September 8, 2022

Zhegang Ma, Ph.D., P.E.
Zhegang.Ma@inl.gov

Fei Xu, Ph.D.
Fei.Xu@inl.gov

Sai Zhang, Ph.D.
Sai.Zhang@inl.gov

INL/MIS-22-68695

AI/ML Support for LPSD Project

A New Approach to Identify Nuclear Plant Shutdown Initiating Events Using Machine Learning Techniques

Introduction

- Probabilistic risk assessment (PRA) is used by the Nuclear Regulatory Commission (NRC) and nuclear industry to estimate risk and provide insights into the strengths and weaknesses of the design and operation of a nuclear plant
  - What can go wrong?
  - How likely is it?
  - What are consequences?
- Earlier focuses have been on at-power, internal event PRA
- Risks from external hazards (such as fire, seismic, flood, high wind) as well as during low-power shutdown (LPSD) operation could be significant
- The needs to develop LPSD PRA and external hazards PRA are on the rise

Introduction (cont.)

- One issue in developing LPSD PRA is the reasonable estimation of shutdown initiating event (IE) frequencies
  - NRC has an at-power IE database based on a licensee event report (LER) review that can be used to estimate at-power IE frequencies
  - NRC has also a shutdown IE database based on LER review
    - However, the shutdown IE database does not include all shutdown IEs as some shutdown IEs may not be reportable for LER
    - The NRC shutdown IE database is only a subset of the shutdown IEs in nuclear plants
  - Manual review operating experience is resource intensive and applying only a small set of source records such as LERs

Introduction (cont.)

- Electrical Power Research Institute (EPRI) has estimated shutdown IE frequencies by using data from the Institute of Nuclear Power Operations (INPO) industry database
  - Keyword search plus manual review
- Manually reviewing thousands of record from the INPO Industry Reporting and Information System (IRIS) database is too resource intensive
- Idaho National Laboratory is investigating a new approach to identify and characterize shutdown IEs from the IRIS database using artificial intelligence/machine learning (AI/ML) techniques

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 SDC 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	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 Shutdown IE categories
  - Apply the relationship to the INPO IRIS database records
- The new approach can be used as a supplemental method to peer review the EPRI updated 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 operating experience data characterization processes for loss-of-offsite-power events, initiating events, common-cause failures (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: about 140 shutdown IEs and 540 non-shutdown IEs
  - Target Dataset: more than 12,000 records from INPO IRIS (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
'SpentFuelPool'	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 (TFIDF) 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

- One issue in developing LPSD PRA is the reasonable estimation of shutdown IE frequencies
  - Manual review operating experience is resource intensive and applying only a small set of source records such as LERs
- We are investigating a new approach to identify and characterize shutdown 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
  - The 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.