May 2, 2023

INL/MIS-23-72503

Zhegang Ma, Ph.D., P.E.
Lead Risk Assessment Engineer
Nuclear Safety and Regulatory Research Division
Idaho National Laboratory, USA

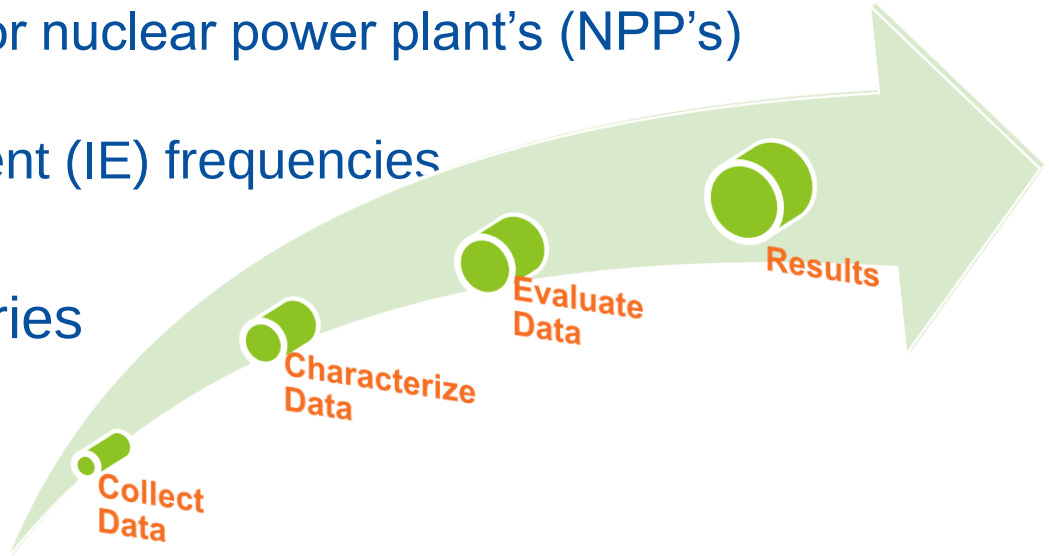
Applying AI/ML Techniques to the Nuclear Operating Experience Program and Low-Power Shutdown Initiating Event Study

Zhegang Ma[1], Fei Xu[1], Sai Zhang[1], Min Xian[2]
[1]Idaho National Laboratory, USA
[2]University of Idaho, USA
Zhegang.Ma@inl.gov, Fei.Xu@inl.gov, Sai.Zhang@inl.gov, mxian@uidaho.edu

Idaho National Laboratory

U.S. Nuclear Operating Experience Program

- Idaho National Laboratory (INL) has provided technical assistance to the Nuclear Regulatory Commission (NRC) in reliability and risk analysis including the operating experience (OpE) program since the 1980s
- The U.S. nuclear OpE program provides input parameters to the NRC's Standardized Plant Analysis Risk models
  - Also used by nuclear industry as generic data for nuclear power plant's (NPP's) probabilistic risk assessment (PRA) models
  - Equipment reliability, unavailability, initiating event (IE) frequencies
  - Common cause failure (CCF) probabilities
- The nuclear OpE program includes two categories
  - Data collection and characterization
  - Data computations and presentation



U.S. Nuclear OpE Program — Main Data Sources

- **Institute for Nuclear Power Operations (INPO) Database**
  - INPO is the U.S. industry organization that promotes safety and reliability in nuclear power plants
  - Plants report the equipment failure/unavailability data as well as demand and run time data to INPO
  - INPO provides the above proprietary data to the NRC/INL quarterly for data analysis
- **Licensee Event Reports (LERs)**
  - Plants are required to submit LERs to the NRC within 60 days of a reportable event
  - Example reportable events: plant manual or automatic reactor scram; safety barriers degraded; inadvertent actuation of emergency core cooling system
- **Event Notification (EN) Reports**
  - Immediate notification reports from plants to the NRC
- **Monthly Operating Reports (MORs)**
  - Each plant's online hours, critical hours, and shutdown information

U.S. Nuclear OpE Program — Current Practice

- **Data Characterization**
  - Analysts review OpE data source documents such as LERs and INPO reports and code the events into different studies
    - IE: initial plant fault, functional impact group
    - Shutdown IE
    - Loss of offsite power (LOOP): category, timeline, operating status, cause
    - Component failure: system, component type, failure mode, p-value, cause, detection, recovery
    - CCF: shared cause factor, time delay factor, failure mode applicability, failure cause, coupling factor, shock type, defense
- **Data Computations**
  - Use number of failures and valid demands to estimate failure probability
  - Use number of failures and run time to estimate failure rate or IE frequency
  - Use number of CCF events and total failure events to estimate CCF alpha factor

U.S. Nuclear OpE Program — Challenges

- Manually review source documents
- Established coding system and data fields in existing database and studies
  - What about new studies or expanding existing studies?
    - Digital system
    - Low-power shutdown (LPSD) IE
- Decreasing number of CCF events
- Classic statistical approaches to estimate failure probabilities/rates
  - How good is previous performance to predict current or future performance?

