Contents

1   Notation

Type 1: Side Effects: 0, Drugs: 0, Patients: 0
Type 2: Side Effects: 0, Drugs: 0, Patients: 1
Type 3: Side Effects: 0, Drugs: 1, Patients: 1
Type 4: Side Effects: 1, Drugs: 1, Patients: 1
Type 5: Side Effects: 1, Drugs: 0, Patients: 1
Type 6: Side Effects: 1, Drugs: 0, Patients: 0

The labels 0 and 1 with Side Effects, Drug and Patients represent:
0: Constraint not Applied
1: Constraint Applied

Constraint on Side Effect:
Every evaluations with rating of side effect must report the side caused.

Constraint on Drugs:
Each drug must be evaluated by at least 4 unique patients

Constraint on Patients:
Each patient must be under treatment with at least 3 unique drugs

* **Bold** values inside a table represent the best performing algorithm.

2   Dataset Details

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
Drugs	1297	1159	692	660	1141	1274
Patients	33013	8051	7948	8006	8011	31092
Latest Evaluations	76368	45645	37134	41298	42075	70938
Evaluations	102888	67034	...	.....	....	95791

Evaluations: Refers to total evaluation of all the drugs by patient including the multiple evaluations of a drug by a patient.

Latest Evaluations: Refers to total evaluation of all the drugs by patient considering only latest evaluation of a patient for a particular drug and hence excluding the multiple evaluations of a drug by a patient

3 Librec Results

3.1 Type 1, 2 and 3 Results

Algorithm	Type 1	Type 2	Type 3
Biased MF	0.856	0.908	0.836
PMF	0.919	**0.884**	**0.817**
NMF	0.989	1.168	1.007
SVD++	0.859	0.915	0.833
User Average	0.892	0.919	0.884
Item Average	**0.852**	0.944	0.830

Table 1: Side Effect Prediction: Mean Absolute Error

Algorithm	Type 1	Type 2	Type 3
Biased MF	1.057	1.149	1.059
PMF	1.304	1.239	1.144
NMF	1.293	1.581	1.355
SVD++	1.052	1.146	1.056
User Average	1.153	**1.101**	1.140
Item Average	**1.022**	1.222	**1.005**

Table 2: Side Effect Prediction: Root Mean Square Error

3.2 Type 4 and 5 Results

3.2.1 Type 4 Results

Dim	BiasMF	PMF	NMF	SVD++	BPMF	User	Item
64	0.817	**0.781**	0.868	0.818		0.879	0.839
32	0.821	**0.786**	0.908	0.819		0.878	0.840
16	0.825	**0.791**	0.964	0.827		0.881	0.840
8	0.833	**0.806**	1.004	0.835		0.882	0.840
4	0.846	**0.816**	0.990	0.843		0.881	0.839

Table 3: Root Mean Absolute Error Values

Dim	BiasMF	PMF	NMF	SVD++	BPMF	User	Item
64	1.017	1.137	1.157	1.016		1.141	1.015
32	1.027	1.136	1.220	1.024		1.140	**1.016**
16	1.044	1.132	1.306	1.044		1.143	**1.016**
8	1.070	1.140	1.373	1.072		1.144	**1.016**
4	1.099	1.140	1.354	1.095		1.145	**1.016**

Table 4: Root Mean Square Error Values

3.2.2   Type 5 Results

Dim	BiasMF	PMF	NMF	SVD++	BPMF	User	Item
64	0.818	**0.780**	0.868	0.819	1.012	0.877	0.841
32	0.816	**0.782**	0.905	0.823	0.972	0.881	0.840
16	0.826	**0.786**	0.962	0.827	0.941	0.881	0.841
8	0.835	**0.803**	1.015	0.835	0.919	0.876	0.840
4	0.843	**0.813**	0.985	0.841	0.892	0.879	0.841

Table 5: Root Mean Absolute Error Values

Dim	BiasMF	PMF	NMF	SVD++	BPMF	User	Item
64	**1.017**	1.141	1.156	**1.017**	1.319	1.139	1.020
32	1.023	1.133	1.217	1.028	1.260	1.142	**1.019**
16	1.043	1.129	1.306	1.045	1.218	1.144	**1.020**
8	1.072	1.142	1.382	1.072	1.187	1.138	**1.019**
4	1.095	1.140	1.350	1.093	1.152	1.139	**1.020**

Table 6: Root Mean Square Error Values

4 Regression Algorithm Results

4.1 Feature Details

4.1.1 Type 2 Feature Set

The patient specific features were extracted by scraping the personal data from patient profiles. We did not scrape profiles of each and every patient but scraped profiles of only those constrained patients: which had evaluated at least three different drugs. This was done as the data was huge and scraping it would take time, so focus on a smaller batch and get results for it.

