Teaching Pharmacovigilance to Machines

Provided by Tata Consultancy Services

Authors - Dr. Ashish Indani, Aparna Patkar, Devraj Goulikar, Prita Venkateswaran and Divya Vasudevan

Introduction

The horizon of Artificial Intelligence (AI) is broadening in every field with its presence now felt in the medical field as well. The application of AI now caters to diverse clinical and non-clinical requirements of the medicine industry and its offshoots i.e. life sciences and healthcare. While AI Systems are aiding in multiple healthcare aspects like medical image processing; going forward, the process of vigilance especially pharmacovigilance would find the most promising application. Currently, multiple applications in the industry leverage AI in pharmacovigilance including the recently deployed real world processing of the pharmacovigilance (PV) cases.

In an attempt to bring automation, PV underwent multiple static transformations guided by human-built and algorithm induced logic, from macros to previous generation of extensive rule based robotic automation. However, dynamic AI systems differ from static ones in multiple ways including progressive logic building and continuous learning for development of own intelligence. Hence, robotic and point automations majorly suit to stereotype work, and AI systems tend to take-over most of the bulk of case processing and aggregate reporting, contributing to faster and precise informed decisions by PV professional (PVP). However, PVPs must establish a symbiotic relationship with sophisticated AI systems for effective co-working. The foundational aspect of this relationship is the process of learning. In pharmacovigilance with AI (AI@PV), both human and machine learning are interdependent and of utmost importance.

PVP training

Training of PVPs include three main sources, medical or paramedical education, PV training and on-the-job experiences. Hence, a PVP processes a case by extracting, applying knowledge and executing multiple complex transformations to complete entire case for submission. The transformations include Medical Coding, Analysis of case validity, seriousness, causal relationship between drugs and adverse events, etc. In addition, it involves a series of decision-making before reaching the final transformational expression (figure 1).
Machine Training on Pharmacovigilance

Training AI@PV system is also an extremely complex process. Unlike humans, machines do not have a qualification in medicine or paramedical sciences. Hence, AI@PV system needs to complement PVPs by mimicking process, accomplished by assessing relationship between components of natural intelligence (human) and artificial intelligence. Natural intelligence comprises of memory, processing and expression. Amongst these, processing is extremely complicated since, at a time, there are more than 100 processes running in parallel in a human brain to reach a decision. The intelligence continues to improvise by development of logic aided by reminiscence, contextualization and application. Higher intelligence, is majorly characterized by ability to make complex decision in shorter time and effort. This journey of humans evolve from a neophyte’s attempt to understand the subject to an established level of understanding, enabling decision-making capability through “learning”. Natural intelligence augments learning mainly by conditioning and instruction. The primary modes of learning for a human hence includes (1) Associative learning leading to classical conditioning, (2) Consequential learning, leading to operant conditioning and (3) Observational learning leading to development of logic and memory\(^4\).
AI systems are expected to possess abilities of par-human decision-making and continuous learning. Machine learning, an AI based system that has the ability to automatically learn and improve from experience, is comparable with the human learning process. However, there are fundamental difference between the natural Intelligence and artificial intelligence, with respect to its composition. Hence, the process of learning is an extremely critical aspect for AI systems. Machines, however, miss out on the parallel aspect of mind, emotion, conditioning, etc. Hence, machine learning is majorly dependent upon observational processing. The conditioning is partially covered by technology, enabling contextualization. The default memory aspect is missing in a typical AI system, which needs to be established with aspects such as ontology. (Figure 2)

In AI systems, models behave as a unit of intelligence. Hence, effective learning management plays a key role in development and maintenance of highly accurate and performing AI system to maintain high model credentials. In the PV system, multiple Models are at the core, e.g.:
1. **Categorization model**: categorizes the document into various types
2. **Medical terminology model**: identifies and codes medical terms
3. **Drug nomenclature model**: identifies and codes medical terms
4. **Case classification model**: identifies validity and seriousness of the case
5. **Causal relationship Model**: establishes the relationship between drug and adverse incident

The models and ontologies like MedDRA, WHODD and DDI, products ontology, study index and others function to process and complete the whole one-touch case. (Figure 3)

AI @ PV are complex systems. Hence, their learning and training management is a critical aspect. Effective learning management includes described objectives, methods and desired outcomes of the learning. Key objective of this learning is to improve the system’s accuracy, precision and recall, achieved through multiple Machine Learning modes (figure 4) resulting in Improvement of-

1. ontologies and dictionaries
2. NER patterns
3. transformation logic like medical coding, test report interpretation and case classification
4. complex decision like causal relationship

---

**Figure 4 Modes of Machine learning**

**Machine Training Methodology**

In order to build an effective AI@PV system, learning has to incorporate three major phases, similar to human’s learning - from basic vocabulary to PV expertise.

