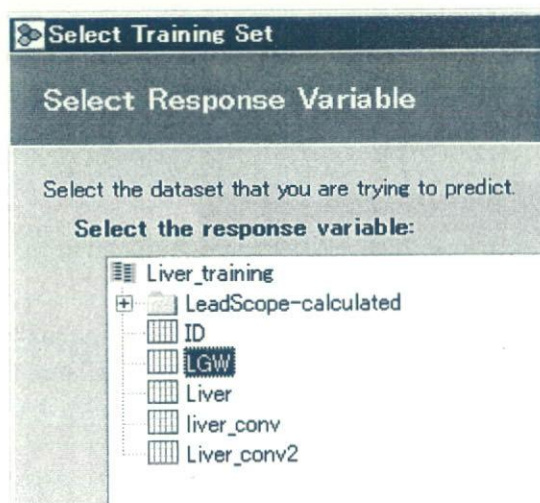


(14) Import Structures And Data : 確認して **Finish** ボタンをクリックした。

3-4-3. 予測モデルの構築

条件6: Liver, GOT/GPT, Weight_Liverデータの組み合わせ

- (1) Leadscope Task Wizardで、“Build Predictive Model” を選択した。
- (2) Predictive Data Miner画面で (下図)、各ステップにおいてモデル構築条件を選択してモデル構築を行った。



以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LGW (Liver, Weight_Liver, GOT/GPTのいずれも0の時は0, それ以外は1とした値)
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional	Rule名 : Scaffold Generation(8) (Default)

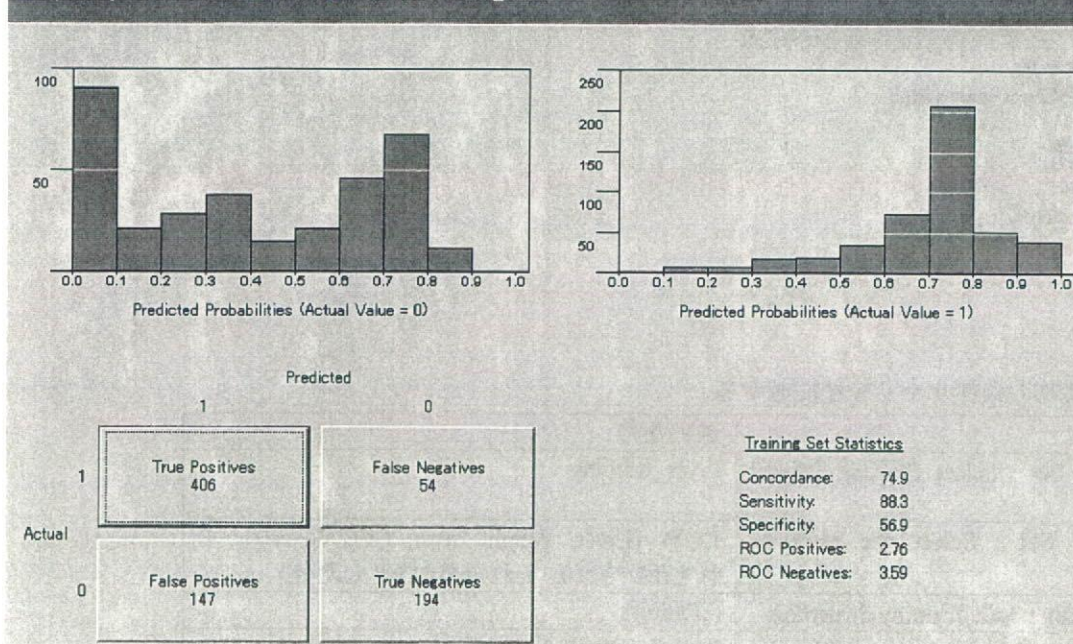
Structural Rules	
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Automated Feature Selection - Automatically select features (Default)
Build Model :	Automated Model Building (Default)

