University of Southampton

Msc Cyber Security

Project Preparation

Risk profiling using NLP

Author:
Gerard Tio Nogueras

Supervisor:
Dr. Sylvain Frey

1 What is the reader supposed to learn?

- main people/technologies in the field (check for a bible and standard technologies/libraries)
- recent major advances and discoveries (check for latest updated technology)
- gaps in the research (nlp sentiment alaysis doesn't have a risk based decision section or a risk assessment modelling)
- current debates (might not come into our review)
- where the subject might go next (explain the different obj of the project: present the paper with new evidences, expand the game, use the game in companies as training but also as an evaluation of the risk based decisions taken by the employees)

2 Our job

Provide an insight by taking the knowledge gained and explain the subject of the review to the reader.

3 Our contribution

Cover the major players in the field and give your critical opinion about these players regarding the context of the subject.(matrix)

4 Importance of Critical thinking

Be very critical of the references used in the review. Be critical when reading the and their results. Not only their results but their methods are correct.

This is important when finding disagreements in different references and then talk about debate in the field.

5 The matrix

To present the references used and their content; as well as the different possibilities that exist for a specific technology/field.
To do so we create a matrix:

- FEATURES(x) and REFERENCES(y)
- POSSIBILITIES(x) and FEATURES(y)

Then you use these matrix to present your review. You can do so following to strategies:

- chronological: present the technologies or references by chronological order.
- Feature: present the different features and for each feature present which references talk about it and the differences found there.

IMPORTANT: Technology reviews are best done by taking each reference and exploring the key features. You can use a scoring for each feature to provide a ranking.
Surveys of a discipline with a strong narrative of the discipline itself are best done by talking about each feature in turn with all relevant references.

ADD THIS TABLE AT THE END TO SUMMARIZE THE REVIEW.

6 List of features with concerned papers

7 Idea

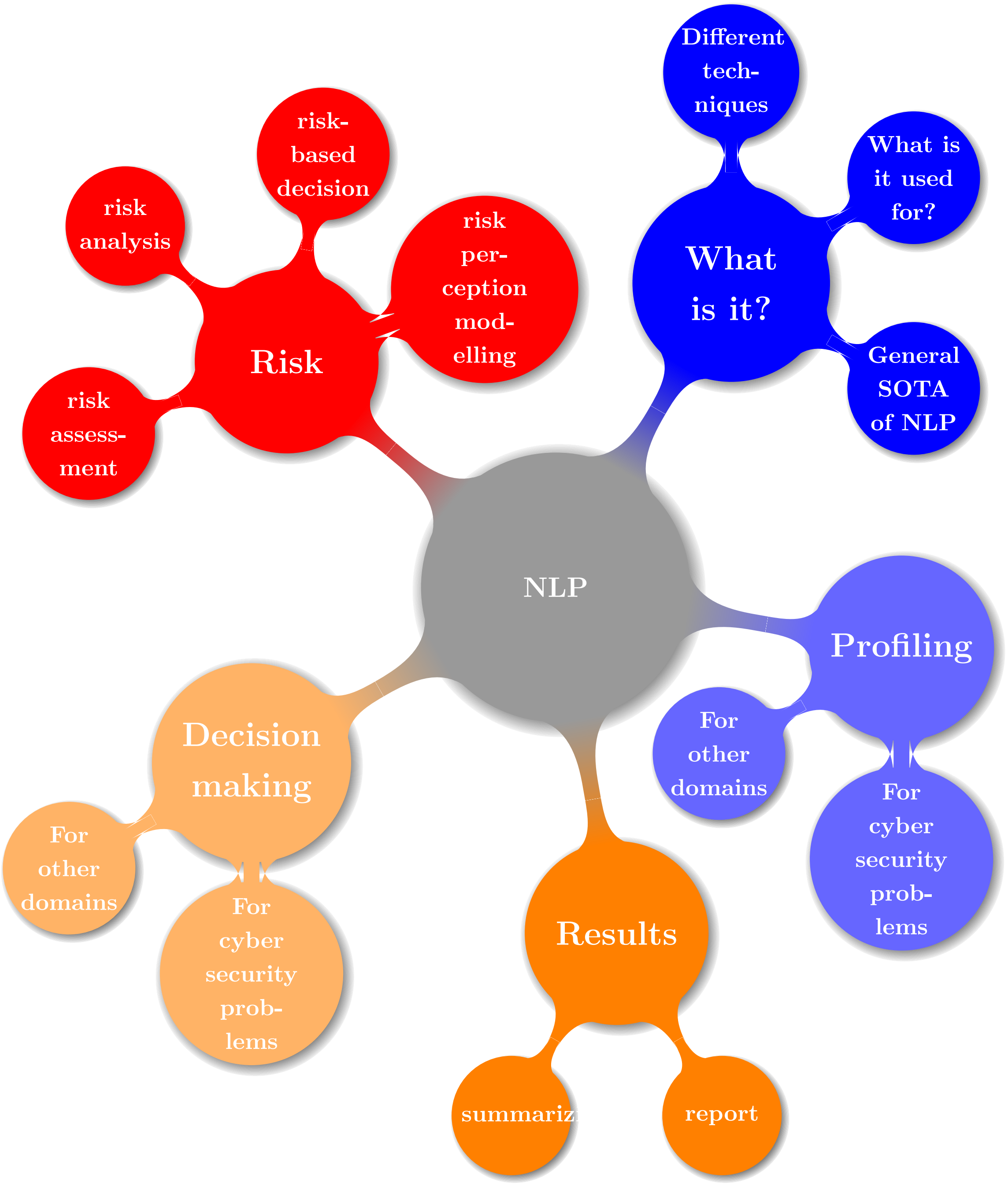
List all the interesting references, highlight the top 10 to create a state of the art for my project and give access to abstracts of the entire list of references used.