1. **Baseline phase** (similar to basic medical vocabulary) – trains Medical NLP for Named PV Entity Extraction
2. **Enhancement phase** (for basic PV Understanding) – trains for identification and transformation of special medical terms, drugs and medical coding
3. Proficiency phase – real-time in production training including complex decision

Baseline phase
The baseline training includes creation of a corpus and various AI models for disease, therapy and drug administration from medical literature (e.g. PubMed) and historical production data (e.g. AERS). The corpus subsequently undergoes compression using standard ontology like MedDRA, COSTART (Figure 5). The compressed corpus and models are orientated for medical knowledge using outcome-based training with NLP techniques to identify the similarities of words and terminologies (e.g. word2Vec, Know your Neighbor (KNN), n-grams, etc.). (Figure 6)

![Figure 5 spread of words in the corpus for training](image)

Most similar word for ‘pulmonary’
- ‘lung’, 0.6447716951370239
- 'static_inflation', 0.6369317770004272
- 'atelectasis', 0.6350029706954956
- 'alveolar_proteinosis', 0.6312357187271118
- 'interstitial_pneumonia', 0.6299742460250854
- 'lung_parenchymal', 0.6234893798828125
- 'interstitial_and_alveolar', 0.620290994644165
- 'bronchial', 0.6184309720993042
- 'bronchial_obstruction', 0.617244958875635
- 'respiratory', 0.6167011260986328

![Figure 6 representative outcome of word vectors with the baseline model trained](image)

However, this layer includes limited words. Hence, this model remains vague and less accurate.

Enhancement phase
Actual calibration and preparation of the AI system takes place in the Enhancement layer. In this layer, multiple cycles of training, testing and retraining adds to the accuracy of the model. There are three main aspects of preparation in this layer.

1. Training on large PV data (AERS, historical data from production, other public sources)
2. Identifying and deploying standard resources
   a. Input – standard formats like CIOMS, E2B, Mail, Literature, Clinical trial AE forms, etc.
   b. Processing aid - ontology, dictionaries, rules and convention statements
   c. Outputs - E2B, CIOMS,
3. Reviewing and retraining from the business SMEs
After a few iterations, the system acquires better accuracy, precision & recall and can further be enhanced by providing additional dictionaries and ontologies for improvement of the model. Processes like medical coding, check on validity & seriousness of the case, causal relationship between drug & event are trained and enhanced in the system. Hence, the system gets ready for deployment after a stable accuracy, precision and recall are achieved. The system is also trained on some algorithmic decision-making with use of dictionaries. (Figure 7)

Proficiency Phase

Proficiency phase is a layer of training the system on real-time basis and covers the remaining lifetime of the system and trains on continuous basis. Hence, this is machine learning in PV in its truest sense. In this layer, training happens in multiple ways.

1. Learning from observation – Learning originates from system’s own observation while processing the case. This is multi-instance, multi-tasked, active learning that includes recognition of new NER patterns, new synonyms for drugs, medical terms etc.
2. Learning from reconciliation – Learning originates from comparing machine output with human processed / checked outputs and their trends. This is Statistical Inference based learning that includes learning new conventions, word relationship or named entity relationships, contexts, user understanding etc. based upon deductive and transductive logic.
3. Learning identified by humans – Learning is initiated through PV professionals by comparing the discrepancies or identifying errors and performing its root cause analysis. This is a typical supervised learning model, based upon inductive statistical logic. This type of learning is more towards enhancement of ontology, induction of some conventions or rules, highly skilled thought algorithms like case merger, symptom conjugation to a nosological head, etc. This method helps rapid enhancement of the system for better accuracy and deterministic intelligence.

Figure 7 example of algorithmic compound training of system for medical coding

PV being highly regulated, AI system requires it to qualify par-human regulatory standards. This includes
a. Qualifications – Human qualification is justified by education, training and experience records. The system must be able to justify equivalency of qualification by well-maintained and recorded learnings, reviewed and approved by qualified PV professionals.

b. Reasoning - During audits and inspections, humans can justify their decisions, from memory and logic. Similarly, machines need to produce detailed (not just audit trail) log of all decisions and processes. Hence, the standardization is facilitated by invoking control of PVP on facilitating deep learning of the system (Figure 8), so PVPs can validate the learning with the source and decide to approve or reject, supported by statistical analysis of each identified learning.

Figure 8 Natural Language processing and rule mining

Cross-skilling PVPs on Machine Interaction

AI@PV systems depend significantly on a PVPs for various procedures like learning approvals, change requests and monitoring. Therefore, man to machine interaction needs to be seamless, that are majorly enabled through PVP in various roles such as

a. Techno-domain ‘Learning Champion’ (LC): reviews and approves learning based upon evidence, root-cause analysis (RCA), statistical evaluation, etc.

b. AI@PV case processor (PVP@AI): Receives AI processed E2B case, reviews, QC’s and makes requisite changes and feed in to reconciliation process).

However, PVPs need to take up some digital skills to act as tutors to AI@PV machines. The LC needs to acquire specialized skills pertaining to RCA, ML, understanding of NER pattern / rule etc.