以下にModel構築の結果を記載した。

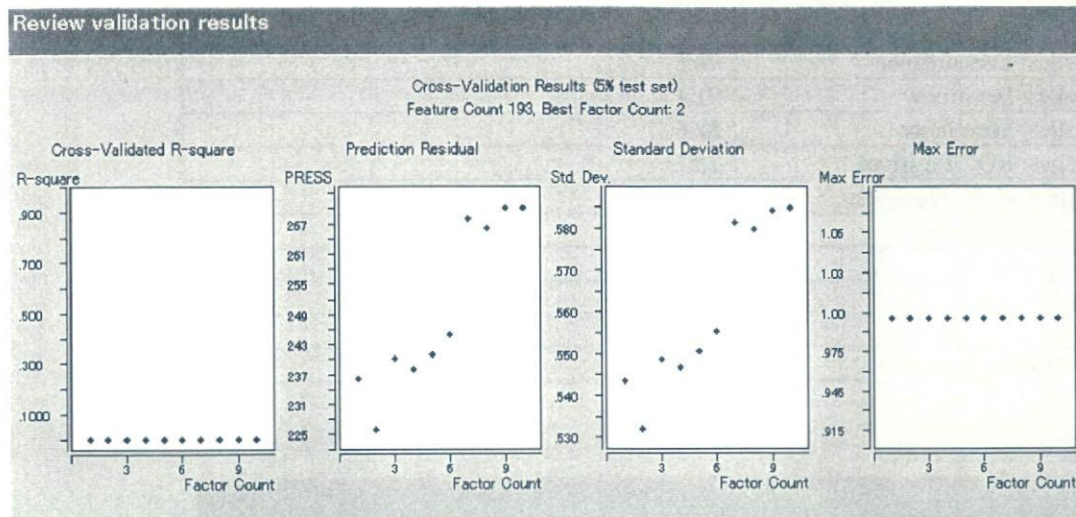
項目	選択条件
Build Knowledge	91 Rules (Scaffold Generation(8))
Select Features	193 Selected Features
Prediction Model Name	PredictiveModel (Run#14)
Type	Partial Logistic Regression
Factor Count	2
Training Set Statistics: Concordance	74.9
Training Set Statistics: Sensitivity	88.3
Training Set Statistics: Specificity	56.9
Training Set Statistics: ROC Positives	2.76
Training Set Statistics: ROC Negatives	3.59
Test Set Statistics: Concordance	71.8
Test Set Statistics: Sensitivity	85.4
Test Set Statistics: Specificity	53.4
Test Set Statistics: ROC Positives	2.47
Test Set Statistics: ROC Negatives	2.72

予測モデルの精度

Compare predicted values with actual (training set)



Cross-Validation Results



条件7：条件6に対してMean LGWが0.3-0.6となるFeaturesを除外

“Edit Predictive Model”機能を使って条件6のモデルから、Mean LGWが0.3以上～0.6未満のものを除外した。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LGW※
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(8) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Manually Selected Features 上記参照
Build Model :	Automated Model Building (Default)

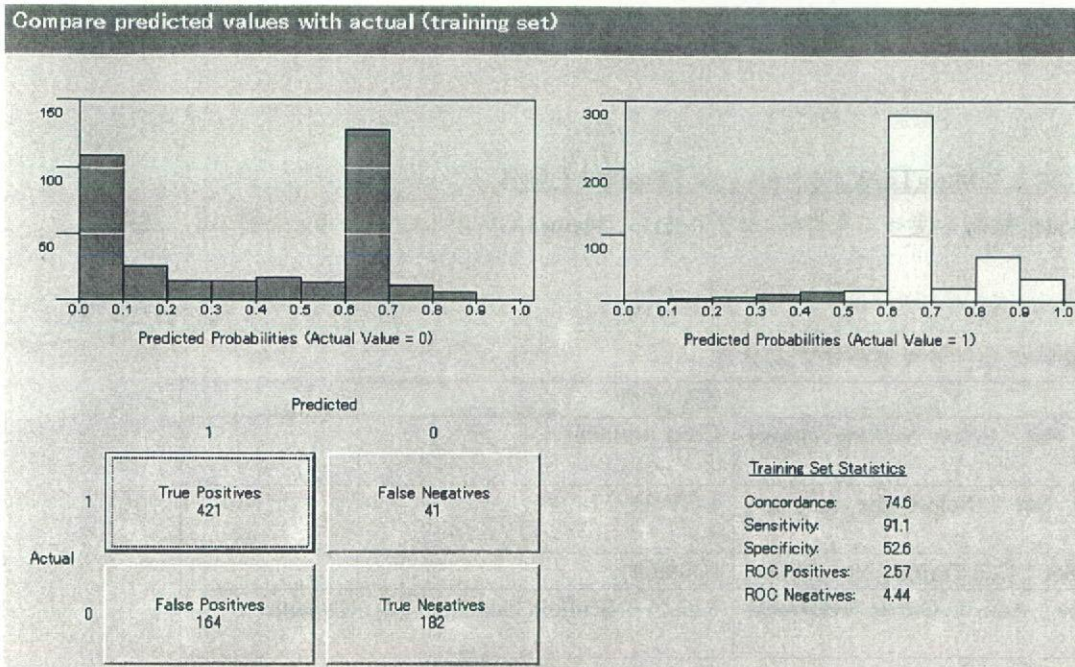
※Liver, Weight_Liver, GOT/GPTのいずれかが0の時は0, それ以外は1とした値。

以下にModel構築の結果を記載した。

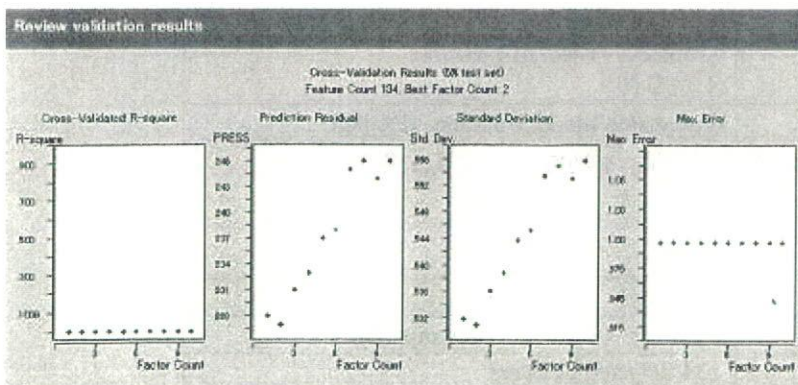
項目	選択条件
Build Knowledge	68 Rules (Scaffold Generation(8))
Select Features	134 Selected Features

