

Prediction of Adverse Side-Effects using recommender-system

Diogo Pinto and Pedro Costa *

INESC Tec.

Abstract. Using one of the best algorithms in use for recommender systems, we find the most probable adverse side effects that a drug may have in addition to a set of ones previously detected. The efficiency and dependence on previously known adverse side effects of this algorithm may be further boosted by adding the most important chemical structures. This enables drugs industry to predict in side-effects even before the drug entering the market, saving millions of dollars in the process.

1 Data

ADReCS¹ is a database, maintained by Xiamen University, of adverse drugs' reactions ontologies, that enables the standardization and hierarchization of ADR terms.

This database divides in two: one with the information on drugs and ADRs related, and another with ontologies about ADRs and drugs that may lead to a certain ADR. Both are originally kept in XML format, but were imported into an relational database for efficiency of querying.

Drugs Each drug has:

- Id *e.g.*: *BADD D00002*
- Name *e.g.*: *Abarelix*
- Description
- ATC (Anatomical Therapeutic Chemical classification system) *e.g.*: *L02BX01*
- Synonyms list *e.g.*: *Plenaxis, Plenaxis depot, PPI-149, Abarelix-Depot-M.*
- Indications
- CAS (Chemical Abstracts Service registry number) *e.g.*: *183552-38-7*
- ADR List. Each ADR has:
 - Term *e.g.*: *Hypersensitivity*
 - Id *e.g.*: *10.01.03.003*
 - Frequency *e.g.*: *> 5%* (this information may be omitted)

* Special thanks to Prof. Vtor Santos Costa and Prof. Rui Camacho

¹ <http://bioinf.xmu.edu.cn/ADReCS>

ADRs Each ADR has:

- Id *e.g.*: 01.01.01.003
- Term *e.g.*: Factor I Deficiency
- Synonyms List.
- Description
- Code Meddra *e.g.*: 10016075
- List of drugs that may lead to this ADR. Each drug entry has:
 - Id *e.g.*: BADD D00937
 - Name *e.g.*: Oxytocin

1.1 Data Structure

The data structure used was derived from the ADReCS. From this database was built a matrix, with the **rows** being the drugs and **columns** being the ADRs.

For a given row (*i.e.* drug), a value of 5 is given to a column if that drug has the corresponding ADR. The value 5 is used so that we are able to scale (*i.e.* measure of its importance) the importance of the features that characterize the drugs. That enables that, in the future, if we add to the columns more features from a different nature (*e.g.* main chemical substructures), it is possible to give a different scale, higher or lower than the first one.

2 Singular Values Decomposition

The singular value decomposition is a factorization of a real or complex matrix.

Formally, the singular value decomposition of an $m \times n$ real or complex matrix M is a factorization of the form $M = U\Sigma V^T$, where U is an $m \times \min(m, n)$ real or complex unitary matrix, Σ is an $\min(m, n) \times \min(m, n)$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and V^T (the conjugate transpose of V , or simply the transpose of V if V is real) is an $\min(m, n) \times n$ real or complex unitary matrix.²

2.1 Dimensionality Reduction

There is a technique which consists on removing the rows and columns from U , Σ and V that corresponds to the lower singular values (constant in the Σ matrix), enabling to lower the dimension of the data while abstracting the features to a higher level. The criteria for stopping removing columns and rows is

$$\sum_{i=1}^k NewSigma[i, i]^2 \geq 0.9 \times \sum_{i=1}^{\min(m, n)} OriginalSigma[i, i]^2$$

² http://en.wikipedia.org/wiki/Singular_value_decomposition

After this step, M is a $m \times k$ matrix, Σ is a $k \times k$ matrix, and V is a $n \times k$, where $k < \min(m, n)$.

It is possible to multiply the obtained matrices in the same way as before and identify that the obtained matrix is very similar to the original one.

3 Gradient Descent

After obtaining the SVD, we obtain $P = M \times \sqrt{\Sigma}$ and $Q = V \times \sqrt{\Sigma}$, so that we only have two matrices to work with, keeping all the information intact.

This two matrices are approximated to the original one using gradient descent, by minimizing the distance for each element of the $P \times Q^\top$ to its desired value.

It starts with a predefined learning rate, and if the distance from the obtained matrix from the original is greater than in the last iteration, it is lowered by a factor of 3.³

4 Resulting artifacts

From the sequence of steps described above, we obtain two main matrices. For normalization, we have to perform

$$P = P \times \sqrt{\Sigma}^{-1}$$

and similarly for Q .

5 Possible operations

The following operations are possibly applied to the inverse cases.

- By computing the cosine distance between two rows (drugs), it is possible to determine the similarity between them;
- Predict ADRs. By $Drug \times Q$, we obtain the latent factors for this drug, and, as such, $Drug \times Q \times Q^\top$ gives the prediction of ADRs for that drug.
- By changing how many rows we remove in the dimensionality reduction step, it is possible to obtain clusters of ADRs (by selecting, for each latent factor of a given ADR, the one that presents the highest value), and next determine the most prominent ADR from the group (by selecting the ADR that has the highest latent value from them all)

³ Although implemented, this step has only 2 iterations, as the previous step produces already a very optimized solution. By adding new features, it is possible that the relevance of this step is enhanced.

6 Testing

6.1 Real World Simulation

For simulating what really happens, we provide a test set (independent from the training set), and remove randomly between 0.3 and 0.7 of its ADRs. By running this for a random row, for a given number of iterations...