key search words: "profiling" "risk assessment" "cyber security" "summarizing" "report" "risk-based decision" "decision making" "risk analysis" "perception modelling" "risk perception modelling" "opinion" "emotion" "sentiment"

NEXT STEP: Commencer crire qu'est-ce que NLP, quoi a sert, et pourquoi on veut l'utiliser dans ce cas prcis a aidera focaliser la recherche / lecture de nouveau papiers et identifier des angles morts dans l'tat de l'art, tel que NLP for risk perception.

Possible final subject: Cyber security Risk perception using NLP

8 Mindmap


NLP
What is it?
Different techniques
What is it used for?
General SOTA of NLP
Risk
risk-based decision
risk analysis
risk perception modelling
risk assessment
Profiling
For other domains
For cyber security problems
Decision making
For other domains
For cyber security problems
Results
summariz
report

9 Intro

9.1 Where does this research come from

"Stakeholders in security decisions play a fundamental role in determining security requirements, yet, little is currently understood about how different stakeholder groups within an organisation approach security and the drivers and tacit biases underpinning their decisions"[1]. To have a better understanding about this subject my supervisor Dr. Sylvain FREY created a board game that simulates these interactions. The problem he/his team encountered when proposing their results to the main cyber security conferences (not sure of the term to use here) is that their results can not be accepted because of the low amount of games played. Therefore we are trying to find other ways to support their results.

9.2 What is NLP?

Natural processing language,summarize wiki and what is interesting to us

9.3 Why are we interested in NLP?

We possess scripted recordings of 12 game-plays. Using NLP we hope to achieve a report of the interactions by profiling the players and their risk-based decisions. We wish to visualize the starting state of the players, their adaptations, their knowledge, the interactions with a "champion", how risk is taken into account, how different backgrounds affect the decision-making process and ADD HERE MORE RELEVANT GOALS + present some papers that will be the base ground of our research

9.4 What do we hope to achieve?

The final goal of this project/research is to support the results found previously and additionally to be able to create reports with constructive criticism of the flaws of the players methodology and the approaches taken during the game to enhance the risk based decisions of the different entities of the company.

Another aspect that we hope to achieve is to create a starting ground for future research on risk-based profiling for cyber security decisions. Because right now there is a lot of research done on medical risk but not on cyber security risk decisions.

10 TOP 10 research papers

- The Good, the Bad and the Ugly: A Study of Security Decisions in a Cyber-Physical Systems Game [1]
- A COMPARATIVE STUDY OF SENTIMENT ANALYSIS TECHNIQUES [15]

11 TOP 50 research papers

- The Good, the Bad and the Ugly: A Study of Security Decisions in a Cyber-Physical Systems Game [1]
- Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research[?]
- The role of behavioral research and profiling in malicious cyber insider investigations[?]
- Cyber security risk assessment for SCADA and DCS networks [?]
- Profiling Enterprise Risks in Large Computer Companies Using the Leximancer Software Tool[3]
- The challenges of automatic summarization [4]
- Mining and summarizing customer reviews [?]
- Natural language processing for information retrieval [?]
- Sentiment analysis: capturing favorability using natural language processing [?]
- Ontology-based parser for natural language processing [?]
- Systems for natural language processing of sentence based queries [?]
- Automated Product Profiling through NLP [2]
- Natural Language Processing [?]
- Profiling Academic Research on Digital Games Using Text Mining Tools [?]
- Semantic Web Mining: State of the art and future directions [?]
- Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions [?]
- Deep Learning for Natural Language Processing (without Magic) [5]
- Stanford CoreNLP [6]
- Natural Language Processing: State of the Art and Prospects for Significant Progress, A workshop sponsored by the National Library of Medicine [7]
- The linguistic approach to the natural language requirements quality: benefit of the use of an automatic tool [9]
- Natural language processing: an introduction [10]
- Pattern Recognition and Natural Language Processing: State of the Art [11]