Prediction Model Name	PredictiveModel (Run#15)
Type	Partial Logistic Regression
Factor Count	2
Training Set Statistics: Concordance	74.6
Training Set Statistics: Sensitivity	91.1
Training Set Statistics: Specificity	52.6
Training Set Statistics: ROC Positives	2.57
Training Set Statistics: ROC Negatives	4.44
Test Set Statistics: Concordance	71.9
Test Set Statistics: Sensitivity	88.1
Test Set Statistics: Specificity	50.3
Test Set Statistics: ROC Positives	2.37
Test Set Statistics: ROC Negatives	3.16

予測モデルの精度



Cross-Validation Results



条件8: LiverとGOT/GPTデータの組合せ

上記LGの値を肝毒性データとしてモデルを構築した。手順は条件6の時と全く同様。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LG※
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(9) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Automated Feature Selection - Automatically select features (Default)
Build Model :	Automated Model Building (Default)

※Liver, GOT/GPTのいずれかが0の時は0, それ以外は1とした値。

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	113 Rules (Scaffold Generation(9))
Select Features	207 Selected Features
Prediction Model Name	PredictiveModel (Run#16)
Type	Partial Logistic Regression
Factor Count	3
Training Set Statistics: Concordance	73.5
Training Set Statistics: Sensitivity	82.4
Training Set Statistics: Specificity	65.7
Training Set Statistics: ROC Positives	2.12
Training Set Statistics: ROC Negatives	4.24
Test Set Statistics: Concordance	69.2
Test Set Statistics: Sensitivity	77.3
Test Set Statistics: Specificity	62.0
Test Set Statistics: ROC Positives	1.79
Test Set Statistics: ROC Negatives	3.11

条件9: LiverとGOT/GPTデータの組合せ

条件8のモデルを作成した際のFeaturesのセットからMean LGが0.3以上、0.6未満となるFeaturesを除外してモデルを構築した。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LG※
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(9) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Manually Selected ※下記参照
Build Model :	Automated Model Building (Default)

※"Edit Predictiv Model"により、Mean LGが0.3以上0.6未満のFeaturesを除外した。

These are the features that will be included in the model.

Name	Frequency	LG	Z-Score: LG	Chi-Square: LG	Mean: LG	Std Dev: LG	LG
carboxyl, alkyl, acyl-	20		1.896	2.076	0.28	0.4666	
Scaffold 274	32		-2.143	3.856	0.2813	0.4568	
carboxamide(NHR)	52		-2.66	6.941	0.2885	0.4575	
Scaffold 166	65		-2.935	7.881	0.2923	0.4584	
Scaffold 60	51		-2.549	5.787	0.2941	0.4602	
carboxylate	73		-2.365	8.087	0.3014	0.462	
Scaffold 23	33		-1.922	3.044	0.303	0.4667	
Scaffold 93	59		-2.581	5.938	0.3051	0.4644	
sulfonate	39		-2.038	3.513	0.3077	0.4676	
Scaffold 46	39		-2.038	3.513	0.3077	0.4676	
oxycarbonyl, O-alkyl-	80		-2.909	7.797	0.3125	0.4664	
1-benzene-carboxylic acid	35		-1.345	2.8	0.3143	0.471	
Scaffold 52	63		-2.169	5.473	0.3175	0.4692	
Scaffold 110	40		-1.34	2.817	0.325	0.4743	
ether, alkyl, cyc-	46		-1.366	3.293	0.3261	0.474	
carboxamide	76		-2.525	5.79	0.3289	0.473	
Scaffold 72	73		-2.227	4.434	0.3425	0.4778	
carbonyl, phenyl-	107		-2.587	6.681	0.3458	0.4779	
Scaffold 116	69		-2.066	3.769	0.3478	0.4798	
Scaffold 108	149		-3.184	9.579	0.349	0.4783	
oxycarbonyl, O-(alkyl, acyl)-	67		-1.856	2.989	0.3582	0.4831	
benzene, 1-amino-	123		-2.431	5.447	0.3659	0.4836	
carbonyl, aryl-	130		-2.427	5.443	0.3692	0.4845	
carbonyl, alkyl, acyl-	102		-1.323	2.952	0.3824	0.4884	
Parent Atom Count	808				0.4666	0.4992	
Rotatable Bonds	808				0.4666	0.4992	
Hydrogen Bond Acceptors	808				0.4666	0.4992	
Hydrogen Bond Donors	808				0.4666	0.4992	
Parent Molecular Weight	808				0.4666	0.4992	
ALogP	801		0.9328		0.4682	0.4993	
Polar Surface Area	801		0.9328		0.4682	0.4993	
Lipinski Score	801		0.9328		0.4682	0.4993	
Scaffold 119	141		2.032	3.951	0.5461	0.4936	
halide, phenyl-	96		2.877	7.67	0.6042	0.4916	
Scaffold 147	96		2.877	7.67	0.6042	0.4916	
amine(NH2), phenyl-	43		1.864	2.917	0.6047	0.4947	
Scaffold 21	30		1.865	2.82	0.6333	0.4901	
4	28		1.864	2.917	0.6333	0.4901	

