Interpretation of PSA results using semantic analysis of minimal cutsets

Gennadi Loskoutov, Per Hellström
Swedish Radiation Safety Authority

Importance

- What makes an event to be important for understanding the results?
  - One of possible definitions: ”*Frequent event combinations show the important events*”
  - Another definition: “*Important event combinations influence the overall result*”

- What is a frequent event?
  - Frequent = occurs often = occurs many times
  - Current identification of *frequent event combination* using importance measures in PSA includes event probability information **on top** of the fact that the combination occurs many times

Importance measures

- Fussell-Vesely (FV)
- Risk Reduction or Risk Decrease Factor (RDF)
- Risk Increase Factor or Risk Achievement Worth (RIF or RAW)

- All measures are defined using probability information for events

Corner cases

- Consider an analysis, where very few event combinations are much more probable than the others
  - Only these combinations are identified as important using importance measures
  - This conclusion is true only if the probability information is correct. In PSA, the probability information is associated with uncertainties.
- Consider an analysis, where risk profile is smooth, and the values of importance measures are almost the same for all event combinations
  - Modelling and results are too fragmented to provide a useful answer for overall picture

Suggested approach

- Is it possible to identify important event combinations without considering their actual probabilities?
  - When reading a book or a text, you can identify its subject (and therefore its major features) by considering frequent words and phrases
- If an MCS list were a text, it would be possible to identify its major features by considering frequent words and phrases
  - MCS list = text
  - MCS = sentence, phrase (event combination = phrase)
  - Event ID = word

MCS list as a word frequency matrix

Frequency = n/N
n is number of MCS where basic event participates,
N is a total number of MCS

	BE1	**BE2**	**BE3**	**…**	**BE-n**
IE1	1	0,92	0,2	…	0
IE2	0	0,3	1	…	0
IE3	0	0,4	0,001	…	0,05
…	…	…	…	…	…
IE-n	0	0	0,7	…	0,46

Semantic analysis of an MCS list

- Count occurrences of events per MCS list and use these counts to identify frequent events and event combinations
- Remove very frequent events as noise (e.g. the initiating event present in each cutset)
- Use number of event occurrences per MCS list as basis for identification of major features of the MCS list (topics)

Available techniques for topic identification in text documents

- **Non-negative matrix factorization (NMF)**

  Matrix V is factorized into two matrices W and H, with the property that all three matrices have no negative elements

  Distribution of analyses over events (matrix V) is factorized into two – distribution of analyses over topics (matrix W) and distribution of topics over events (matrix H)

- **Latent Dirichlet Allocation (LDA)**

  Distributions of documents over topics and topics over words are generated using Bayesian inference, where a sparse Dirichlet prior is used for the distribution of topics over words.

Semantic analysis of an MCS list by topic identification

- Use number of event occurrences per MCS list as basis for identification of major features of the MCS list (topics)
  - Each text can be considered as a mixture of topics it describes (topic distribution)
  - Each topic can be considered as a frequency distribution over words
  - *Examples*:
    - (reactor, fission, neutron flux, moderator) ≈ “nuclear reactor”
    - (scram, boron, control rod, reactivity) ≈ “reactor control and protection system”
    - A text, where the words above are repeated frequently, probably describes something related to reactor control and protection system in a nuclear reactor.

