SOME RESEARCH CHALLENGES FOR BIG DATA ANALYTICS OF INTELLIGENT SECURITY

Yuh-Jong Hu
hu at cs.nccu.edu.tw

Emerging Network Technology (ENT) Lab.
Department of Computer Science
National Chengchi University, Taipei, Taiwan

April-10-2015

Seminar at CS&E, Yuan Ze Univ.

Motivations

1. My original expertise is Semantic Web for privacy protection in the Cloud.
2. Here, we are exploiting the *structured machine learning* for big data analytics.
3. We hope this might be helpful for the security intrusion detection problem.
4. So we intend to apply the structured machine learning for intelligent security and other domains.

Motivations

1. My original expertise is Semantic Web for privacy protection in the Cloud.
2. Here, we are exploiting the *structured machine learning* for big data analytics.
3. We hope this might be helpful for the security intrusion detection problem.
4. So we intend to apply the structured machine learning for intelligent security and other domains.

Motivations

1. My original expertise is Semantic Web for privacy protection in the Cloud.
2. Here, we are exploiting the *structured machine learning* for big data analytics.
3. We hope this might be helpful for the security intrusion detection problem.
4. So we intend to apply the structured machine learning for intelligent security and other domains.

Motivations

1. My original expertise is Semantic Web for privacy protection in the Cloud.
2. Here, we are exploiting the *structured machine learning* for big data analytics.
3. We hope this might be helpful for the security intrusion detection problem.
4. So we intend to apply the structured machine learning for intelligent security and other domains.

Research Challenges

1. How to apply the structured machine learning for big data?
2. Why collecting the security big datasets is hard?
3. What is the big data analytics lifecycle for intelligent security?
4. How do we extract the security features to model, classify, and detect the malicious (or outlier) behaviors?
5. Which modelling and analytics paradigms for recognizing intrusion patterns?
6. What are the possible core technologies?
    - Knowledge representation and discovery (or query)?
    - Machine learning algorithms?
    - How to combine structured knowledge with machine learning?
    - Which big data analytics platform? Spark vs. Hadoop

Research Challenges

1. How to apply the structured machine learning for big data?
2. Why collecting the security big datasets is hard?
3. What is the big data analytics lifecycle for intelligent security?
4. How do we extract the security features to model, classify, and detect the malicious (or outlier) behaviors?
5. Which modelling and analytics paradigms for recognizing intrusion patterns?
6. What are the possible core technologies?
   - Knowledge representation and discovery (or query)?
   - Machine learning algorithms?
   - How to combine structured knowledge with machine learning?
   - Which big data analytics platform? Spark vs. Hadoop

Research Challenges

1. How to apply the structured machine learning for big data?
2. Why collecting the security big datasets is hard?
3. What is the big data analytics lifecycle for intelligent security?
4. How do we extract the security features to model, classify, and detect the malicious (or outlier) behaviors?
5. Which modelling and analytics paradigms for recognizing intrusion patterns?
6. What are the possible core technologies?
   - Knowledge representation and discovery (or query)?
   - Machine learning algorithms?
   - How to combine structured knowledge with machine learning?
   - Which big data analytics platform? Spark vs. Hadoop

Research Challenges

1. How to apply the structured machine learning for big data?
2. Why collecting the security big datasets is hard?
3. What is the big data analytics lifecycle for intelligent security?
4. How do we extract the security features to model, classify, and detect the malicious (or outlier) behaviors?
5. Which modelling and analytics paradigms for recognizing intrusion patterns?
6. What are the possible core technologies?
   - Knowledge representation and discovery (or query)?
   - Machine learning algorithms?
   - How to combine structured knowledge with machine learning?
   - Which big data analytics platform? Spark vs. Hadoop

Research Challenges

1. How to apply the structured machine learning for big data?
2. Why collecting the security big datasets is hard?
3. What is the big data analytics lifecycle for intelligent security?
4. How do we extract the security features to model, classify, and detect the malicious (or outlier) behaviors?
5. Which modelling and analytics paradigms for recognizing intrusion patterns?
6. What are the possible core technologies?
   - Knowledge representation and discovery (or query)?
   - Machine learning algorithms?
   - How to combine structured knowledge with machine learning?
   - Which big data analytics platform? Spark vs. Hadoop

Research Challenges

1. How to apply the structured machine learning for big data?
2. Why collecting the security big datasets is hard?
3. What is the big data analytics lifecycle for intelligent security?
4. How do we extract the security features to model, classify, and detect the malicious (or outlier) behaviors?
5. Which modelling and analytics paradigms for recognizing intrusion patterns?
6. What are the possible core technologies?
    - Knowledge representation and discovery (or query)?
    - Machine learning algorithms?
    - How to combine structured knowledge with machine learning?
    - Which big data analytics platform? Spark vs. Hadoop

Possible Approaches

1. Machine learning algorithms in inductive reasoning
2. Logic programming in deductive reasoning
3. Inductive with deductive reasoning? e.g. (probabilistic) inductive logic programming (ILP), statistical relational learning (SRL), structured machine learning.
4. **C**omposite **B**ig **D**ata **A**nalytics and **M**odelling (**CBDAM**)
5. Establishing CBDAM framework on Spark.
6. Verifying CBDAM for intelligent security and other domains.

Possible Approaches

1. Machine learning algorithms in inductive reasoning
2. Logic programming in deductive reasoning
3. Inductive with deductive reasoning? e.g. (probabilistic) inductive logic programming (ILP), statistical relational learning (SRL), structured machine learning.
4. **C**omposite **B**ig **D**ata **A**nalytics and **M**odelling (**CBDAM**)
5. Establishing CBDAM framework on Spark.
6. Verifying CBDAM for intelligent security and other domains.

