International Journal of Community Medicine and Public Health
Chen AW. Int J Community Med Public Health. 2018 Mar;5(3):901-904
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: http://dx.doi.org/10.18203/2394-6040.ijcmph20180744

Original Research Article

Predicting adverse drug reaction outcomes with machine learning

Andy W. Chen*

University of British Columbia, 2053 Main Mall, Vancouver, BC, V6T 1Z2, Canada

Received: 31 December 2017
Accepted: 02 February 2018

***Correspondence:**
Dr. Andy W. Chen,
E-mail: andywchenca99@gmail.com

ABSTRACT

Background: Adverse drug reactions are a drug safety issue affecting more than two million people in the U.S. annually. The Food and Drug Administration (FDA) maintains a comprehensive database of adverse drug reactions reported known as FAERS (FDA adverse event reporting system), providing a valuable resource for studying factors associated with ADRs. The goal of the project is to build predictive models to predict the outcome given patient characteristics and drug usage. The results can be valuable for health care practitioners by offering new knowledge on adverse drug reactions which can be used to improve decision making related to drug prescriptions.
Methods: In this paper I present and discuss results from machine learning models used to predict outcomes of ADRs. Machine learning models are a popular set of models for prediction. They have gained attention recently and have been used in a variety of fields. They can be trained on existing data and retrained when new data become available. The trained models are then used to make predictions.
Results: I find that the supervised learning models are work similarly within groups, with accuracy between 65% and 75% for predicting deaths and 70% to 75% for predicting hospitalizations. Across groups the models predict hospitalizations better than deaths.
Conclusions: The predictive models I built achieve good accuracy. The results can potentially be improved when more data become available in the future.

Keywords: Drug-related side effects, Adverse reactions, Adverse drug reaction reporting systems, Drug interactions, Statistical models

INTRODUCTION

Adverse drug reactions (ADR), defined as 'appreciably harmful or unpleasant reaction[s], resulting from an intervention related to the use of a medicinal product' by Edwards et al affect millions of people worldwide.[1] ADRs are a huge burden of financial resources and labor. It was estimated by Lazarou et al that more than 100,000 die from ADRs annually.[2] Moreover, Sultana et al found that around $30.1 billion are spent annually on ADRs in the U.S., and nearly half of these costs can be prevented based on a study by Bates et al.[3,4] The unnecessary high financial costs and labor spent on ADRs provide strong motivation to learn more about ADRs and be able to predict the probability of ADRs accurately.

The project's goal is to build predictive models that can predict ADR outcomes given patient demographics and drug prescription information with good accuracy. These models include supervised machine learning models logistic regression, support vector machine, as well as ensemble models random forest and gradient boosted tree. These models can then be used predict outcomes on an individual basis.

Related works on ADR include one by Kadoyama et al, who found strong association between hypersensitivity

and anticancer agents.[5] In another study, Gurwitz and Avorn found that physiological and functional characteristics of patients to be strong predictors of ADR outcomes.[6] Lassila et al studied the relationship between patient demographics and ADRs and found that more elder people suffer from ADRs due to inappropriate prescriptions.[7] Cooper found that patients with ADRs take an average of 7.8 drugs in contrast to 3.3 among patients without ADRs, concluding that there exist significant associations between ADRs and number of drugs taken.[8] In another study using ensemble machine learning methods, Tsymbal et al found significant relationship between antibiotic resistance and ADRs.[9]

METHODS

I use the FAERS data maintained by FDA each quarter from 2012 to 2017. For each quarter, seven datasets are available: DEMO (patient demographic and administrative information), DRUG (drug/biologic information), REAC (adverse events), OUTC (patient outcomes), RPSE (report sources), THER (drug therapy start and end dates), and INDI (indications for use for the reported drugs). There are around 4 million events with outcomes reported, where each event is tracked by a unique identifier called primary id. To create the dataset used by the model, I first merge all events reported for the same patient. This creates the total number of drugs used by each patient. I then standardize the continuous variables, age, weight, and drug dosage, by subtracting the mean from each value and dividing by the standard deviation. I also create dummy variables for categorical variables drug role, drug name, and route of drug intake.