Details of the features:

Patient Specific Features:
Patient Id, Age, Gender, Primary Conditions

Drug Specific Features:
Drug Id, Purpose, Adherence, Cost, Other Drugs Taken Id
Other Drugs Purpose, Other Drugs Adherence, Other Drugs Cost, Other Drugs Side Effect Rating

All the features are categorical features, so each feature is like an id assigned for that particular category. The features for each patient would vary in terms of the number of Other Drug features. So, for the cases where there are no other drugs, the labels or id assigned is -1.

4.1.2 Type 3 Feature Set

Patient Specific Features:
Patient Id, Age, Gender, Location, Primary Conditions

Drug Specific Features:
Drug Id, Purpose, Adherence, Cost, TimeStamp of Drug, Other Drugs

Again, all the features are categorical features (apart from TimeStamp of Drug) So each feature is like an Id assigned for that particular category.

Time Stamp of a drug is a normalised number between 0 and 1 which represents the evaluation time of that drug with respect to the evaluation time of the latest drug the patient has consumed. It keeps a track of relative time difference between consumption of multiple drugs by the patient.

So, the varying information about the Other Drug features for a patient is expressed in terms of a single feature Time-Stamp. We do not explicitly give purpose, adherence, cost of Other Drugs that a patient might have take, because model would itself learn it as the part of patient features would be similar for

different cases of drugs that a patient would have taken. Additionally, feature like Time-Stamp also keeps a track of relative difference which was not included in the previous set of feature.

A new feature Other Drugs in this set is a binary feature that has value 0 and 1. It takes value 1 for patient-drug pair if the patient had taken any other drug before 1 month of taking the current drug, else it sets to zero. The relative time difference knowledge from the Time Stamp feature helped in generating features like Other Drugs. This binary feature can be important in terms that not only taking some other drug prior to current treatment matters but the time duration before which it was taken also matter.

Feature	Total Distinct Categories
Primary Condition	429
Other Condition	1671
Purpose	1973
Side Effects	3754
Location	3578
Age	5
Gender	3
Other Drugs	2
Cost	6
Adherence	4

Table 7: Feature Statistics

4.1.3 Type 5 Feature Set

The feature set is same as for Type 3 case.
Patient Specific Features:
Patient Id, Age, Gender, Location, Primary Conditions

Drug Specific Features:
Drug Id, Purpose, Adherence, Cost, TimeStamp of Drug, Other Drugs

Feature	Total Distinct Categories
Primary Condition	434
Other Condition	1675
Purpose	2169
Side Effects	2813
Location	3602
Age	6
Gender	3
Other Drugs	2
Cost	6
Adherence	4

Table 8: Feature Statistics

4.2 Type 1, 2 and 3 Results

The results for Side Effect Prediction are obtained by using the complete feature set of each type described as above.

In Feature Selection, the performance of Random Forest Regressor is compared over different subsets of features set for type 3.

4.2.1 Side Effect Prediction

Algorithm	Type 1	Type 2	Type 3
Random Forest Regressor		0.993	**0.848**
Decision Tree Regressor		1.225	0.857
Linear Regression		**0.898**	0.883
Gaussian Naive Bayes Regressor		2.036	0.904

Table 9: Side Effect Prediction: Mean Absolute Error

Algorithm	Type 1	Type 2	Type 3
Random Forest Regressor		1.149	**1.007**
Decision Tree Regressor		1.624	1.026
Linear Regression		**1.063**	1.042
Gaussian Naive Bayes Regressor		2.295	1.384

Table 10: Side Effect Prediction: Root Mean Square

4.2.2 Feature Selection

1. Results for **Random Forest Regressor** for **Type 3 Dataset**
2. Estimators refers to the number of trees in the forest.
3. The bold value in this table represent the Set on which best performance is attained among all the 5 sets.

Estimators	Set1	Set2	Set3	Set4	Set5
100	0.854	0.850	0.851	0.850	**0.848**
300	0.853	0.850	0.851	0.850	**0.848**
500	0.853	0.850	0.851	0.850	**0.848**

Table 11: Mean Average Error Values

Estimators	Set1	Set2	Set3	Set4	Set5
100	1.011	1.008	1.010	1.010	**1.007**
300	1.010	1.008	1.010	1.009	**1.007**
500	1.010	1.008	1.010	1.009	**1.007**

Table 12: Root Mean Square Values

Set1 Features:
Patient Id, Age, Gender, Location, Primary Condition, Drug Id, Purpose, Adherence, Cost, Time of Evaluation, Other Drugs

Set2 Features:
Patient Id, Age, Gender, Primary Condition, Drug Id, Purpose, Adherence, Cost, Time of Evaluation

Set3 Features:
Patient Id, Age, Gender, Primary Condition, Drug Id, Purpose, Time of Evaluation

Set4 Features:
Patient Id, Primary Condition, Drug Id, Purpose, Time of Evaluation, Other Drugs

Set5 Features:
Patient Id, Drug Id, Purpose, Adherence, Cost, Time of Evaluation, Other Drugs

Comments:
1. The best performance is on Set 5, the case with no patient specific features.