Possible Approaches

1. Machine learning algorithms in inductive reasoning
2. Logic programming in deductive reasoning
3. Inductive with deductive reasoning? e.g. (probabilistic) inductive logic programming (ILP), statistical relational learning (SRL), structured machine learning.
4. **C**omposite **B**ig **D**ata **A**nalytics and **M**odelling (**CBDAM**)
5. Establishing CBDAM framework on Spark.
6. Verifying CBDAM for intelligent security and other domains.

Possible Approaches

1. Machine learning algorithms in inductive reasoning
2. Logic programming in deductive reasoning
3. Inductive with deductive reasoning? e.g. (probabilistic) inductive logic programming (ILP), statistical relational learning (SRL), structured machine learning.
4. **C**omposite **B**ig **D**ata **A**nalytics and **M**odelling (**CBDAM**)
5. Establishing CBDAM framework on Spark.
6. Verifying CBDAM for intelligent security and other domains.

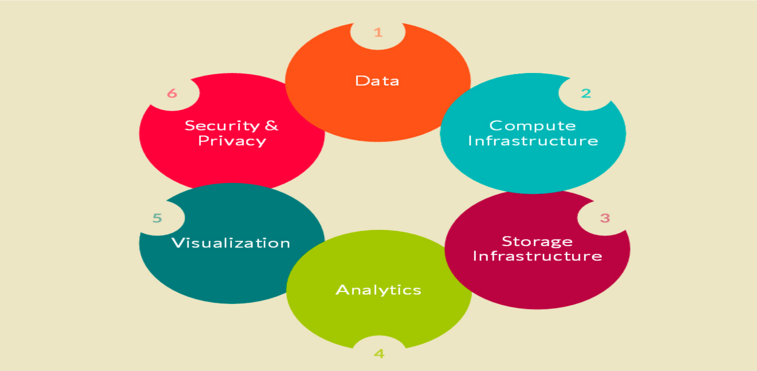
Possible Approaches

1. Machine learning algorithms in inductive reasoning
2. Logic programming in deductive reasoning
3. Inductive with deductive reasoning? e.g. (probabilistic) inductive logic programming (ILP), statistical relational learning (SRL), structured machine learning.
4. **C**omposite **B**ig **D**ata **A**nalytics and **M**odelling (**CBDAM**)
5. Establishing CBDAM framework on Spark.
6. Verifying CBDAM for intelligent security and other domains.

Possible Approaches

1. Machine learning algorithms in inductive reasoning
2. Logic programming in deductive reasoning
3. Inductive with deductive reasoning? e.g. (probabilistic) inductive logic programming (ILP), statistical relational learning (SRL), structured machine learning.
4. **C**omposite **B**ig **D**ata **A**nalytics and **M**odelling (**CBDAM**)
5. Establishing CBDAM framework on Spark.
6. Verifying CBDAM for intelligent security and other domains.

Big Data 6-Dimension Taxonomy

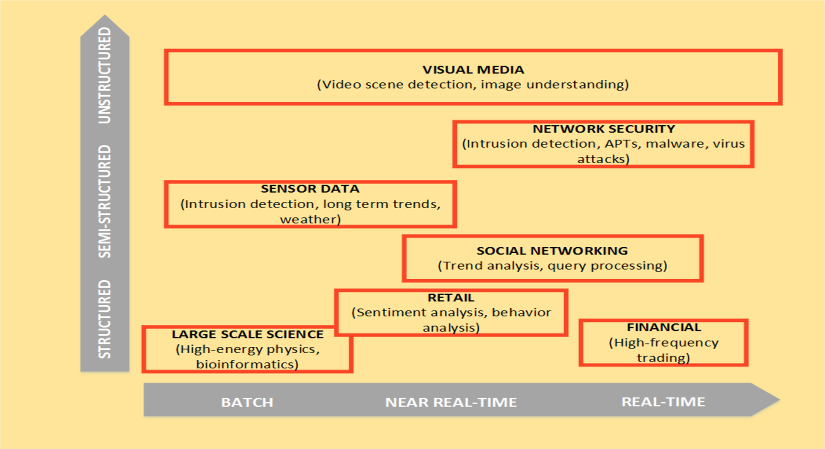


–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Possible Data Domains for Analytics

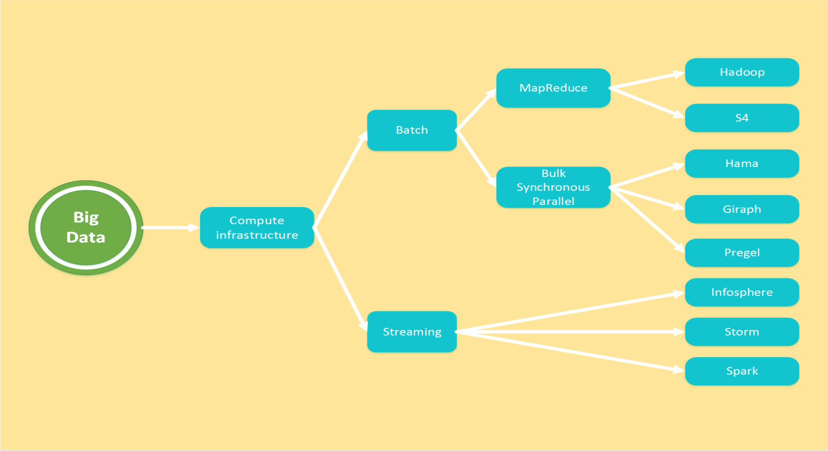
–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Mapping the Big Data Verticals



