Monday 11th September 2017

Stream: Health

Calculating a Severity Score of an Adverse Drug Event Using Machine Learning on the FAERS Database

Rob Schumaker, Michael Veronin, Rohit Dixit, Pooja Dhake, Danya Manson
University of Texas Tyler
USA
rob.schumaker@gmail.com, mveronin@uttyler.edu, rdixit@patriots.uttyler.edu,
pdhake@patriots.uttyler.edu, dmason4@patriots.uttyler.edu

An Adverse Drug Event (ADE) is medical injury that can result from a prescription or over the counter drug that causes an allergic reaction, overdose, reaction with other drugs or is the result of a medication error (health.gov 2017). There are approximately two million ADEs every year (Lazarou, Pomeranz, and Corey 1998) with 1,320,389 reported to the FDA in 2015 alone. ADEs are directly responsible for over 100,000 deaths annually and is the fourth leading cause of death with the number of reported deaths increasing as shown in Figure 1.

Aside from the number of ADEs occurring, another important aspect of ADEs is the medical costs associated with hospital visits, rehabilitation and follow-up care. It is believed that adverse drug events cost $4.7billion annually (Romeril 2017) and could be as high as $30.1billion when taking into account unreported events (Sultana, Cutroneo, and Trifirò 2013). Many of these costs can be attributed to hospital admissions where ADEs are linked to 148,000 admissions annually and 6.5% of all admissions in the UK are ADE-related (Davies et al. 2009). Furthermore, mean hospital stays increase from 8 to 20 days for ADE-related admissions.

Vulnerable populations such as children and the elderly are most susceptible to ADEs. It is estimated that between 2.1% and 5.2% of hospital admissions for children were ADE-related where 39% of those admissions were either life-threatening or fatal (Impicciatore et al. 2001). For the elderly, it is estimated that between 11.4% and 35.5% of admissions were ADE-related (Budnitz et al. 2007) with the numbers climbing to 32%-65% in nursing homes (Cooper 1996; Cerety et al. 1993).

To address the problem of post-marketed drugs causing ADEs, the US Food & Drug Administration (FDA), created a repository in 1969 to collect and disseminate adverse reaction data. This repository, the FDA Adverse Reporting System (FAERS) has undergone several transformations over the years, however, its mission remains the same. To voluntarily collect adverse reaction data on drugs, medication errors and adverse events concerning medical devices from manufacturers, medical professionals and the general public. Many of the records are directly submitted by pharmaceutical manufacturers, however, anyone can submit data using MedWatch (Form FDA 3500) which collects patient demographic data, the adverse event, product availability, suspect products, suspect medical device, other medical products that might be involved and information about the reporter. FAERS data is publicly available1 and contains nearly 112,000,000 records over the past 12 years. This makes it difficult to find salient information between drugs that may be adversely affecting one another or finding those drugs that are connected to a disease state. Also, because of the variety of inputs into the FAERS system, many fields contain a non-trivial amount of non-standardized data such as misspelled drug names, use of multiple brand names for the same drug, abbreviations, extraneous information, excessive punctuation/formatting, mixed capitalizations and non-descriptive data such

as “unknown purposes.” This lack of standardized data has kept FAERS from fulfilling its full potential as a pharmacovigilance tool and its limitations have been the subject of numerous studies, which will be discussed later.

Early and accurate identification of ADEs and their severity is critically important for public health. Rather than waiting for sufficient post-market evidence to accumulate for a given ADE, a predictive approach that relies on leveraging existing, contextual drug safety information would have the potential to identify certain ADEs earlier.

Some studies have begun to use FAERS data in a bid to construct their own pharmacovigilence systems, however, many of these studies examine specific bivariate drug associations rather than analyze all historical data using data mining algorithms to determine the potential of previously unknown drug interactions.

Our motivation is to improve drug safety by creating a new type of pharmacovigilence system that 1. Performs data cleaning and standardization of FAERS data, 2. Computes a drug reaction severity score for each ADE based on the reported indications and coded using a modified Hartwig Severity scale, 3. Models the data to A) empirically identify drug-interaction events and their relative strength of event in specific symptom-related incidents and to B) identify drug-disease event severity for specific indications such as hypertension, stroke and cardiac failure, 4. Computes a predicted severity score for the models using machine learning algorithms such as Back-Propagation Neural Networks, Extreme Learning Machines, Tensor Deep Stacking Networks and Support Vector Regression, 5. Evaluates the accuracy of the predicted severity score versus actual severity on a holdout dataset, and 6. Builds a predictive clinical tool for physicians that can interact with a patient’s EHR and identify adverse reaction potential at the point of prescription.

We propose a global data-driven approach with the TylerADE System. This system uses advanced machine-learning techniques to sift through data and uncover potentially unknown drug events. This research has the potential to 1) improve the efficiency of pharmacological research by identifying potentially unknown n-drug events that merit further study; 2) create a risk score of potential medication events that physicians can use in a clinical setting; and 3) improve patient safety.

Keywords: Pharmacovigilance; BioInformatics; Machine Learning; Adverse Drug Events