For example, machine presents a verbatim “A 54-year-old man developed TEN 4 week after beginning lamotrigine” with its respective rule as “Subj_Patient; Act_Developed; Object_Symptom; CausePrep_After_Drug”. This means

A. Subject of the verb ‘developed’ is always a patient
B. Object of the verb ‘developed’ is always a symptom
C. Causal relation of preposition ‘after’ with verb ‘developed’ is a medicinal product.

The LC checks appropriateness of presented part of speech (Subject) in the verbatim and named entity in the pattern (Patient). When LC finds the appropriate pair (Subject, Patient) and proper context of the sentence, this learning is approved.
AI@PV performs case intake, triaging, data entry, medical coding, medical review and causal analysis in one-step generating an E2B xml (typical E2B fields and additional fields as per the customer’s requirement). For a PVP, this resembles a typical E2B case (e.g. EMA case) which is ready for review, checked for duplicity, quality and if everything is correct, submitted. However, one-step processing has a mystic ‘black box’, requiring decoding for regulatory and comprehension purposes which is enabled by the LC and PVP@AI, with their detailed analysis and resonance check with system’s logic and process.

For humans to interact with the machine, humans need to connect with the machine. The LC and PVP@AI are required to understand how machines operate, learn and process, unravel each subconscious effort and associate it with the machine process (Figure 9). Human brain performs many complex activities including decisions within a process subconsciously, while the AI machine performs four major activities.

- **Extraction** – Perform NER by reading document, identifying and populating the information in the named fields
- **Derivation** – Populate some more information formula-based calculated information e.g. age from date of birth
- **Determination** – Look up and populate the information e.g. city or country name from the ZIP code, Drug Coding
- **Decision** – Make complex decision of selection, transformation of the data from one field to populate based upon learnings by the system e.g. MedDRA coding, causal analysis, Event listedness in labeling etc.

**Figure 9 comparator of view - Machine and human case processing from human’s eye**

**Discussions**

Pharmacovigilance is a critical process in Life sciences, not only due to its regulatory importance but also because of the need towards having a qualified and experienced personnel. The extent and effort of pharmacovigilance has widened, with approximately 20% year-on-year growth of case volume over last decade, largely pertaining to:

- **Regulatory or safety purposes**
b. Business outcomes such as product repurposing, newer indications, new drug forms etc.
c. Broadening spread of reach of medicine in terms of number of drugs, increasing outreach of
terms of number of drugs, increasing outreach of medical services to newer geographies and increasing world population
d. Regulations becoming sensitive and stringent, leading a need to capture all-source complaints and adverse reactions data.

Revenue-wise, PV spend is expected to reach approximately 4.9Bn US$ by 2024 (Figure 10), regulatory
decisions and updates continue to be major contributors to pharmacovigilance case volume and many other business operations. Recent addition of social media, scientific communications, publications, forums, conference presentations, hospital, healthcare professional or patient’s attendant report etc. have also contributed to an increase of real-world data of adverse drug events. Hence, regulatory bodies do now allow a sample-based search and validation or quality processing from these unstructured data which majorly comprises of irrelevant information, referred to as noise. Before processing this data for PV, noise must be cleared correctly and effectively. Noise elimination is both a waste of time & effort and beyond capability of humans as well as static and rule-based algorithms due to the sheer amount of data. These are among the most important reasons to invoke AI@PV solution in PV processing.

Listening for a Shot in the Dark
AI@PV bundles the proprietary best-in-class natural language processing (NLP) and artificial intelligence (AI) technologies capable of cleaning and contextually analyzing PV data identifying trends and outliers including signals that can impact patient safety. Leveraging a combination of rule-based deterministic and probabilistic approaches, it processes structured and unstructured ICSR case documents to abstract patient, medical events, drug, and reporter information.

In AI systems, a major challenge is posed by the paradoxical nature of machine learning and regulations, learning cannot be static and regulations cannot be dynamic. In view of keeping the systems perpetually audit-ready, records of machine training and qualification (equivalent to humans) is critical. Such records
are generated from machine processing logs. Conversion of all learnings into a validated deterministic rules or conventions provide partial control of machine learning to trained and qualified PVPs.

Conclusion

Artificial Intelligence in Pharmacovigilance is amongst the most recent advancement that can disrupt life sciences and healthcare. However, unlike human PV professionals, machines lack qualification records, training and experience. Hence, there is a need to control the learning process of machine not only from compliance perspective, but also for improving accuracy, maintaining positive and correct learning. Subsequently, the role of stakeholders such as Learning Champions becomes critical. A Learning Champion is a human trained to understand machine learning, PV process and possess medical as well as technical knowledge. To interact with machines and to monitor the learning activities, humans including learning champions need to be cross-trained. The training record of techno-domain role and machine’s learning records need to be validated by learning champion. This way, AI@PV systems will be highly accurate, efficient and perpetually validated, to be ready for audits by regulatory authorities.

To know more about our first-of-its-kind cognitive automation solution for pharmacovigilance, please visit https://www.tcs.com/ai-solution-for-pharmacovigilance

---

2 Data on file with TCS
4 https://faculty.washington.edu/robinet/Learning.htm
5 https://machinelearningmastery.com/types-of-learning-in-machine-learning/