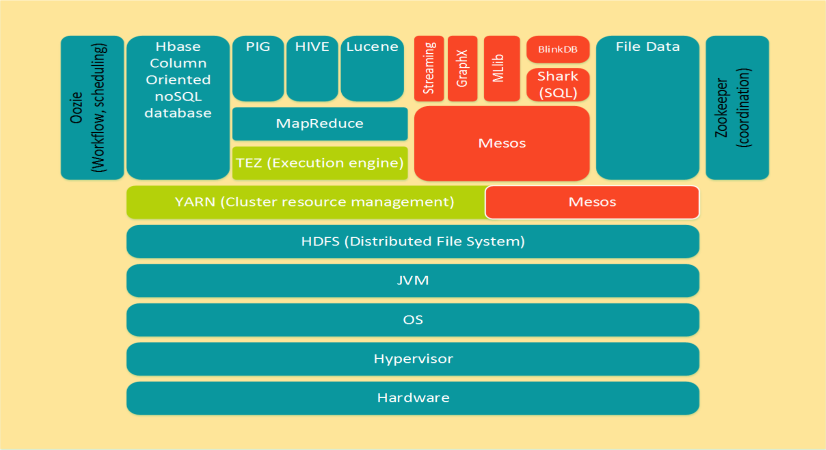
–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Big Data Computing Infrastructure



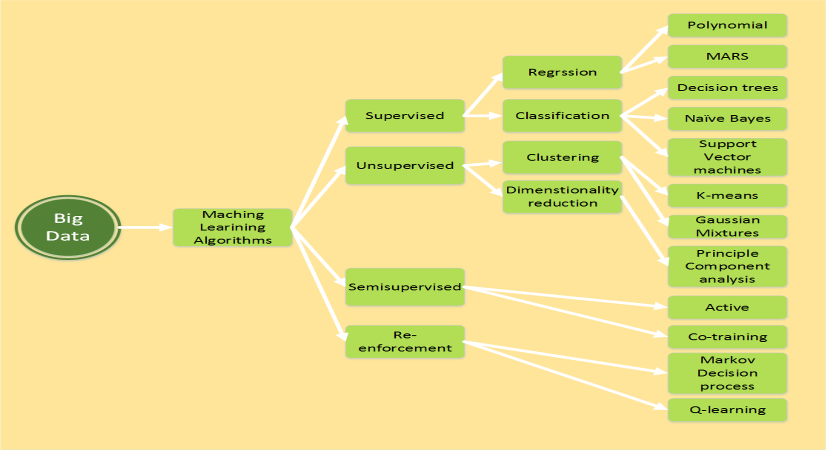
–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Spark in the Hadoop 2.0



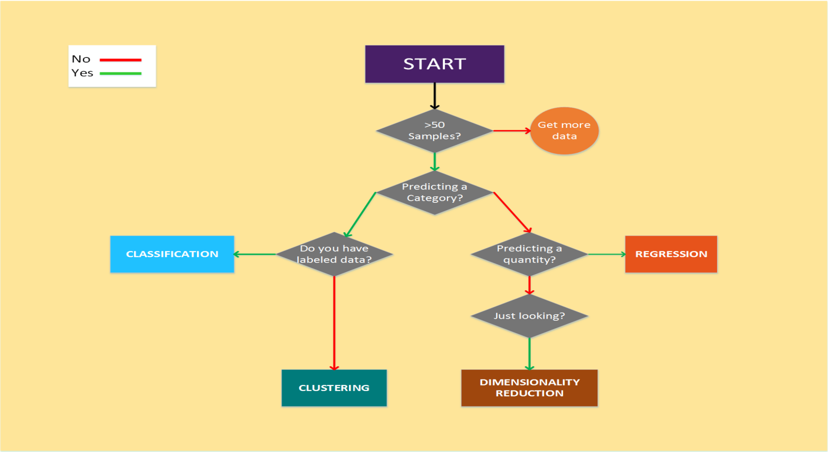
–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Possible Machine Learning Algorithms



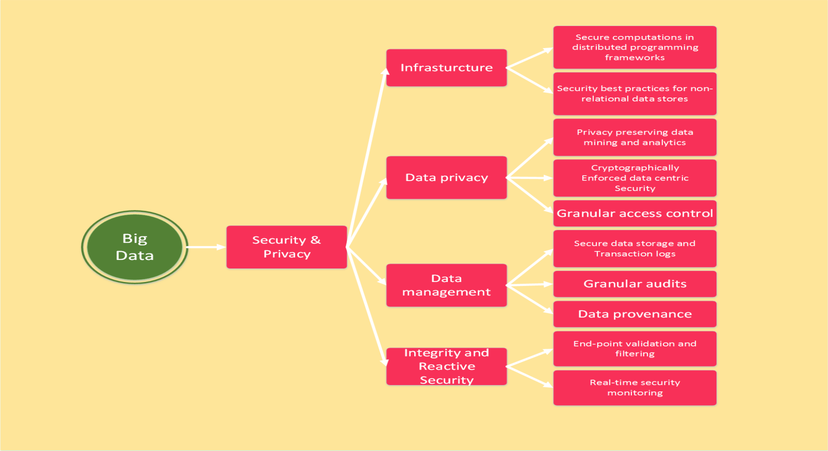
–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Machine Learning Flow Chart



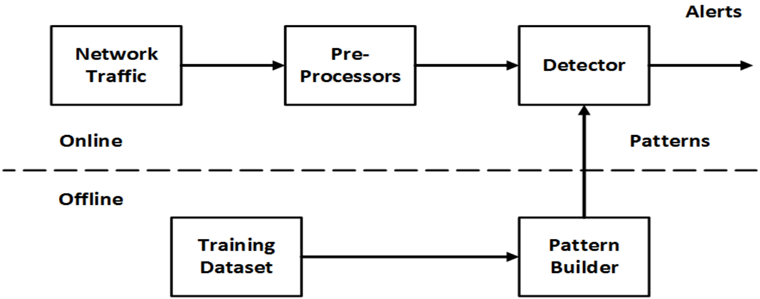
–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Security and Privacy Research Problems



