HASTE:
Human-centric Active-learning for decision Support in Threat Exploration

Presentation by

Dr. Phil Legg
Prof. Jim Smith
Dr. Richard Preen

UWE Bristol | University of the West of England

19/09/2018

Research Questions

- How can interactive machine learning and visualisation techniques aid analysis and understanding in complex threat exploration tasks?

- Can the machine facilitate better data exploration and understanding by learning and exploiting multi-modal interactions of the user?

- What can the user learn about the machine’s capability of decision-making through the inspection of how decisions are computed?

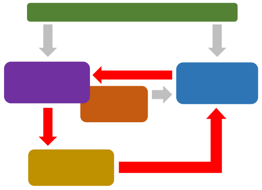
- In contrast to traditional batch learning, can an active learning approach help improve accuracy, time required, and trust, for both parties?

UWE Bristol University of the West of England

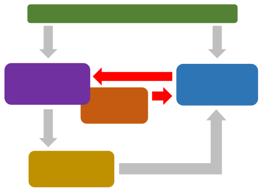
HASTE Concept

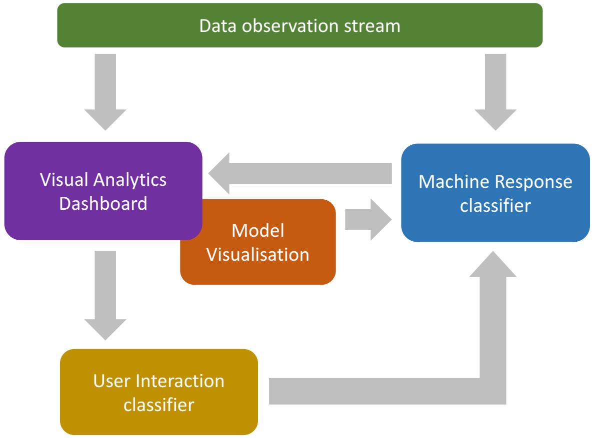
Given incoming data, HASTE has two forms of utility:

- If the machine is unconfident, query class with user. User can inspect data using visual tools and provide response. Machine observes user interactions to learn **how** response was formulated.



- If the machine is confident in classification, assign class to observation and inform user. User can inspect decision and refine if needed. Machine to try learn **why** it was incorrect.





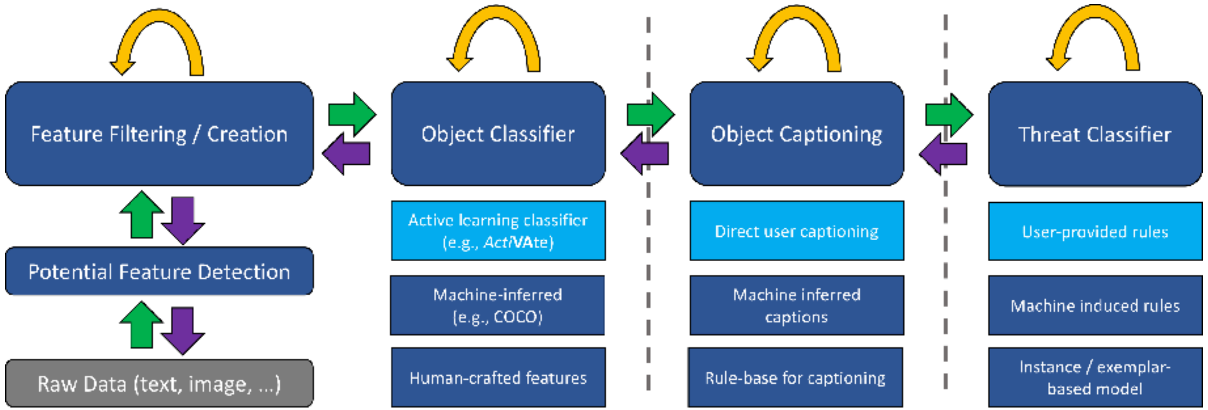
Approach

- **DSTL Phase 1**: Developing a proof-of-concept tool that can support research and demonstrate the HASTE concept

- Phase 1 use cases:
  - Image-based Road Hazard Exploration
  - Text-based exploration of news articles
  - Active learning for exploration of object (mis-) classification

- With richer datasets and use cases, we can envisage different modes of utility for how data observations may require rapid analysis and response
  - To be explored for later TRL development phases

Approach



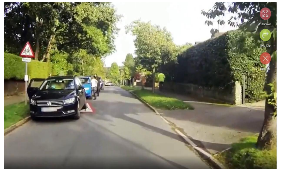
How can a low-level data observation be transformed into a high-level concept such as whether a threat is posed?

