Self Generation of Policies for Training Data Curation

Contributors: **A. Abu Jabal** (Purdue), **S. Witherspoon, I. Manotas (**IBM US**),**
E. Bertino (Purdue**), S. Calo, S. Chakraborty, D. Verma (**IBM US**),**
G. Cirincione, A. Swami (ARL**), G. De Mel (**IBM UK**), G. Pearson (Dstl**).

Scope: We present an approach and architecture for generating policies required for sharing training data needed to build a machine learning model in a coalition environment. Self-generation of policies increases autonomy and reduces information management burden on humans. The demo considers a coalition of three countries: US, UK, and the fictional country Kish.

Description: We demonstrate the effectiveness of self-generated policies for data curation versus policies defined by humans in the context of three scenarios.

In the first scenario, a human administrator has manually authored/deployed policies to allow the (US hosted) data curator to only accept training data from the US and UK, thereby rejecting any training data from Kish by default (because Kish is less trusted than UK). The data curator evaluates the data sources via a registry, and then accepts or denies the data as stated in the policy. The accuracy and F1 score are then returned to the curator, quantifying the impact on the model when accepting data from only two of three countries.

The second scenario showcases the impact on the model when policies are self-generated to accept *selected* data from countries deemed "trusted" (UK) and "untrusted" (Kish); US data is accepted by default. Statistics about data distribution among different classes is used to determine the Value of Information (VoI) provided by each country, and policies are generated to selected class-specific data from different countries depending on their trust level. To determine VoI, micro services, such as data evaluation services, are used to assist in generating policies. The system can now generate policies at class granularity.

Finally, the last scenario once again shows the model's effect by the use of self-generated policies. Microservices are used to assess the Quality of Information (QoI) of data provided by each partner. This allows the system to assess the trustworthiness of partners on the fly instead of relying on pre-existing notions. Additional criteria are also evaluated prior to generating the policy, which is to evaluate the data's effect on the F1 score. If the increase is above some percentage, then a policy is generated to accept class-specific data from that source.