Features:207 Selected Features:28 Compounds:808 Selected Compounds:808

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	104 Rules (Scaffold Generation(9))
Select Features	179 Selected Features
Prediction Model Name	PredictiveModel (Run#18)
Type	Partial Logistic Regression
Factor Count	2
Training Set Statistics: Concordance	71.3
Training Set Statistics: Sensitivity	82.8
Training Set Statistics: Specificity	61.3
Training Set Statistics: ROC Positives	1.87
Training Set Statistics: ROC Negatives	4.06
Test Set Statistics: Concordance	69.4
Test Set Statistics: Sensitivity	81.7
Test Set Statistics: Specificity	58.7
Test Set Statistics: ROC Positives	1.73
Test Set Statistics: ROC Negatives	3.67

条件10: LiverとGOT/GPTデータの組合せ

条件8のモデル構築の際に利用したFeaturesから、下記手順によりFeaturesを除外してモデルを再度構築した。手順は条件6の時と全く同様。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LG
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(9) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Manually Selected ※
Build Model :	Automated Model Building (Default)

※Mean LGが0.3-0.7の間でZ-Scoreが灰色(-2~2の範囲)にあるFeaturesを削除した。

Appendix 9

These are the features that will be included in the model

Name	Frequency	LG	Z-Score: LG	Chi-Square: LG	Mean: LG	Std Dev: LG	LG
Scaffold 46	39		-2.038	3.513	0.3077	0.4676	
oxycarbonyl, O-alkyl-	80		-2.909	7.797	0.3125	0.4664	
1-benzene-carboxylic acid	35		-1.845	2.8	0.3143	0.471	
Scaffold 52	63		-2.469	5.473	0.3175	0.4692	
Scaffold 110	40		-1.84	2.817	0.325	0.4743	
ether, alkyl, cyc-	46		-1.966	3.293	0.3261	0.474	
carboxamide	76		-2.525	5.79	0.3289	0.473	
Scaffold 72	73		-2.227	4.434	0.3425	0.4778	
carbonyl, phenyl-	107		-2.687	6.681	0.3458	0.4779	
Scaffold 116	69		-2.066	3.769	0.3478	0.4798	
Scaffold 108	149		-3.184	9.579	0.349	0.4783	
oxycarbonyl, O-(alkyl, acyl)-	67		-1.856	2.989	0.3582	0.4831	
benzene, 1-amino-	123		-2.431	5.447	0.3659	0.4836	
carbonyl, aryl-	130		-2.427	5.443	0.3692	0.4845	
carbonyl, alkyl, acyl-	102		-1.823	2.952	0.3824	0.4884	
Parent Atom Count	808				0.4666	0.4992	
Rotatable Bonds	808				0.4666	0.4992	
Hydrogen Bond Acceptors	808				0.4666	0.4992	
Hydrogen Bond Donors	808				0.4666	0.4992	
Parent Molecular Weight	808				0.4666	0.4992	
ALogP	801		0.9628		0.4692	0.4993	
Polar Surface Area	801		0.9628		0.4682	0.4993	
Lipinski Score	801		0.9628		0.4682	0.4993	
Scaffold 119	141		2.082	3.961	0.5461	0.4996	
halide, phenyl-	96		2.877	7.67	0.6042	0.4916	
Scaffold 147	96		2.877	7.67	0.6042	0.4916	
amine(NH2), phenyl-	43		1.854	2.917	0.6047	0.4947	
Scaffold 21	30		1.865	2.82	0.6333	0.4901	
benzene, 1,2-dihalo-	26		1.944	3.048	0.6538	0.4852	
benzene, 1-chloro-3-halo-	26		1.944	3.048	0.6538	0.4852	
Scaffold 154	26		1.944	3.048	0.6538	0.4852	
Scaffold 149	26		1.944	3.048	0.6538	0.4852	
Scaffold 56	19		1.923	2.861	0.6842	0.4776	
Scaffold 140	20		2.117	3.579	0.7	0.4702	
Scaffold 300	22		2.483	5.146	0.7273	0.4558	
benzene, 1-alkyl-2-halo-	15		2.089	3.346	0.7333	0.4577	
Scaffold 159	15		2.089	3.346	0.7333	0.4577	

Features:207 Selected Features:20 Compounds:808 Selected Compounds:808

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	108 Rules (Scaffold Generation(9))
Select Features	187 Selected Features
Prediction Model Name	PredictiveModel (Run#20)
Type	Partial Logistic Regression
Factor Count	4
Training Set Statistics: Concordance	73.8
Training Set Statistics: Sensitivity	84.4
Training Set Statistics: Specificity	64.5
Training Set Statistics: ROC Positives	2.08
Training Set Statistics: ROC Negatives	4.71
Test Set Statistics: Concordance	70.5
Test Set Statistics: Sensitivity	80.9
Test Set Statistics: Specificity	61.5
Test Set Statistics: ROC Positives	1.84
Test Set Statistics: ROC Negatives	3.68

条件11: LiverとGOT/GPTデータの組合せ

条件8のモデル構築の際に利用したFeaturesから、下記手順によりFeaturesを除外してモデルを再度構築した。手順は条件6の時と全く同様。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LG
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(9) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Manually Selected ※
Build Model :	Automated Model Building (Default)

※Z-Scoreが灰色(-2~2の範囲)で表示されるFeaturesを全て削除した。また、Z-Scoreが無いものは、Mean LGが0.3-0.6の間になるFeaturesを除外した。

Appendix 9

These are the features that will be included in the model.