U.S. Nuclear OpE Program — Opportunities

- Advanced computational tools and techniques with artificial intelligence (AI) and machine learning (ML)
  - Recognize and process both structured data and unstructured data
    - Data classification for new or expanded data fields, studies, or source documents
  - Train and enhance data-driven models and reflect the relationships between model inputs and outputs
    - Causal learning, causal inference
  - Develop predictive models for parameter estimations
  - Other applications in supporting and optimizing nuclear power plant operation and maintenance including advanced diagnostic (such as detecting failure cause) and prognostic models (such as predicting the remaining useful life)
- NUREG/CR-7294, *Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants*

LPSD Initiating Event Study

- One issue in developing LPSD PRA is the reasonable estimation of LPSD IE frequencies
  - NRC has a shutdown IE database based on LER review
    - However, the shutdown IE database does not include all LPSD IEs as many LPSD IEs may not be reportable events and thus not included in LER
    - The NRC shutdown IE database is only a subset of the LPSD IEs in nuclear plants
  - Manually review OpE data is resource intensive

LPSD Initiating Event Study (cont.)

- Electrical Power Research Institute (EPRI) did conduct LPSD IE study using data from the INPO industry database
  - Keyword search plus manual review
- We are investigating a new approach to identify and characterize LPSD IEs from the INPO database using AI/ML techniques
  - Independent from the EPRI study
  - Good benchmark

NRC Shutdown IE Database – Categories and Event Counts

Shutdown IE Category	Description	Details	Event Count
ISOL	Trip or Isolation of Shutdown Cooling Loop	Primary isolation, does not include low level trip due to LOCA	22
LOAC	Loss of Safety or Vital Bus for Shutdown Cooling Equipment	Loss of vital bus due to LOOP or local fault	62
FLOW	Diversion or Loss of Cooling Water Flow	Blockage or diversion of primary coolant or service/closed cooling water flow path such that heat removal is no longer accomplished, does not include primary isolations or losses of primary coolant from the primary system	18
LOCA	Loss of Coolant Accident (LOCA)	Includes inadvertent drain-down of primary system where sufficient coolant is no longer available for the normal decay heat removal process	3
LOOP	Loss of Offsite Power		37
SPF	Spent Fuel Pool		2
Total			144

New Approach

- The main process in the AI/ML approach is to
  - Find out the relationship between keywords in the event descriptions and the LPSD IE categories
  - Apply the relationship to the INPO database records
- The new approach can be used as a supplemental method to peer review the EPRI study results if we receive their report
- If the new approach is proved to be accurate and efficient, it could be applied to other NRC OpE data characterization processes for LOOP events, IEs, CCFs, and multi-unit CCFs

Preliminary Results

- Three popular classification models
  - Support Vector Machine (SVM), with two different feature selecting strategies
  - Naïve Bayes
  - Random Forest
- Data
  - Training/Testing Dataset: 145 shutdown IEs and 542 non-shutdown IEs
  - Target Dataset: more than 12,000 records from INPO (both at-power and shutdown events)

Categories information	Category #	total # of records
'ISOL'	1	24
'LOAC'	2	64
'FLOW'	3	17
'LOCA'	4	3
'LOOP'	5	35
'SPF'	6	2
'NONSDIE'	7	542
Total of the records		687

Preliminary Results (cont.)

- Input: event description field of records
- Output: one of seven event categories
- Process:
  1) Remove stop words from the event description
  2) Create term frequency-inverse document frequency features
  3) Separate the training and testing dataset (e.g., 70% of the data is randomly selected as a training set with all remaining as testing set)
  4) Apply training and testing process to the models and obtain the overall performances
  5) Generate the accuracy of each subcategory

Preliminary Results (cont.)

- Performance comparisons of the four models

Models	Average Accuracy	Subcategories Accuracy of the Test Dataset							Subcategories Accuracy of the Training Dataset						
		1	2	3	4	5	6	7	1	2	3	4	5	6	7
		5	16	4	0	13	0	169	19	48	13	3	22	2	373
SVM	**0.8647**	0	11	0	0	2	0	167	19	45	13	3	21	2	373
Naive Bayes	0.8164	0	0	0	0	0	0	169	0	2	0	0	0	0	373
Random Forest	0.8454	2	5	0	0	0	0	168	19	45	13	3	21	2	373
SVM with Chi-Square Feature	**0.8599**	1	9	1	0	9	0	138	17	39	9	2	18	1	373

→ Category

→ "Known" Events

ML Predicted Events

Accuracy = Sum of ML Predicted Events (from Cat. 1 to Cat. 7) / Sum of Known Events (from Cat. 1 to Cat. 7)

Preliminary Results (cont.)

- Applying the AI/ML models to the Target Dataset

	Certainty Degree	ISOL	LOAC	FLOW	LOCA	LOOP	SFP	NonSDIE	Total
High Certainty	1	1	287	1	0	1	0	542	832
	0.83	6	182	8	0	5	1	0	202
Low Certainty	0.66	190	276	16	0	1	0	0	483
	0.5	645	266	39	2	3	1	0	956
	0.33	187	30	0	0	1	1	0	219
Total		1029	1041	64	2	11	3	542	2692
Certainty = 1 or 0.83		7	469	9	0	6	1	0	492
Known Events		24	64	17	3	35	2	542	687

- The above application is not ideal as it labels about 2,100 shutdown IEs, while there are a total of about 2,700 shutdown events (IE or not) from the 12,000 IRIS records

Summary of LPSD IE Study

- One issue in developing LPSD PRA is the reasonable estimation of LPSD IE frequencies
  - Manually review OpE is resource intensive
- We are investigating a new approach to identify and characterize LPSD IEs using AI/ML techniques
- Preliminary models were developed and applied
  - Preliminary results show that the models need to be improved
- Issues in the study
  - Training/testing datasets are too small
  - Six shutdown IE categories are not distributed in the training/testing dataset uniformly
  - Event descriptions and terms in the records are not standardized

Idaho National Laboratory

Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy. INL is the nation's center for nuclear energy research and development, and also performs research in each of DOE's strategic goal areas: energy, national security, science and the environment.