4.3   Type 5 Results

Algorithm	Set 1	Set 2	Set 3	Set 4	Set 5
Random Forest Regressor	**0.876**	**0.871**	**0.874**	**0.870**	**0.860**
Decision Tree Regressor	0.887	0.881	0.885	0.876	0.863
Linear Regression	0.902	0.913	0.908	0.910	0.917
Gaussian Naive Bayes Regressor	0.877	0.873	0.874	0.876	0.868

Table 13: Side Effect Prediction: Mean Absolute Error

Algorithm	Set 1	Set 2	Set 3	Set 4	Set 5
Random Forest Regressor	**1.030**	**1.029**	**1.032**	**1.026**	**1.026**
Decision Tree Regressor	1.052	1.047	1.049	1.041	1.030
Linear Regression	1.058	1.066	1.067	1.064	1.072
Gaussian Naive Bayes Regressor	1.386	1.384	1.385	1.385	1.379

Table 14: Side Effect Prediction: Root Mean Square

Set1 Features: All Features
Patient Id, Age, Gender, Location, Primary Condition, Drug Id, Purpose, Adherence, Cost, Time of Evaluation, Other Drugs

Set2 Features: All Important Features
Patient Id, Primary Condition, Drug Id, Purpose, Time of Evaluation

Set3 Features: Patient Specific Features Only
Patient Id, Age, Gender, Location, Primary Condition, Drug Id

Set4 Features: Drug Specific Features Only
Patient Id, Drug Id, Purpose, Adherence, Cost, Time of Evaluation, Other Drugs

Set5 Features: No extra features
Patient Id, Drug Id

Comments:
1. Random Forest Regressor performs best among all algorithms on all 5 sets.
2. The performance of Random Forest Regressor is better on Set 4 and Set 5 where we do not include patient specific features.
3. Best performance for Random Forest Regressor is in Set 5 where do not include any drug or patient specific features

4.4 Side Effect + Rating Prediction

This contains result for Type 5 Dataset for multi label classification.

Model	Iterations	Precision	Recall	F1 Score
Perceptron	1000	0.0004	0.0008	0.0005
Logistic Regression	..	0.0006	0.0011	0.0007

5 Deep Learning Model

5.1 Side Effect Rating Prediction

5.1.1 Details of the Model

Input:
User, Item: Patient Id and Drug Id respectively

Embedding Layers:
Matrix Factorization Embedding : MF User, MF Item
Multi Layer Perceptron Embedding: MLP User, MLP Item

Merge:
Using the features generated for User and Item Id by embedding layers, we merge them to create different features.

User Vector 1: MF User * MLP User
Item Vector 1: MF Item * MLP Item
User Vector 2: MF User + MLP User
Item Vector 2: MF Item + MLP Item
MF vector: MF User*MF Item
MLP Vector: MLP User + MLP Item $->$ Dense Layer (Reduce to MF Dim)

Predict Vector
There are multiple ways to generate a final prediction vector. Listed below the different combinations of prediction vectors tried.

Predict Vector 1: MF Vector * MLP vector
Predict Vector 2: User Vector 1 + Item Vector 1
Predict Vector 3: User Vector 2 + Item Vector 2
Predict Vector 2*: Predict Vector 2 $->$ Dense Layer (Reduce to MF Dim)
Predict Vector 3*: Predict Vector 2 $->$ Dense Layer (Reduce to MF Dim)

Prediction 1:
Dense Layer: Predict Vector $->$ Side Effect Rating

5.1.2 Type 5 Results

Notation
MF Dim: Output Dim of MF Embedding Layer
Vec : Predict Vector being used for Prediction 1

MF Dim	Vec	Type 4	Type 5	Type 6
64	1	0.844	0.834	
32	1	0.844	0.831	
16	1	**0.840**	0.828	
8	1	0.842	0.810	
8	2		0.798	
16	2		0.795	
16	3		0.797	
64	3		0.797	
32	2*		**0.774**	
128	2*		**0.774**	
32	3*		0.776	

Table 15: Mean Absolute Error Values

MF Dim	Vec	Type 4	Type 5	Type 6
64	1	**1.014**	**1.018**	
32	1	1.017	1.023	
16	1	1.021	1.029	
8	1	1.035	1.054	
8	2		1.096	
16	2		1.104	
16	3		1.105	
64	3		1.108	
32	2*		1.093	
128	2*		1.107	
32	3*		1.090	

Table 16: Root Mean Square Error Values

5.2 Side Effect + Rating Prediction

For predicting the Side effect along with the rating, we implement a different deep learning model.