Name	Frequency	LQ	Z-Score LQ	Chi-Square LQ	Mean LQ	Std Dev LQ	LQ
benzene, 1-bromo-4-fluoro-	4		2.142	2.694	1.0	0.0	
benzene, 1-halo-4-C(=O)acethoxy-	4		2.142	2.694	1.0	0.0	
hydrazone(RNHNH2)	4		2.142	2.694	1.0	0.0	
benzene, 1-bromo-3-fluoro-	4		2.142	2.694	1.0	0.0	
bromide, alkyl-	4		2.142	2.694	1.0	0.0	
Scaffold 140	20		2.117	3.679	0.7	0.4702	
benzene, 1-alkyl-2-halo-	15		2.089	3.345	0.7333	0.4577	
Scaffold 169	15		2.089	3.345	0.7333	0.4577	
Scaffold 119	141		2.082	3.961	0.5461	0.4906	
Scaffold 117	7		2.079	2.889	0.9571	0.378	
benzene, 1-alkyl-3-amine(NH2)-	7		2.079	2.889	0.9571	0.378	
sulfone	12		1.982	2.99	0.75	0.4523	
Scaffold 154	26		1.944	3.048	0.6538	0.4852	
Scaffold 148	26		1.944	3.048	0.6538	0.4852	
benzene, 1-chloro-3-halo-	26		1.944	3.048	0.6538	0.4852	
benzene, 1,2-dihalo-	26		1.944	3.048	0.6538	0.4852	
Scaffold 59	19		1.923	2.861	0.6842	0.4776	
Scaffold 21	30		1.885	2.82	0.6303	0.4801	
amine(NH2), phenyl-	43		1.854	2.917	0.6047	0.4847	
Polar Surface Area	801		0.928		0.4682	0.4993	
Lipinski Score	801		0.928		0.4682	0.4993	
AllogP	801		0.928		0.4682	0.4993	
carbonyl, alkyl, acyl-	102		-1.023	2.952	0.3324	0.4894	
Scaffold 110	40		-1.84	2.817	0.326	0.4743	
1-benzene-carboxylic acid	35		-1.845	2.8	0.3149	0.471	
oxycarbonyl, O-alkyl, acyl-	27		-1.956	2.999	0.2582	0.4831	
carbonyl, alkyl, cyc-	25		-1.898	2.876	0.28	0.4593	
carbonyl, alkyl, cyc-	25		-1.898	2.876	0.28	0.4593	
Scaffold 292	14		-1.908	2.685	0.2143	0.4258	
Scaffold 145	14		-1.908	2.685	0.2143	0.4258	
Scaffold 23	33		-1.922	3.044	0.303	0.4667	
Scaffold 237	17		-1.991	2.844	0.2253	0.4372	
Scaffold 30	17		-1.991	2.844	0.2253	0.4372	
ether, alkyl, cyc-	45		-1.966	3.293	0.3201	0.474	
sulfonate	39		-2.038	3.613	0.3077	0.4676	
Scaffold 46	39		-2.038	3.613	0.3077	0.4676	
carboxylate, aryl-	26		-2.049	3.425	0.2082	0.4523	
Scaffold 172	26		-2.049	3.425	0.2082	0.4523	

Features:202 Selected Features:23 Compounds:808 Selected Compounds:801

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	103 Rules (Scaffold Generation(9))
Select Features	179 Selected Features
Prediction Model Name	PredictiveModel (Run#21)
Type	Partial Logistic Regression
Factor Count	4
Training Set Statistics: Concordance	73.0
Training Set Statistics: Sensitivity	85.1
Training Set Statistics: Specificity	62.4
Training Set Statistics: ROC Positives	1.98
Training Set Statistics: ROC Negatives	4.8
Test Set Statistics: Concordance	70.0
Test Set Statistics: Sensitivity	81.7
Test Set Statistics: Specificity	59.9
Test Set Statistics: ROC Positives	1.78
Test Set Statistics: ROC Negatives	3.74

条件12: LiverとWeight_Liverデータの組合せ

Liver、Weight_Liverの値を肝毒性データ(LW)としてモデルを構築した。手順は条件6の時と全く同様。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LW※
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(10) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Automated Feature Selection - Automatically select features (Default)
Build Model :	Automated Model Building (Default)