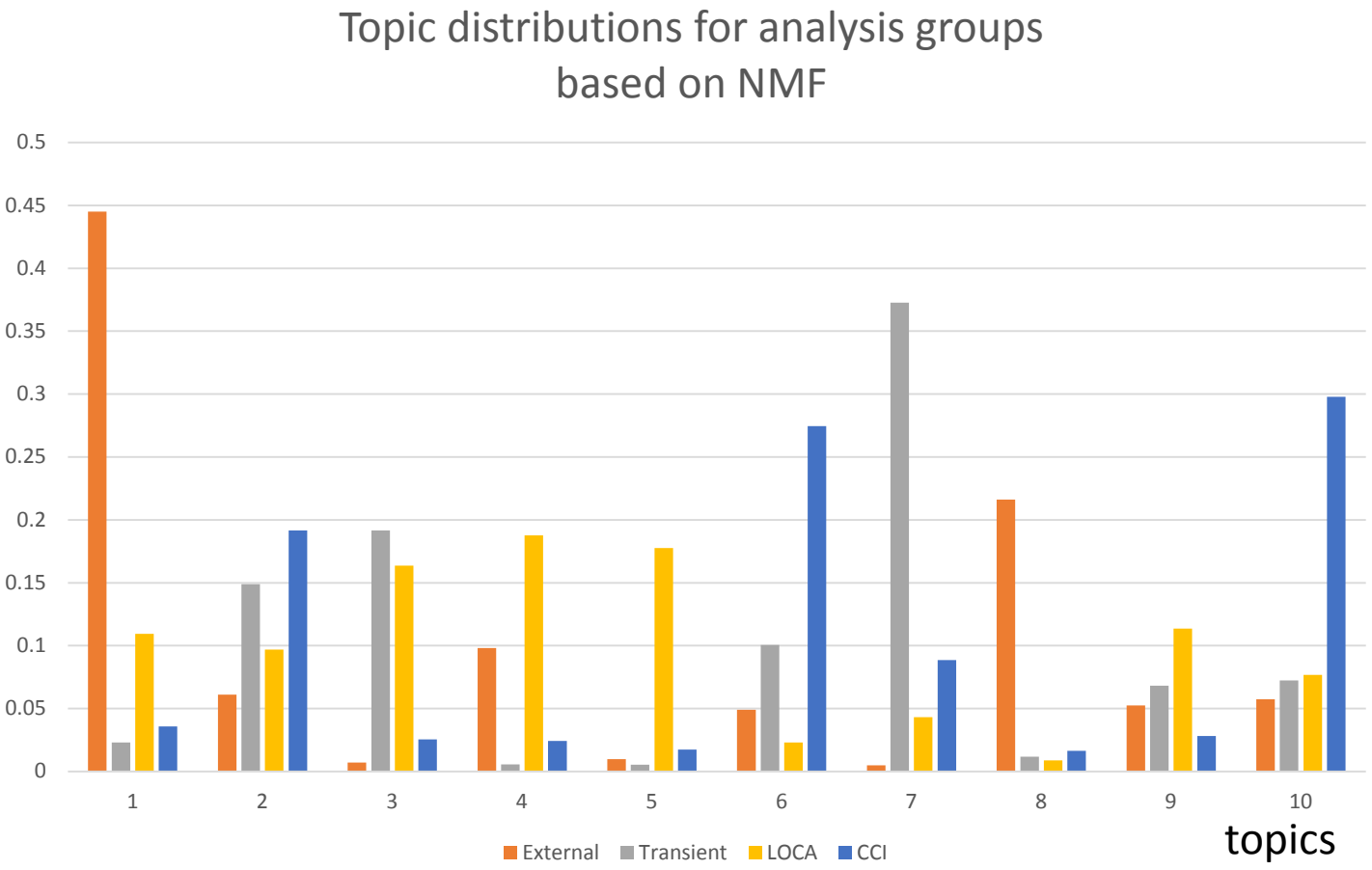
Semantic analysis by topic identification in MCS lists

- Problem of finding important event combinations and understanding their meaning becomes explicitly two-fold:
  - First, identify topics that appear in the MCS lists
    - It roughly corresponds to finding safety functions that are important for the given plant design (e.g. heat removal from reactor core, containment integrity, power supply and so on)
  - Second, identify events that are most frequent for specific topic
    - It roughly corresponds to finding events and component failures that are important for specific safety functions (e.g. diesel generators for power supply)