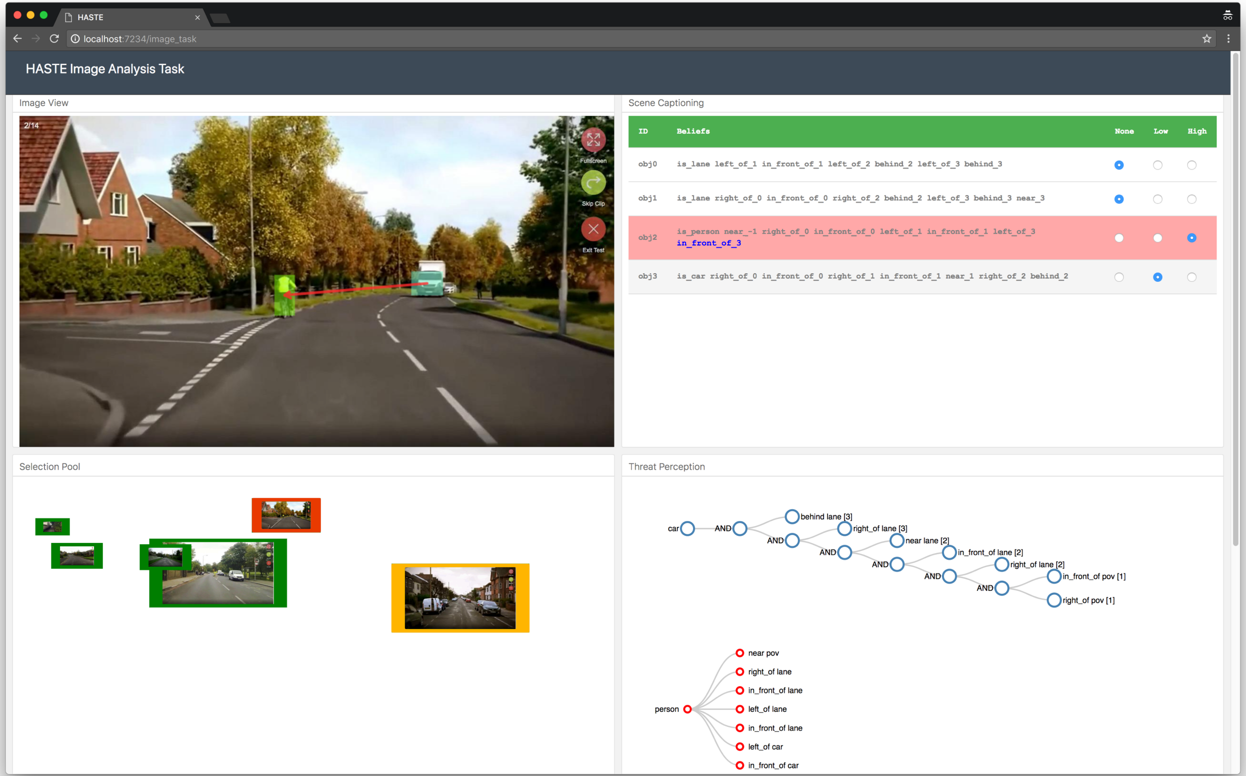
Modular system design to allow interchangeable use of different components (e.g., different object classifiers, data types, feature types, etc.).

Road Hazard Exploration

- Which “objects” are threats and why?
  - How do humans identify hazards and how can machines mimic?

- **Object detection** – using a combination of detection models (to integrate both common + bespoke objects)
- **Relationship detection** – spatial / temporal / behavioural.
- **Semantic graph** – descriptive model of the image: objects and relationships.
- **Threat classifier** – receiving a unique description of each object in the image.
- **Human-in-the-loop** – selecting, labeling, filtering, creating --> understanding