–Big Data Taxonomy, Big Data Working Group, CSA, Sep. 2014

Misuse Intrusion Detection



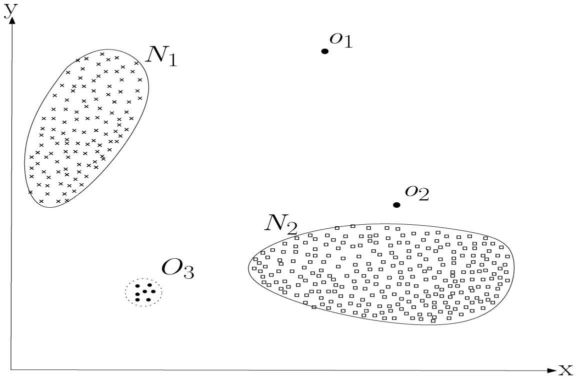
–J. Zhang, et al., Random-Forests-Based Network Intrusion Detection Systems, Sep. 2008

Misuse Intrusion Detection (Conti.)

- Signature rule-based SIEM systems on misuse detection
- Discovering attacks from intrusions *known features*
- Low false positive rate but cannot detect new attacks
- Zero-day and APTs are novel new attacks.

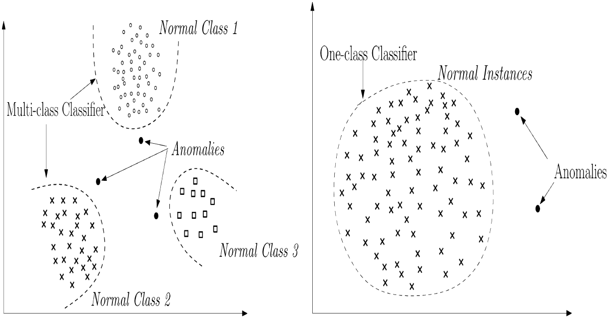
–J. Zhang, et al., Random-Forests-Based Network Intrusion Detection Systems, Sep. 2008

Anomaly Intrusion Detection



–V. Chandola, et al., Anomaly Detection: A Survey, ACM Computing Surveys, July 2009

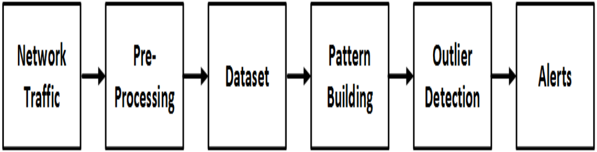
Anomaly Intrusion Detection (Conti.)



(a) Multi-class Anomaly Detection (b) One-class Anomaly Detection

–V. Chandola, et al., Anomaly Detection: A Survey, ACM Computing Surveys, July 2009

Anomaly Intrusion Detection (Conti.)



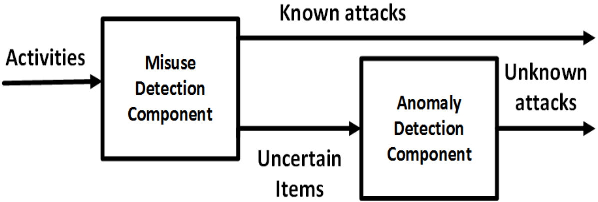
–J. Zhang, et al., Random-Forests-Based Network Intrusion Detection Systems, Sep. 2008

Anomaly Intrusion Detection (Conti.)

- Identifying attacks with significant deviations from normal.
- Extracting features to represent normal activity is hard.
- Can detect new attacks, but a high false positive rate.
- Use attack free training datasets to learn.
- How about hybrid intrusion detection?

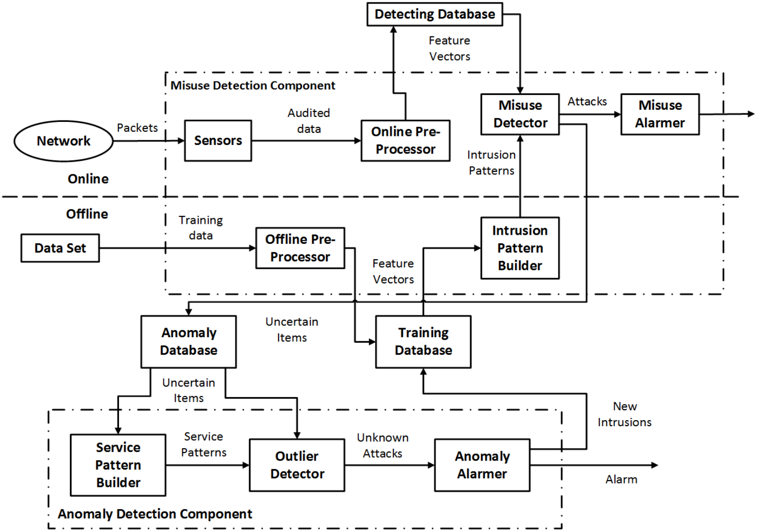
–J. Zhang, et al., Random-Forests-Based Network Intrusion Detection Systems, Sep. 2008

Hybrid Intrusion Detection



–J. Zhang, et al., Random-Forests-Based Network Intrusion Detection Systems, Sep. 2008

Hybrid Intrusion Detection (Conti.)

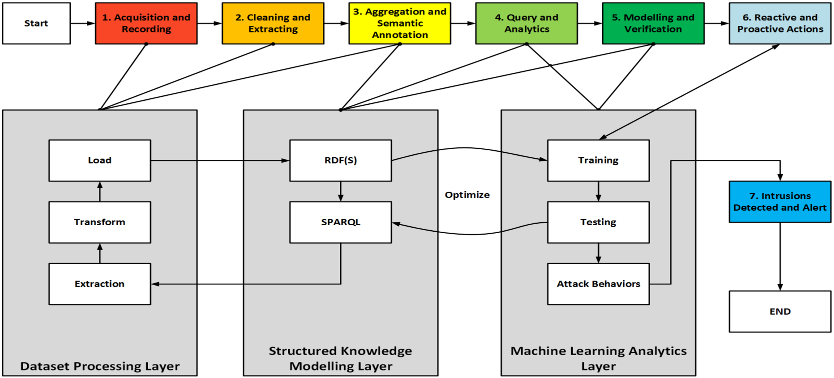


–J. Zhang, et al., Random-Forests-Based Network Intrusion Detection Systems, Sep. 2008

Hybrid Intrusion Detection (Conti.)