※Liver、Weight_Liverのいずれかが0の時は0、それ以外は1とした値。

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	111 Rules (Scaffold Generation(10))
Select Features	189 Selected Features
Prediction Model Name	PredictiveModel (Run#17)
Type	Partial Logistic Regression
Factor Count	3
Training Set Statistics: Concordance	73.8
Training Set Statistics: Sensitivity	83.8
Training Set Statistics: Specificity	62.8
Training Set Statistics: ROC Positives	2.47
Training Set Statistics: ROC Negatives	3.53
Test Set Statistics: Concordance	68.9
Test Set Statistics: Sensitivity	79.2
Test Set Statistics: Specificity	57.6
Test Set Statistics: ROC Positives	2.05
Test Set Statistics: ROC Negatives	2.53

条件13: LiverとWeight_Liverデータの組合せ

条件12のモデルから**Mean LW**が**0.3**以上**0.6**未満の**Features**を除外してモデルを構築した。手順は条件6の時と全く同様。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LW
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(10) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Manually Selected ※
Build Model :	Automated Model Building (Default)

※

Appendix 9

These are the features that will be included in the model.

Name	Frequency	LW	Z-Score: LW	Chi-Square: LW	Mean: LW	Std Dev: LW	LW
Scaffold 87	25		-2.03	3.341	0.32	0.4761	
benzene, 1-arylamino-	28		-2.137	3.787	0.3214	0.4756	
Scaffold 152	65		-3.308	10.12	0.3231	0.4713	
Scaffold 190	34		-2.339	4.688	0.3235	0.4749	
Scaffold 160	37		-2.435	5.144	0.3243	0.4746	
carboxylate, alkyl-	27		-1.971	3.157	0.3333	0.4804	
Scaffold 53	51		-2.752	6.807	0.3333	0.4761	
Scaffold 22	33		-2.188	4.045	0.3333	0.4787	
1-benzene-carboxylic acid	35		-2.141	3.878	0.3429	0.4816	
ether, alkyl, cyc-	46		-2.403	5.072	0.3478	0.4815	
Scaffold 189	34		-1.989	3.292	0.3529	0.4851	
Scaffold 33	70		-2.848	7.426	0.3571	0.4826	
benzene, 1-sulfonyl-	50		-2.334	4.792	0.36	0.4849	
carboxamide(NHR)	52		-2.303	4.669	0.3654	0.4862	
methane, 1,1-diphenyl-	38		-1.912	3.052	0.3684	0.4889	
tert-amine	46		-2.089	3.796	0.3696	0.488	
carboxylic acid, aryl-	48		-2.069	3.692	0.375	0.4892	
Scaffold 37	80		-2.494	5.652	0.3875	0.4903	
Scaffold 81	59		-2.074	3.764	0.3898	0.4919	
carbonyl, phenyl-	107		-2.827	7.427	0.3925	0.4906	
Scaffold 46	63		-2.033	3.623	0.3968	0.4932	
carboxylate	73		-2.196	4.303	0.3973	0.4927	
oxycarbonyl, O-alkyl-	80		-2.023	3.632	0.4125	0.4954	
carboxylic acid	100		-2.133	4.109	0.42	0.496	
carboxy	100		-2.133	4.109	0.42	0.496	
Scaffold 125	118		-2.259	4.669	0.4237	0.4963	
Scaffold 96	149		-2.078	3.953	0.443	0.4994	
Hydrogen Bond Acceptors	808				0.5198	0.4999	
Hydrogen Bond Donors	808				0.5198	0.4999	
Parent Atom Count	808				0.5198	0.4999	
Parent Molecular Weight	808				0.5198	0.4999	
Rotatable Bonds	808				0.5198	0.4999	
ALogP	801		2.004		0.5231	0.4998	
Polar Surface Area	801		2.004		0.5231	0.4998	
Lipinski Score	801		2.004		0.5231	0.4998	
Scaffold 3	60		1.828	2.874	0.6333	0.486	
halide, phenyl-	96		2.414	5.32	0.6354	0.4838	
benzene, 1-chloro-	56		1.909	3.14	0.6429	0.4835	

Features:189 Selected Features:44 Compounds:808 Selected Compounds:808

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	96 Rules (Scaffold Generation(10))
Select Features	145 Selected Features
Prediction Model Name	PredictiveModel (Run#19)
Type	Partial Logistic Regression
Factor Count	2
Training Set Statistics: Concordance	70.5
Training Set Statistics: Sensitivity	86.9
Training Set Statistics: Specificity	52.8
Training Set Statistics: ROC Positives	1.99
Training Set Statistics: ROC Negatives	3.73
Test Set Statistics: Concordance	68.1
Test Set Statistics: Sensitivity	85.5
Test Set Statistics: Specificity	49.2
Test Set Statistics: ROC Positives	1.82
Test Set Statistics: ROC Negatives	3.13

条件14: LiverとWeight Liverデータの組合せ

条件12のモデルから下記の手順でFeaturesを除外してモデルを構築した。手順は条件6の時と全く同様。

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	Liver_training
Select Training Set : Select the response variables	LW
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation(10) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Manually Selected ※
Build Model :	Automated Model Building (Default)

※Mean LWが0.294-0.7の範囲内でZ-Scoreが灰色(-2~2の範囲))にあるFeaturesを除外して、モデルを構築した。下図の表で選択したfeaturesを削除してモデルを構築した。