- Semantic Analysis for Monitoring Insider Threats [12]
- Learning to Connect Language and Perception [13]
- Precisiated Natural Language [14]
- A COMPARATIVE STUDY OF SENTIMENT ANALYSIS TECHNIQUES [15]
- Jumping NLP Curves: A Review of Natural Language Processing Research [16]
- NLTK: the Natural Language Toolkit [17]
- Identifying Expressions of Opinion in Context [18]
- Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques [19]
- OpinionFinder: a system for subjectivity analysis [20]
- Annotating Expressions of Opinions and Emotions in Language [21]
- Thumbs up? sentiment classification using machine learning techniques [22]
- Direction-based text interpretation as an information access refinement [23]
- On the computation of point of view [24]

Possible interesting papers:
Sentiment classifiers:

S. Das and M. Chen. Yahoo! for amazon: Extracting market sentiment from stock message boards. In Proc. of the 8th APFA, 2001. `http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.202.6418`
R. M. Tong. An operational system for detecting and track- ing opinions in on-line discussion. In SIGIR Workshop on Operational Text Classification, 2001.
Affect analysis:
P. Subasic and A. Huettner. Affect analysis of text using fuzzy semantic typing. IEEE Trans. on Fuzzy Systems, Special Issue, Aug., 2001.
C. Whissell. The dictionary of affect in language. Emotion: Theory, Research, and Experience, pages 113131.
automatic survey analysis:
H. Li and K. Yamanishi. Mining from open answers in questionnaire data. In Proc. of the 7th ACM SIGKDD Conf. , 2001.
opinion extraction:
S. Morinaga, K. Yamanishi, K. Teteishi, and T. Fukushima. Mining product reputations on the web. In Proc. of the 8th ACM SIGKDD Conf. , 2002.
Recommender systems:

L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter. PHOAKS: A system for sharing recommendations. CACM, 40(3):5962, 1997.

12 Non-research information

- OpenDNS Uses Natural Language Processing to Detect APTs [?]
- Using Natural Language Processing to Identify Malicious Domains
- BASICS OF NLP [?]
- NLP Ideas [?]
- NLP wiki [?]
- SOTA nlp 2015 [8]
- NLP english [?]

References

[1] Sylvain Frey, Awais Rashid, Pauline Anthonysamy, Maria Pinto-Albuquerque, and Syed Asad Naqvi. 2016. The Good, the Bad and the Ugly: A Study of Security Decisions in a Cyber-Physical Systems Game.

[2] https://nlp.stanford.edu/courses/cs224n/2011/reports/jihohan-gogo9th.pdf

[3] http://link.springer.com/article/10.1057/palgrave.rm.8250030

[4] http://ieeexplore.ieee.org/abstract/document/881692/

[5] https://nlp.stanford.edu/courses/NAACL2013/ + http://nlp.stanford.edu/~socherr/DeepLearning-ACL2012-tutorial.pdf

[6] http://stanfordnlp.github.io/CoreNLP/ + http://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf

[7] https://www.researchgate.net/publication/243966799_Natural_Language_Processing_State_of_the_Art_and_Prospects_for_Significant_Progress_A_workshop_sponsored_by_the_National_Library_of_Medicine

[8] https://www.reddit.com/r/MachineLearning/comments/4020ek/state_of_the_art_dec2015_natural_language/

[9] http://ieeexplore.ieee.org/abstract/document/992662/

[10] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168328/

[11] http://www.temjournal.com/content/52/TemJournalMay2016_236_240.pdf

[12] http://link.springer.com/chapter/10.1007/978-3-540-25952-7_40

[13] http://www.aaai.org/Papers/AAAI/2008/AAAI08-271.pdf

[14] http://www.aaai.org/ojs/index.php/aimagazine/article/view/1778

[15] https://pdfs.semanticscholar.org/3f10/b006bab60c7f363bc03e72ad405d264b8d42.pdf

[16] http://ieeexplore.ieee.org/abstract/document/6786458/?part=1

[17] http://dl.acm.org/citation.cfm?id=1118117

[18] http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-431.pdf

[19] http://ieeexplore.ieee.org/abstract/document/1250949/ http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1250949

[20] http://dl.acm.org/citation.cfm?id=1225751

[21] http://link.springer.com/article/10.1007%2Fs10579-005-7880-9?LI=true

[22] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In Proc. of the 2002 ACL EMNLP Conf. , pages 7986, 2002. https://www.cs.cornell.edu/home/llee/papers/sentiment.pdf

[23] M. Hearst. Direction-based text interpretation as an informa-tion access refinement. Text-Based Intelligent Systems, 1992. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2037E600F4275AC38D0DCDE1010C71B3?doi=10.1.1.40.9124&rep=rep1&type=pdf or http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.9124

[24] W. Sack. On the computation of point of view. In Proc. of the 12th AAAI Conf., 1994. http://www.aaai.org/Papers/AAAI/1994/AAAI94-282.pdf