Once the feature set and target variables are extracted and formatted, I build supervised machine learning models. In particular, I use logistic regression, support vector machine, random forest, and gradient boosted tree. The features are mentioned above and the target variable is the outcome of an ADR event. It is possible for an event to have multiple outcomes associated with it. For example, a patient may be hospitalized and then die. In other words, the outcome categories are not mutually exclusive for an event, so multi-class models will not work. Therefore, I build models separately for each outcome type. In other words, I build a model for predicting the probability of death, another model for predicting hospitalization, and so on. I evaluate the models using accuracy, precision, recall, and F1 score.

I predict the possibility of each outcome for each patient using a variety of features. From the DEMO dataset, I use a patient's age, weight, sex, and country of treatment. From the DRUG dataset, I include the drug's role, drug name, route of drug intake, and dosage. I filter out the events that do not have all the features available. To create the dataset used by the model, I first merge all events reported for the same patient. This creates the total number of drugs used by each patient. I then standardize the continuous variables, age, weight, and drug dosage, by subtracting the mean from each value and dividing by the standard deviation. I also create dummy variables for categorical variables drug role, drug name, and route of drug intake.

Inspecting the dataset reveals that it is unbalanced, especially for the death outcome. There are only 9.6% of positive instances (deaths) in the data. Training models on unbalanced data yields unreliable results due to the large number of negative instances (non-deaths). Our initial results show that such unbalanced models have high accuracy, but very low precision, recall, and F scores. The high accuracy is due to the high number of correct predictions of negative data points. However, precision and recall depend on true positives. Low accuracy on predicting positive data points will yield low precision and recall. F score depends on precision and recall, so it is also low for unbalanced data. Since it is preferable to predict deaths than obtain high classification accuracy, it is more important to have a higher sensitivity and F score. My solution is to balance this dataset by taking several random subsamples of the dataset with replacement. Each sample contains an equal proportion of positive and negative instances. I train models on these balanced subsamples using an ensemble approach. That is, I train separate models using each random balanced subsample. The prediction is determined by taking the majority vote of these models as in an ensemble. The same approach is used to predict the test set. The metrics are reported for the test set.

Once the feature set and target variables are extracted and formatted, I build supervised machine learning models. In particular, I use logistic regression, support vector machine, random forest, and gradient boosted tree. The features are mentioned above and the target variable is the outcome of an ADR event. It is possible for an event to have multiple outcomes associated with it. For example, a patient may be hospitalized and then die. In other words, the outcome categories are not mutually exclusive for an event, so multi-class models will not work. Therefore, I build models separately for each outcome type. In other words, I build a model for predicting the probability of death, another model for predicting hospitalization, and so on. I evaluate the models using accuracy, precision, recall and F1 score.

RESULTS

I start with some data exploration. For the demographics of patients in 2012 and 2017, 46.6% males and 53.4% females, so ADRs affect both genders roughly equally. Figures 1 and 2 show the distribution of age and weight. Most patients are in the 50 to 90 year old range with a mean of 59 years and weigh 50 to 100 kg with a mean of 72 kg.

I now present the results of the models. As mentioned before, each model is for predicting one type of outcome. Table 1 shows the models for each of the three most common outcomes evaluated by accuracy, precision, recall, and F1 score.

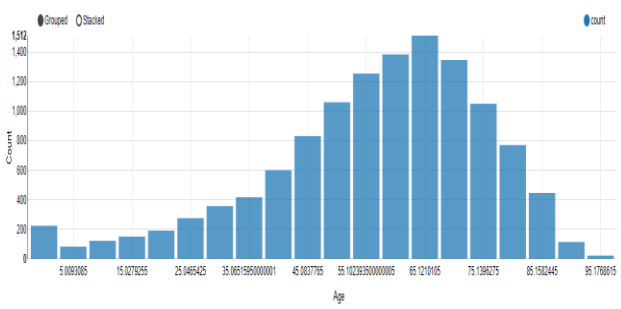

Figure 1: Distribution of patient age.