These are the features that will be included in the model

Name	Frequency	LW	Z-Score LW	Chi-Square LW	Mean LW	Std Dev LW	LW
Scaffold 22	38		-2188	4.045	0.3333	0.4787	
1-benzene-carboxylic acid ether, alkyl cyc-	35		-2141	3.878	0.3429	0.4816	
ether, alkyl cyc-	46		-2403	5.072	0.2479	0.4815	
Scaffold 189	34		-1959	3.292	0.3523	0.4851	
Scaffold 33	70		-2348	7.428	0.5571	0.4828	
benzene, 1-sulfonyl-carboxamide(OH)O	60		-2334	4.792	0.36	0.4849	
carboxamide(OH)O	52		-2303	4.669	0.3654	0.4802	
methane, 1,1-diphenyl-	30		-1912	3.052	0.3694	0.4839	
tert-amine	46		-2069	3.756	0.3695	0.488	
carboxylic acid aryl-	48		-2069	3.662	0.375	0.4892	
Scaffold 37	90		-2404	5.652	0.3875	0.4903	
Scaffold 81	59		-2074	3.764	0.3898	0.4919	
carbonyl phenyl-	107		-2827	7.427	0.3925	0.4906	
Scaffold 46	63		-2033	3.623	0.3968	0.4922	
carboxylate	73		-2196	4.303	0.3973	0.4927	
oxycarbonyl, O-alkyl-	80		-2023	3.622	0.4125	0.4954	
carboxylic acid	100		-2139	4.109	0.42	0.496	
carboxy	100		-2133	4.109	0.42	0.496	
Scaffold 125	118		-2259	4.669	0.4237	0.4963	
Scaffold 96	149		-2078	3.953	0.443	0.4934	
Hydrogen Bond Acceptors	908				0.5193	0.4999	
Hydrogen Bond Donors	908				0.5193	0.4999	
Parent Atom Count	908				0.5193	0.4999	
Parent Molecular Weight	908				0.5193	0.4999	
Rotatable Bonds	608				0.5193	0.4999	
AlecP	801		2.004		0.5231	0.4998	
Polar Surface Area	801		2.004		0.5231	0.4998	
Lipinski Score	801		2.004		0.5231	0.4998	
Scaffold 3	60		-1828	2.874	0.5333	0.495	
halide, phenyl-	96		-2414	5.32	0.6354	0.4938	
benzene, 1-chloro-	56		-1909	3.14	0.6429	0.4936	
benzene, 1-fluoro-	36		-1833	2.569	0.6667	0.4791	
Scaffold 294	42		-1956	3.223	0.6667	0.4771	
benzene, 1-chloro-3-halo-	26		-2386	5.703	0.7692	0.4297	
Scaffold 308	22		-2405	4.801	0.7727	0.4288	
1-benzene-nitrile	15		-2191	3.732	0.8	0.414	
Scaffold 21	21		-1254	3.451	0.8065	0.4016	

Features:189 Selected Features:17 Compounds:608 Selected Compounds:808

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	Rules (Scaffold Generation(10))
Select Features	Selected Features
Prediction Model Name	PredictiveModel (Run#22)
Type	Partial Logistic Regression
Factor Count	
Training Set Statistics: Concordance	73.5
Training Set Statistics: Sensitivity	83.5
Training Set Statistics: Specificity	62.6
Training Set Statistics: ROC Positives	2.45
Training Set Statistics: ROC Negatives	3.46
Test Set Statistics: Concordance	69.0
Test Set Statistics: Sensitivity	79.2
Test Set Statistics: Specificity	57.9
Test Set Statistics: ROC Positives	2.06
Test Set Statistics: ROC Negatives	2.54

条件15: 条件7のモデルで陽性と予測された化合物群からの予測モデルの作成

条件7のモデルは感度が高い(90%程度)ものの特異性は低い(50%程度)。このモデルにより「陽性」

と予測された化合物群から予測モデルを再構築し、予測精度に改善がみられるかについて以下で検討した。条件7の予測モデルで陽性と予測された586化合物をトレーニングセットとしてモデルを構築した。

		条件7のモデル(Run#15)		Total	Concordance	74.6
		+	-			
LGW	+	421	41	462	Sensitivity	91.1
	-	164	182	346	Specificity	<u>52.6</u>
				808		

以下にモデル構築時の条件サマリーを記載する。

項目	選択条件
Select Training Set : Select Starting Project for Training	condition11_positives
Select Training Set : Select the response variables	Liver_conv2 (1)
Select Training Set : Edit Training Structures	(Default)
Build Knowledge : Add or Create Additional Structural Rules	Rule名 : Scaffold Generation (1) (Default)
Build Knowledge : Filter Structural Rules	Filter Structural Rules (Default)
Build Knowledge : Review Structures Rules Template	No Edit (Default)
Select Features :	Automatically Selected
Build Model :	Automated Model Building (Default)

以下にModel構築の結果を記載した。

項目	選択条件
Build Knowledge	96 Rules (Scaffold Generation(10))
Select Features	207 Selected Features
Prediction Model Name	PredictiveModel
Type	Partial Logistic Regression
Factor Count	3
Training Set Statistics: Concordance	77.3
Training Set Statistics: Sensitivity	68.5
Training Set Statistics: Specificity	84.9
Training Set Statistics: ROC Positives	3.89
Training Set Statistics: ROC Negatives	3.14
Test Set Statistics: Concordance	72.7
Test Set Statistics: Sensitivity	62.9
Test Set Statistics: Specificity	81.0
Test Set Statistics: ROC Positives	2.85
Test Set Statistics: ROC Negatives	2.55