- How to combine misuse with anomaly intrusion detection?
- High performance online *misuse* detection engine runs with offline *anomaly* system.
- How to extracting security features to model abnormal signatures and normal behaviors?
- Possible network security features are *packet size, IP addresses, ports, header fields, time stamps, inter-arrival time, session duration, session volume, etc*.

Big Data Analytics Lifecycle for Intelligent Security



Composite Big Data Analytics and Modelling (CBDAM) (Conti.)

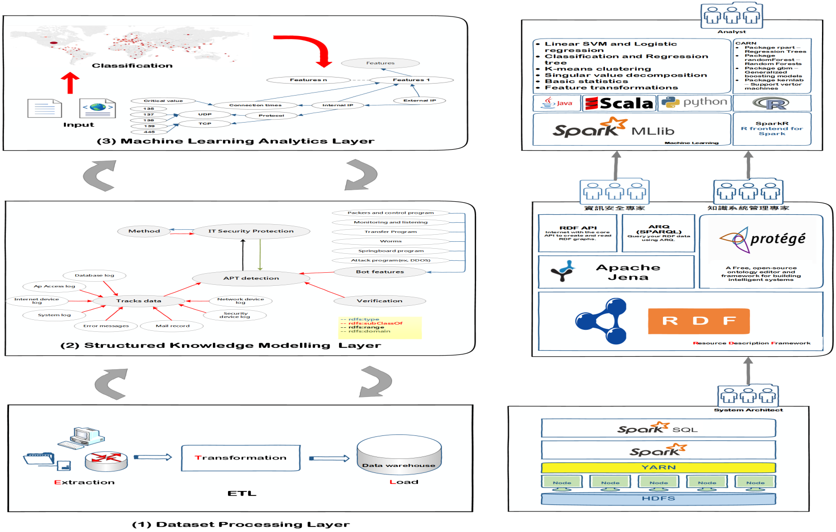


FIGURE: CBDAM Architecture

Composite Big Data Analytics and Modelling (CBDAM) (Conti.)

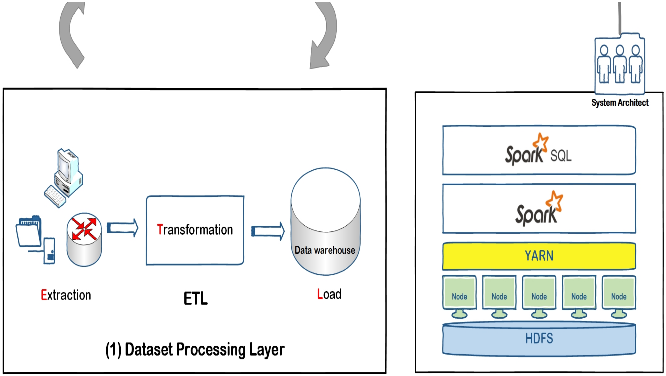


Figure: DAtaset Processing (DAP) layer

Composite Big Data Analytics and Modelling (CBDAM) (Conti.)

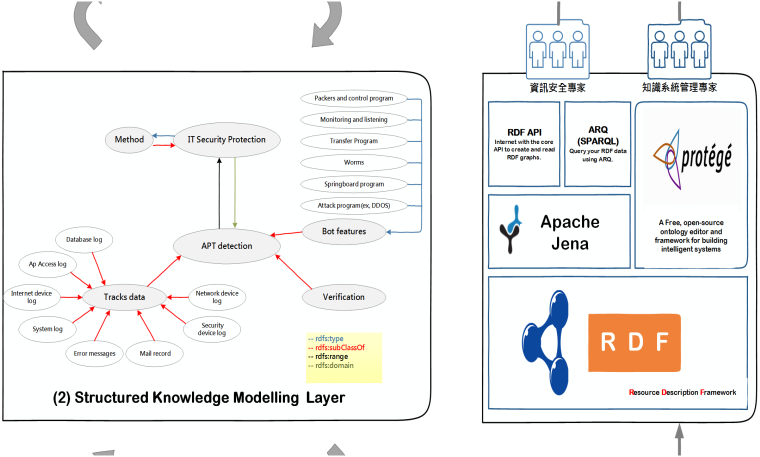


Figure: Structured Knowledge Modelling (SKM) layer

Composite Big Data Analytics and Modelling (CBDAM) (Conti.)