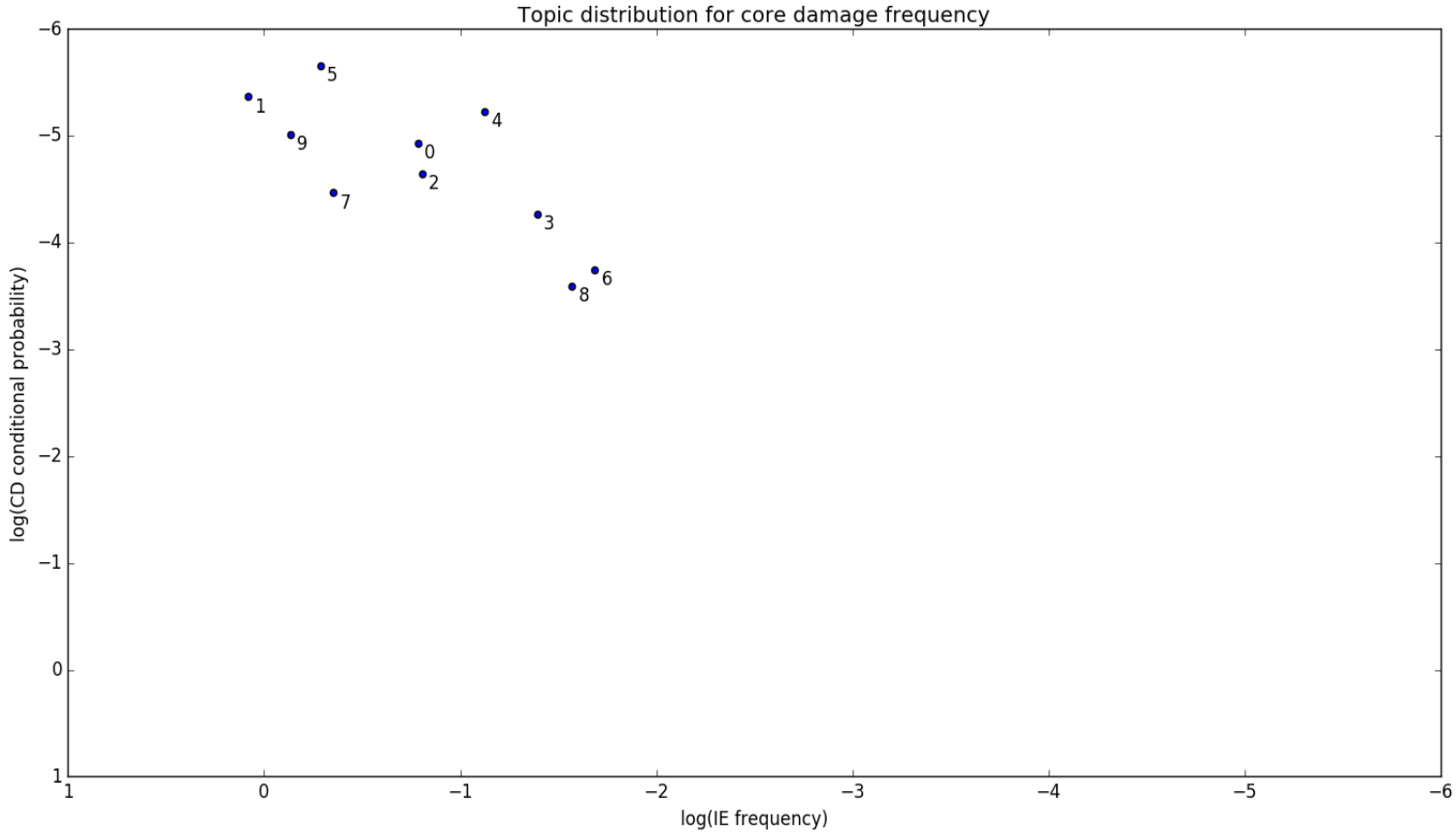
Results of semantic analysis by topic identification in MCS lists

- Each MCS list is considered to be a mixture of topics (topic distribution)
  - ≈ *distribution of analyses over safety functions*
  - Ability to compare MCS results to each other
    - If topic distributions in MCS lists for two different analyses are similar, the plant design is vulnerable to the same function failures for both cases
  - Ability to calculate potential CDF contributions from identified topics
- Each topic can be considered as a frequency distribution over words
  - ≈ *distribution of safety functions over events*

Results of semantic analysis by topic identification in MCS lists (examples)



Results of semantic analysis by topic identification in MCS lists (examples, cont)



Conclusions

- The suggested approach can be used as a complement to existing techniques for importance and sensitivity analysis for identification of important aspects of plant design and operation
- It allows comparison of analyses using event occurrences in MCS lists instead of probabilistic measures
- It allows to reveal important events even in cases, where MCS list is too short because of analysis technicalities such as used cut-off
  - If topic X is important for analysis Y, and event Z is an important word describing topic X, then event Z might be important for analysis Y even if the event is not present in MCS list