条件7のモデルのTraining Setに対する予測精度と比べると、Concordanceは74.6%から77.3%へと全体

的には若干精度が上がっているものの、Sensitivityは91.1%から68.5%へと大幅に下がっている。
Specificityは52.6%から84.9%へと上がっている。

3-4-4. 予測モデルのまとめ 2

各モデルの予測精度の比較 (Training Setに対する予測精度を表示)

条件	トレーニング グセット	モデル名	Feature 数	Features選択方法	Conco rdance	Sensi tivity	Speci ficity
条件1	Liver_conv	Run#10	207	Automatic	87.1	50.3	96.7
条件2	Liver_conv	Run#11	156	Manual(Meanが0.3-0.6となるFeaturesを除外)	87.3	43.6	98.6
条件3	Liver_conv2	Run#12	207	Automatic	77.5	53.4	91.4
条件4	Liver_conv2	Run#4	624	Manual(Meanが0.3-0.6となるFeaturesを除外)	80	62.3	90.2
条件5	Liver_conv2	Run#13	283	Manual(Meanが0.1-0.6となるFeaturesを除外)	80	56.3	93.4
条件6	LGW	Run#14	193	Automatic	74.9	88.3	56.9
条件7	LGW	Run#15	134	Manual(Meanが0.3-0.6となるFeaturesを除外など)	74.6	91.1	52.6
条件8	LG	Run#16	207	Automatic	73.5	82.4	65.7
条件9	LG	Run#18	179	Manual(Meanが0.3-0.6となるFeaturesを除外など)	71.3	82.8	61.3
条件10	LG	Run#20	187	Manual(Meanが0.3-0.7でZ-Scoreが±2範囲内となるFeaturesのみを除外)	73.8	84.4	64.5
条件11	LG	Run#21	179	Manual(Z-Scoreが小さい(±2)Featuresを全て削除)	73.0	85.1	62.4
条件12	LW	Run#17	189	Automatic	73.8	83.8	62.8
条件13	LW	Run#19	145	Manual(Meanが0.3-0.6となるFeaturesを除外など)	70.5	86.9	52.8
条件14	LW	Run#22		Manual(Mean LWが0.294-0.7でZ-Scoreが±2範囲内となるFeaturesのみを除外)	73.5	83.5	62.6
条件15	Liver_conv2	Predictive Model(番号なし)	207	Automatic	77.3	68.5	84.9

3-5. 条件4のモデルと条件7のモデルとの組合せモデルの予測精度の検討

[背景]

条件4のモデルは特異性が非常に高く感度が比較的低い。条件7のモデルは逆に感度が高く特異性が低い。そこでこれらのモデルを相補的に組み合わせて肝毒性を予測することで高い精度が得られると考えられる。両者を組み合わせたモデルの予測精度を以下の手順で検討した。

[方法概略]

- (1) Leadscape Task Wizardで、“Build Predictive Model” をクリックし、条件4(Run#4)あるいは条件7(Run#15)のモデルを選択した。
- (2) Predictive Data Miner画面でReviewボタンをクリックし、該当モデルの各化合物に対する予測結果をスプレッドシート形式で表示させ、そのデータをExcelへコピー・ペーストし陽性、陰性の組み合わせを集計した。集計結果は以下のようなになる。

Liver_conv2	条件4	条件7	件数
-	-	-	196
-	-	+	263
-	+	-	2
-	+	+	48
-	(適用外)	+	6
+	-	-	15
+	-	+	95
+	+	-	10
+	+	+	172
+	(適用外)	+	1
		total	808

LGW	条件4	条件7	件数
-	-	-	180
-	-	+	130
-	+	-	2
-	+	+	29
-	(適用外)	+	5
+	-	-	31
+	-	+	228
+	+	-	10
+	+	+	191
+	(適用外)	+	2
		total	808

[結果]

上記表において条件4と条件7とで陰性/陽性が一致しない場合は予測不可能と見なし、予測可能なケースのみ集計した。また、条件4はLiver_conv2、条件7はLGWをデータセットとしているので、予測対象となる変数はLiver_conv2とLGWの2通りがあるため、各々のデータセットに関する予測結果を以下の表にまとめた。

		条件4と条件7モデルの組合せ			%	
		+	-	total	concordance	85.38
Liver_conv2	+	172	15	187	sensitivity	91.98
	-	48	196	244	specificity	80.33
				431		
予測可能な化合物の割合(%):		53.34				

		条件4と条件7モデルの組合せ	%

Appendix 9

		+	-	total	concordance	86.08
LGW	+	191	31	222	sensitivity	86.04
	-	29	180	209	specificity	86.12
				431		
予測可能な化合物の割合(%):			53.34			

モデルを組み合わせた場合、適用範囲が小さくなる代わりに予測精度が大幅に向上することが分かった。