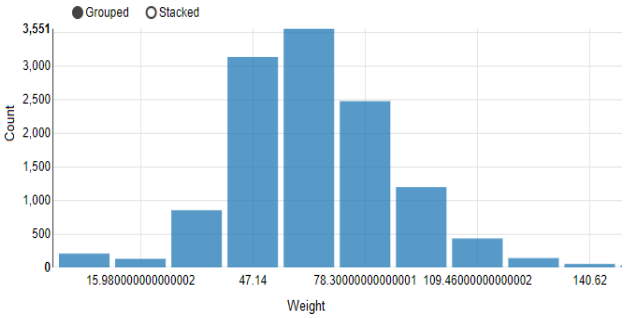

Figure 2: Distribution of patient weight.

Table 1: Metrics of supervised learning models for three most common outcomes.

Outcome	Model	Accuracy	Precision	Recall	F1 Score
Death	Logistic Regression	0.76	0.78	0.71	0.74
Death	SVM	0.69	0.71	0.64	0.67
Death	Random Forest	0.73	0.76	0.68	0.71
Death	Gradient Boosted Tree	0.68	0.75	0.53	0.62
Hospitalization	Logistic Regression	0.75	0.77	0.90	0.83
Hospitalization	SVM	0.73	0.73	0.95	0.83
Hospitalization	Random Forest	0.74	0.73	0.99	0.84
Hospitalization	Gradient Boosted Tree	0.74	0.73	0.97	0.84

DISCUSSION

Overall, the models have good predictive power for all three outcome categories. The metrics are much improved from the experimental results using unbalanced data, which have high accuracy but low precision, recall, and F1 score below 50%. Here, with balanced data, most metrics are above 75% and the F1 scores are mostly in the 80's. Within each outcome category, the models have similar performance. Across the categories, the models predict Hospitalization and Other better than Death. This could still be due to the relative few number of positive instances for deaths. Because of that, each subsample has a relatively small number of data points, despite that they are balanced samples. Training models with small number of data points may result in over-fitting, or less accurate prediction on the test set and real world data. This is a limitation in this data set, but can be improved when more data become available in the future.

CONCLUSION

This paper presents the results of logistic regression, a supervised machine learning model for predicting ADR deaths based on patient demographics and drug usage. The model shows good predictive power when trained using balanced samples. Future extensions may include using other models such as support vector machines to compare the results of the model. Also, when more data become available in the future, it would be interesting to retrain the models and see how the prediction metrics change.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet. 2000;356:1255-9.
2. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. Journal of the American Medical Association. 1998;279(15):1200-5.
3. Sultana J, Cutroneo P, Trifiro G. Clinical and economic burden of adverse drug reactions. J Pharmacol Pharmacother. 2013;4(1):S73-7.
4. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA. 1997;277:307–11.
5. Kadoyama K, Kuwahara A, Yamamori M, Brown JB, Sakaeda T, Okuno Y. Hypersensitivity reactions to anticancer agents: data mining of the public version of the FDA adverse event reporting system, AERS. Journal of Experimental & Clinical Cancer Research. 2011;30:93.
6. Gurwitz JH, Avorn J. The ambiguous relation between aging and adverse drug reactions. Ann Intern Med. 1991;114:956–66.
7. Lassila HC, Stoehr GP, Ganguli M, Seaberg EC, Gilby JE, Belle SH, et al. Use of prescription medications in an elderly rural population: the

MoVIES Project. Ann Pharmacother. 1996;30:589–95.
8. Cooper JW. Probable adverse drug reactions in a rural geriatric nursing home population: a four-year study. J Am Geriatric Soc. 1996;44:194–7.
9. Tsymbal A, Pechenizky M, Cunningham P, Puuronen S. Dynamic integration of classifiers for handling concept drift. Information Fusion. 2008;9(1):56-68.

Cite this article as: Chen AW. Predicting adverse drug reaction outcomes with machine learning. Int J Community Med Public Health 2018;5:901-4.