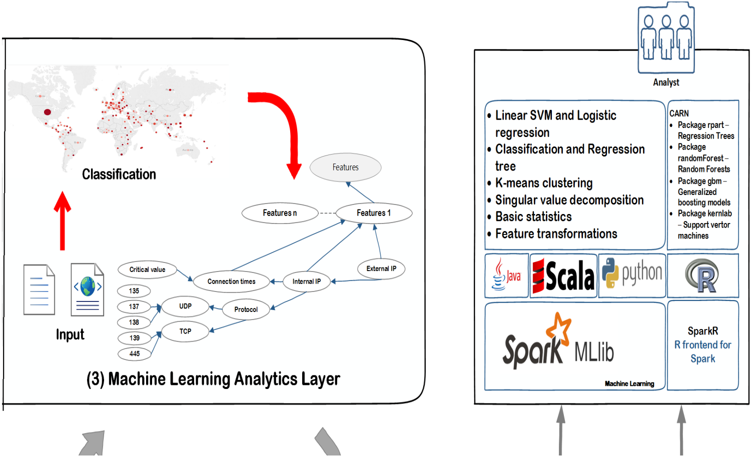


Figure: Machine Learning and Analytics (MLA) layer

Machine Learning with Perfect Domain Knowledge

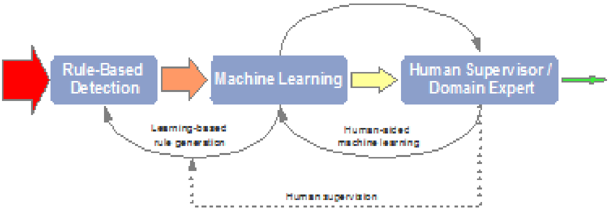


Figure: Security domain knowledge aids machine learning

–Joseph, D. A., et al. Machine Learning Methods for Computer Security. Dagstuhl Perspective Workshop, 2012.

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. What are the important features should be extracted from the big datasets D and model the initial security domain knowledge K to further learn the intrusion behaviors B?
  2. What algorithms exist to learn the security target function $f(x)$ from training instances $x_i \in D_l$, where $D = D_l \cup D_u$,?
  3. How many training datasets D_l are sufficient to offer an acceptable target function $f(x)$ to approximate the true intrusion behaviors B_t?
  4. How to provide and use a minimum amount of labelled training instances $x_l \in D_l$ to learn and classify the intrusion behaviors correctly?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. What are the important features should be extracted from the big datasets D and model the initial security domain knowledge K to further learn the intrusion behaviors B?
  2. What algorithms exist to learn the security target function $f(x)$ from training instances $x_i \in D_l$, where $D = D_l \cup D_u$,?
  3. How many training datasets D_l are sufficient to offer an acceptable target function $f(x)$ to approximate the true intrusion behaviors B_t?
  4. How to provide and use a minimum amount of labelled training instances $x_l \in D_l$ to learn and classify the intrusion behaviors correctly?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. What are the important features should be extracted from the big datasets D and model the initial security domain knowledge K to further learn the intrusion behaviors B?
  2. What algorithms exist to learn the security target function $f(x)$ from training instances $x_i \in D_l$, where $D = D_l \cup D_u$,?
  3. How many training datasets D_l are sufficient to offer an acceptable target function $f(x)$ to approximate the true intrusion behaviors B_t?
  4. How to provide and use a minimum amount of labelled training instances $x_l \in D_l$ to learn and classify the intrusion behaviors correctly?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. What are the important features should be extracted from the big datasets D and model the initial security domain knowledge K to further learn the intrusion behaviors B?
  2. What algorithms exist to learn the security target function $f(x)$ from training instances $x_i \in D_l$, where $D = D_l \cup D_u$,?
  3. How many training datasets D_l are sufficient to offer an acceptable target function $f(x)$ to approximate the true intrusion behaviors B_t?
  4. How to provide and use a minimum amount of labelled training instances $x_l \in D_l$ to learn and classify the intrusion behaviors correctly?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. What are the important features should be extracted from the big datasets D and model the initial security domain knowledge K to further learn the intrusion behaviors B?
  2. What algorithms exist to learn the security target function $f(x)$ from training instances $x_i \in D_l$, where $D = D_l \cup D_u$,?
  3. How many training datasets D_l are sufficient to offer an acceptable target function $f(x)$ to approximate the true intrusion behaviors B_t?
  4. How to provide and use a minimum amount of labelled training instances $x_l \in D_l$ to learn and classify the intrusion behaviors correctly?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. How the prior security domain knowledge K can guide the generalization from the labelled instances $x_i \in D_l$ to correctly predict the unknown instances x_j using labelled training instances $x_i \in D_l$ with noise?
  2. The learner is provided with a perfect security domain knowledge K_p to satisfy the *correct* and *complete* criteria.
  3. What do you mean the learner has the *correct* and *complete* intrusion detection criteria?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. How the prior security domain knowledge K can guide the generalization from the labelled instances $x_i \in D_l$ to correctly predict the unknown instances x_j using labelled training instances $x_i \in D_l$ with noise?
  2. The learner is provided with a perfect security domain knowledge K_p to satisfy the *correct* and *complete* criteria.
  3. What do you mean the learner has the *correct* and *complete* intrusion detection criteria?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. How the prior security domain knowledge K can guide the generalization from the labelled instances $x_i \in D_l$ to correctly predict the unknown instances x_j using labelled training instances $x_i \in D_l$ with noise?
  2. The learner is provided with a perfect security domain knowledge K_p to satisfy the *correct* and *complete* criteria.
  3. What do you mean the learner has the *correct* and *complete* intrusion detection criteria?

Machine Learning with Perfect Domain Knowledge (Conti.)

- Research issues and challenges:
  1. How the prior security domain knowledge K can guide the generalization from the labelled instances $x_i \in D_l$ to correctly predict the unknown instances x_j using labelled training instances $x_i \in D_l$ with noise?
  2. The learner is provided with a perfect security domain knowledge K_p to satisfy the *correct* and *complete* criteria.
  3. What do you mean the learner has the *correct* and *complete* intrusion detection criteria?

Machine Learning with Perfect Domain Knowledge (Conti.)

- K_p can be shown as a combination of ontologies O and rules R to explain the labelled training instances D_l.
- The desired output is a hypothesis $h \in H$ consistent with the labelled training instances $x_i \in D_l$ and the security domain knowledge K_p with acceptable detection capability for unknown instances $x_j \in D_u$.
- Why we need a perfect domain knowledge K_p to model our hypothesis $h_p \in H$?
- However, a perfect domain knowledge K_p with sound and complete criteria is hard to obtain.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Machine Learning with Perfect Domain Knowledge (Conti.)