First, the idea behind predicting Side Effect: We maintain a vocabulary of all the Side Effects from the training data and then for the cases of patient-drug test data pair, we predict a probability for the occurence of each side effect in the vocabulary. So, it is like a Multi-Label Classification problem where each patient-drug pair has a list of side effect labels assigned to it.

5.2.1 Details of the Model

Input:
User, Item: Patient Id and Drug Id respectively

Embedding Layers:
Matrix Factorization Embedding : MF User, MF Item
Multi Layer Perceptron Embedding: MLP User, MLP Item

Merge:
Using the features generated for User and Item Id by embedding layers, we merge them to create different features.

User Vector: MF User * MLP User
Item Vector: MF Item * MLP Item
MF vector: MF User*MF Item
MLP Vector: MLP User + MLP Item
Dense Layer: MLP Vector $->$ MLP Vector (Reduce dimension by half)

Predict Vector
There are multiple ways to generate a final prediction vector. Listed below the different combinations of prediction vectors tried.

Predict Vector 0: MF Vector * MLP vector
Predict Vector 1: MF Vector + MLP vector
Predict Vector 2: MF Vector $->1$ + MLP vector $->1$
Predict Vector 3: MF User $->1$ + MF Item $->1$
Predict Vector 4: User Vector $->1$ + Item vector $->1$
Predict Vector 5: User Vector + Item vector
Predict Vector 6: User Vector $->1$ + Item vector

Prediction 1:
Dense Layer: Predict Vector $->$ Side Effect Rating

Prediction 2:
1. Dense Layer: Predict Vector $->$ Predict Vector
2. Dropout : Predict Vector $->$ Predict Vector
3. Batch Normalisation: Predict Vector $->$ Predict Vector
4. Dense Layer: Predict Vector $->$ Side Effect Vector

5.2.2   Type 5 Results

There are two tables, both correspond to different usage of Loss Functions.

Notation
Dim 1: Output Dim of MF Embedding Layer
Dim 2: Output Dim of Prediction 2 Dense Layer 1
Vec 1: Predict Vector being used for Prediction 1
Vec 2: Predict Vector being used for Prediction 2

Case 1:

Loss Function: MAE (Prediction 1), Binary Cross Entropy (Prediction 2)

Case	Vec 1	Vec 2	Dim 1	Dim 2	MAE	RMSE	F1 Score	Precision	Recall
No Cutoff	2	Mlp	8	100	0.841	1.272	0.568	0.566	0.587
No Cutoff	2	Mlp	64	100	0.850	1.253	0.550	0.606	0.512
Cutoff 1	2	Mlp	64	100	0.841	1.249	0.539	0.588	0.543
Cutoff 1	2	0	64	100	0.800	1.094	0.453	0.585	0.506
Cutoff 1	2	1	64	100	0.836	1.212	0.461	0.615	0.469
Cutoff 1	Mf	Mf	64	100	0.859	1.367	0.524	0.533	0.559
Cutoff 1	3	3	64	100	0.840	1.162	0.579	0.579	0.583
Cutoff 1	4	4	64	100	0.875	1.128	0.428	0.613	0.402
Cutoff 1	4	5	64	100	0.858	1.204	0.516	0.590	0.489
Cutoff 1	4	5	8	100	0.797	1.110	0.548	0.634	0.483
Cutoff 1	4	5	8	500	0.796	1.109	0.525	0.648	0.443
Cutoff 1	5	5	8	500	0.769	1.092	0.537	0.641	0.464
Cutoff 2	5	5	8	500	0.761	1.088	0.543	0.645	0.469
Cutoff 2	6	6	8	100	0.771	1.101	0.531	0.653	0.449
Cutoff 2	7	7	8	100	0.834	1.303	0.585	0.560	0.630
Cutoff 1	8	8	8	100	0.858	1.367	0.593	0.533	0.669
Cutoff 1	9	9	8	100	0.774	1.106	0.555	0.628	0.499

Case 2:
Loss Function: MAE (Prediction 1), Recall (Prediction 2)

Case	Dim 1	MAE	RMSE	F1 Score	Precision	Recall
No Cutoff	8	0.838	1.331	0.002	0.0008	0.471
Cutoff 1	8	0.835	1.300	0.003	0.001	0.476
Cutoff 2	8	0.815	1.258	0.005	0.002	0.540
Cutoff 1	32	0.942	1.093	0.003	0.002	0.491