	15.6 / 13.8	5.3 / 4.8	802.6 / 741.5	0.9% / 1.7%
		50 / 15	31.0 / 27.5	17.4 / 11.3	3.0 / 3.2	699.4 / 666.9	1.4% / 2.9%
		100 / 20	30.2 / 26.4	19.5 / 11.8	0.0 / 1.5	619.3 / 628.1	2.4% / 3.9%
	8000	0*	25.5	3.7	9.7	794.7	0%
		10 / 5	25.6 / 24.6	8.9 / 8.0	5.3 / 5.2	636.6 / 625.4	0.3% / 0.8%
		25 / 10	25.0 / 24.0	10.4 / 8.8	5.5 / 4.3	644.4 / 584.7	0.4% / 0.9%
		50 / 15	22.1 / 20.5	9.2 / 7.6	3.3 / 1.9	541.7 / 487.4	0.6% / 1.4%
		100 / 20	20.8 / 19.3	12.0 / 7.0	0.0 / 1.6	457.2 / 507.2	0.9% / 1.8%

جدول امتیازات در شبکه های متوسط 2: امتیازات



هنگام تجزیه و تحلیل تاثیر محدودیت‌های اجدادی بر کیفیت راه حل‌ها مشاهده می‌شود که تعداد یال‌های گم‌شده و معکوس تمایل به کاهش دارد در حالی که تعداد یال‌های اضافی تغییر نمی‌کند یا بدتر می‌شود. با این حال، در برخی موارد تعداد یال‌های از دست رفته بهبود نمی‌یابند، حتی هنگامی که 100% از محدودیت‌های اجدادی نیز اضافه شده باشند. به نظر تعداد کمی از محدودیت‌های مختلف بهبود قابل‌توجه و پایداری را برای کم کردن یال‌های بیشتر ارایه می‌دهند. دلیل احتمالی این است که یال وجود دارد.

محدودیت‌های نبود یال می‌توانند مستقیماً یک یال یا یال اضافی را ثابت کنند در حالی که محدودیت‌های اجدادی مبهم هستند. اگر یک مسیر نادرست برای برآورده سازی محدودیت اجدادی مورد استفاده قرار گیرد، ممکن است باعث ایجاد یال‌های اضافی شود در حالی که تعداد یال‌های گم شده را بهبود نمی‌دهد. از سوی دیگر به نظر می‌رسد که محدودیت‌های اجدادی به تنهایی به طور قابل‌توجهی (SID) را بهبود می‌بخشد. درحالی که محدودیت‌های اجدادی منجر به دلایل غیر مستقیم می‌شوند و (SID) تفاوت در اظهارات علی را اندازه‌گیری می‌کند، این انتظار را نیز میتوان داشت.

## مزایا و معایب روش MINOBSx

- برای تمام آزمایش‌ها توسط MINOBSx راه حل‌های پذیرفتنی ارائه شد.
- زمان اجرای MINOBSx بسیار کم و قابل ملاحظه بود.
- تمام راه حل‌های تولید شده توسط MINOBSx در حاشیه خطای 4٪ قرار داشتند.
- ابزار و روش MINOBSx عملکرد بسیار قوی به خصوص با افزایش تعداد متغیرها داشت.
- روش MINOBSx در برخی از داده‌ها، هنگام یافتن یال‌های گم شده بهبود نسبی داشت.
- روش MINOBSx توسط محدودیت‌های اجدادی، SID را بهبود چشمگیری داد.
- تعداد یال‌های اضافی در روش MINOBSx هنگام اعمال محدودیت‌های اجدادی، تغییر نمی‌کند و یا بدتر می‌شود.
- در شبکه‌های شامل کمتر از 3 والد روش MINOBSx عملکرد ضعیفی دارد.
- روش MINOBSx نیازمند پیش پردازش داده‌های شبکه توسط روش امتیاز دهی GOBNILP است.