- K_p can be shown as a combination of ontologies O and rules R to explain the labelled training instances D_l.
- The desired output is a hypothesis $h \in H$ consistent with the labelled training instances $x_i \in D_l$ and the security domain knowledge K_p with acceptable detection capability for unknown instances $x_j \in D_u$.
- Why we need a perfect domain knowledge K_p to model our hypothesis $h_p \in H$?
- However, a perfect domain knowledge K_p with sound and complete criteria is hard to obtain.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Machine Learning with Perfect Domain Knowledge (Conti.)

- K_p can be shown as a combination of ontologies O and rules R to explain the labelled training instances D_l.
- The desired output is a hypothesis $h \in H$ consistent with the labelled training instances $x_i \in D_l$ and the security domain knowledge K_p with acceptable detection capability for unknown instances $x_j \in D_u$.
- Why we need a perfect domain knowledge K_p to model our hypothesis $h_p \in H$?
- However, a perfect domain knowledge K_p with sound and complete criteria is hard to obtain.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Machine Learning with Perfect Domain Knowledge (Conti.)

- K_p can be shown as a combination of ontologies O and rules R to explain the labelled training instances D_l.
- The desired output is a hypothesis $h \in H$ consistent with the labelled training instances $x_i \in D_l$ and the security domain knowledge K_p with acceptable detection capability for unknown instances $x_j \in D_u$.
- Why we need a perfect domain knowledge K_p to model our hypothesis $h_p \in H$?
- However, a perfect domain knowledge K_p with sound and complete criteria is hard to obtain.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Machine Learning with Perfect Security Domain Knowledge (Conti.)

Inductive Learning System (ILS)

$$ILS = (\forall < x_i, f(x_i) > \in D_l)(K'_p \wedge h_p \wedge x_i) \vdash f(x_i)$$

- K'_p: background knowledge for security
- D_l: training datasets
- A hypothesis $h_p \in H$
- f: target function
- $f(x_i)$: target value
- x_i: the i^{th} training instance.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Machine Learning with Perfect Domain Knowledge (Conti.)

Deductive Learning System (DLS)

$$DLS = \begin{cases} (\forall < x_i, f(x_i) > D_l)(h_p \wedge x_i \vdash f(x_i)) \\ D_l \wedge K_p \vdash h \\ (\forall < x_i, f(x_i >) \in D_l)(K_p \wedge x_i) \vdash f(x_i) \end{cases}$$

- K_p: domain knowledge for security
- D_l: training datasets
- A hypothesis $h_p \in H$
- f: target function
- $f(x_i)$: target value
- x_i: the i^{th} training instance.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Machine Learning with imperfect Domain Knowledge (Conti.)

How about an imperfect domain knowledge K_{ip}?

This is an optimization problem

Minimize $argmin_{h \in H}\ \alpha_{D_{ip}} error_{D_{ip}}(h) + \beta_{K_{ip}} error_{K_{ip}}(h)$
where

- $\alpha_{D_{ip}}$ and $\beta_{K_{ip}}$: tunable parameters
- $error_{D_{ip}}(h_{ip})$: the ratio of instances misclassified by h_{ip}
- $error_{K_{ip}}(h_{ip})$: the probability that h_{ip} disagrees with K_{ip} on the classification of an instance.

–Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997

Ontology Learning

Ontology Representation

- What do you mean *ontology*?
- What ontology languages are available?
- How ontology can use security features to describe the concepts of intrusions?
- Why we need SPARQL query in the ontology learning process?

Ontology Learning (Conti.)

Feature Extraction and Representation

- Feature is a basic element of a recognized intrusion pattern.
- Possible feature types are:
  - Selector features
  - Order features
  - Hierarchical features
  - Relational features
  - Set-value features
- Above features should be combined with *time, spatial, sequence order's* contextual features to classify intrusion types.

–J. Fürnkranz, et al., Foundations of Rule Learning, Springer, 2012

Ontology Learning (Conti.)

Feature Extraction and Representation

- Feature is a basic element of a recognized intrusion pattern.
- Possible feature types are:
  - Selector features
  - Order features
  - Hierarchical features
  - Relational features
  - Set-value features
- Above features should be combined with *time, spatial, sequence order's* contextual features to classify intrusion types.

–J. Fürnkranz, et al., Foundations of Rule Learning, Springer, 2012

Ontology Learning (Conti.)

Feature Extraction and Representation

- Feature is a basic element of a recognized intrusion pattern.
- Possible feature types are:
  - Selector features
  - Order features
  - Hierarchical features
  - Relational features
  - Set-value features
- Above features should be combined with *time, spatial, sequence order's* contextual features to classify intrusion types.

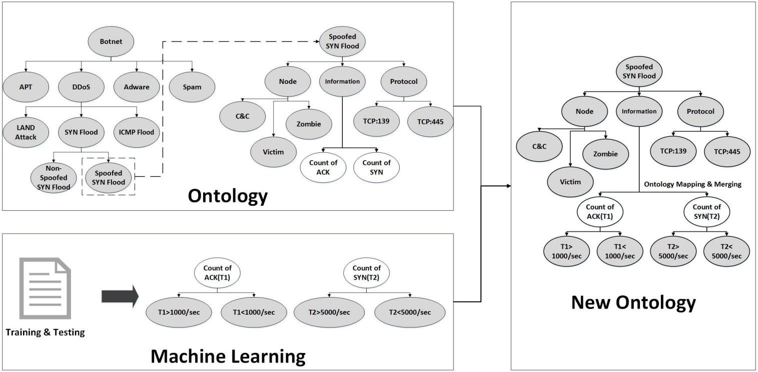
–J. Fürnkranz, et al., Foundations of Rule Learning, Springer, 2012

Ontology Learning (Conti.)

Ontology Learning

- Ontology learning is a concept with instance learning to create, revise, and update the previous ontologies to reflect the status changes.
- Ontology learning is achieved by ontology matching, alignment, and merging from multiple ontologies.
- Ontology learning is an iterative process to reflect and adapt for new datasets.
- Ideally, we need an (semi-)automated ontology learning.

Ontology Learning for Intelligent Security



Rule Learning

Rule Representation

- Rule can be represented as: **If body then head**
  - a **body** contains a conjunction of conditions
  - each condition is a feature satisfaction constraint
  - a **head** contains a prediction with a classification label
- A rule is said to **cover** an positive (or negative) instance if the instance satisfies the conditions of the rule.
- A rule's head is predicted class label or prediction values for an instance if a rule covers this instance.
- If a rule's head only covers the positive instance, so NAF for CWA is assumed.

–J. Fürnkranz, et al., Foundations of Rule Learning, Springer, 2012

Rule Learning (Conti.)

Rule Learning

- A rule learning is (probabilistic) inductive logic program (ILP), statistical relational learning, structured machine learning, etc.
- A single rule learning is for a general to specific principle.
- A ruleset learning is for a specific to general principle.
- How to combine the deductive with inductive reasoning?
- This can be achieved by structured machine learning, how?

Rule Learning (Conti.)

From Ontology to Rule Learning and vice versa

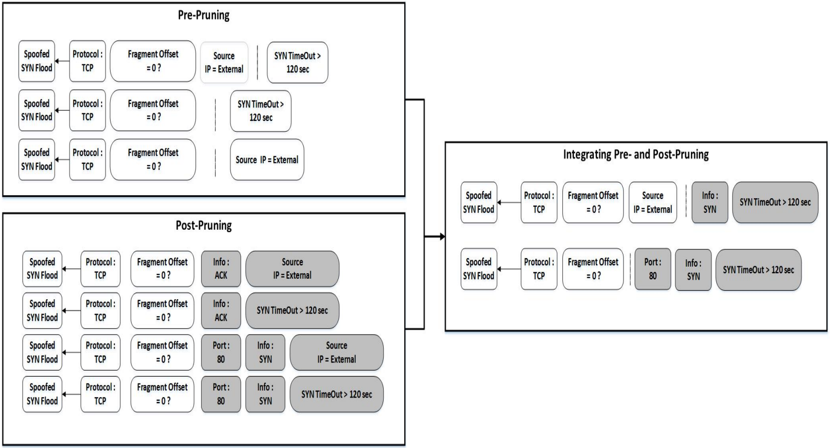
- On ontology learning, the RDF(S)-based ontology schema and instances are imported to the rule module by using SPARQL for classification.
- On rule learning, the rules (or SPARQL queries) enable the ontology module to reformulate the ontologies through approximate query to enable new predicate features.
- RDF(S) graph ontologies with approximate SPARQL query with time and error bounds can learn the new evolving ontologies.

Rule Learning (Conti.)

From Ontology to Rule Learning and vice versa

- Later, we will upgrade to the logic-based ontologies with Datalog rules.
- The evolving ontologies with schema and relationships dynamic creation and removing.
- The rule (or query) learning allows approximate query to discover potential predicate relationships of features with a certain confidence.

Rule Learning for Intelligent Security Enforcement



Conclusion and Future Works

- Preliminary Results:
  1. Structured machine learning in inductive logic program (ILP), since 1990+, has been established at least 20+ years.
  2. Big data analytics is a driving force to rethink about the research challenges.
  3. A combination of ontology with rule learning creates a specific research avenue.
  4. A hybrid intrusion detection application domain is the first problem to verify this concept.

Conclusion and Future Works

- Preliminary Results:
  1. Structured machine learning in inductive logic program (ILP), since 1990+, has been established at least 20+ years.
  2. Big data analytics is a driving force to rethink about the research challenges.
  3. A combination of ontology with rule learning creates a specific research avenue.
  4. A hybrid intrusion detection application domain is the first problem to verify this concept.

Conclusion and Future Works

- Preliminary Results:
  1. Structured machine learning in inductive logic program (ILP), since 1990+, has been established at least 20+ years.
  2. Big data analytics is a driving force to rethink about the research challenges.
  3. A combination of ontology with rule learning creates a specific research avenue.
  4. A hybrid intrusion detection application domain is the first problem to verify this concept.

Conclusion and Future Works

- Preliminary Results:
  1. Structured machine learning in inductive logic program (ILP), since 1990+, has been established at least 20+ years.
  2. Big data analytics is a driving force to rethink about the research challenges.
  3. A combination of ontology with rule learning creates a specific research avenue.
  4. A hybrid intrusion detection application domain is the first problem to verify this concept.

Conclusion and Future Works (Conti.)

- Future Works:
  1. The RDF(S)graph ontology learning with SPARQL rule learning for supervised machine learning, e.g., random forest and boosting, are considered first to verify the intelligent security problem.
  2. The SPARK with RDF(S) and SPARQL platform have been establishing for the CBDAM platform.
  3. Using the logic-based ontology and rules to verify structured machine learning concepts will be the next research challenge.

Conclusion and Future Works (Conti.)

- Future Works:
  1. The RDF(S)graph ontology learning with SPARQL rule learning for supervised machine learning, e.g., random forest and boosting, are considered first to verify the intelligent security problem.
  2. The SPARK with RDF(S) and SPARQL platform have been establishing for the CBDAM platform.
  3. Using the logic-based ontology and rules to verify structured machine learning concepts will be the next research challenge.

Conclusion and Future Works (Conti.)

- Future Works:
  1. The RDF(S)graph ontology learning with SPARQL rule learning for supervised machine learning, e.g., random forest and boosting, are considered first to verify the intelligent security problem.
  2. The SPARK with RDF(S) and SPARQL platform have been establishing for the CBDAM platform.
  3. Using the logic-based ontology and rules to verify structured machine learning concepts will be the next research challenge.