HASTE
localhost:7234/image_task
HASTE Image Analysis Task
Image View
2/14
Fullscreen
Skip Clip
Exit Test
Scene Captioning
ID
Beliefs
None
Low
High
obj0
is_lane left_of_1 in_front_of_1 left_of_2 behind_2 left_of_3 behind_3
obj1
is_lane right_of_0 in_front_of_0 right_of_2 behind_2 left_of_3 behind_3 near_3
obj2
is_person near_-1 right_of_0 in_front_of_0 left_of_1 in_front_of_1 left_of_3 in_front_of_3
obj3
is_car right_of_0 in_front_of_0 right_of_1 in_front_of_1 near_1 right_of_2 behind_2
Selection Pool
Threat Perception
car
AND
behind lane [3]
right_of lane [3]
near lane [2]
in_front_of lane [2]
right_of lane [2]
in_front_of pov [1]
right_of pov [1]
person
near pov
right_of lane
in_front_of lane
left_of lane
in_front_of lane
left_of car
in_front_of car

UWE Bristol University of the West of England

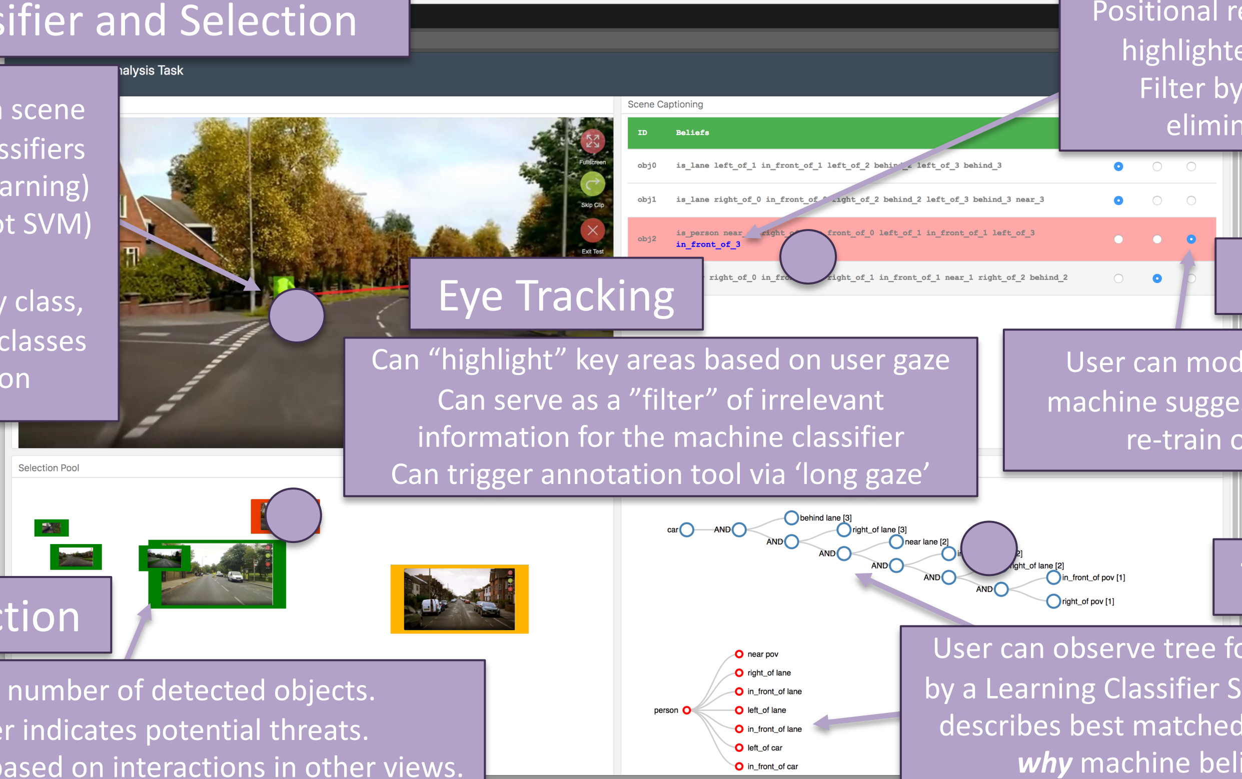
Scene / Object Captioning

Positional relationships – key aspects highlighted and shown on image
Filter by coloured selection to eliminate weak indicators

Object Classifier and Selection

Objects detected in scene using ensemble classifiers
(e.g., COCO deep learning)
(e.g., bespoke 1-shot SVM)

objects coloured by class, annotation of new classes via user selection



Eye Tracking

Can “highlight” key areas based on user gaze
Can serve as a ”filter” of irrelevant information for the machine classifier
Can trigger annotation tool via ‘long gaze’

Threat Class

User can modify if they disagree with machine suggestion – machine will then re-train on new information

Threat Reasoning

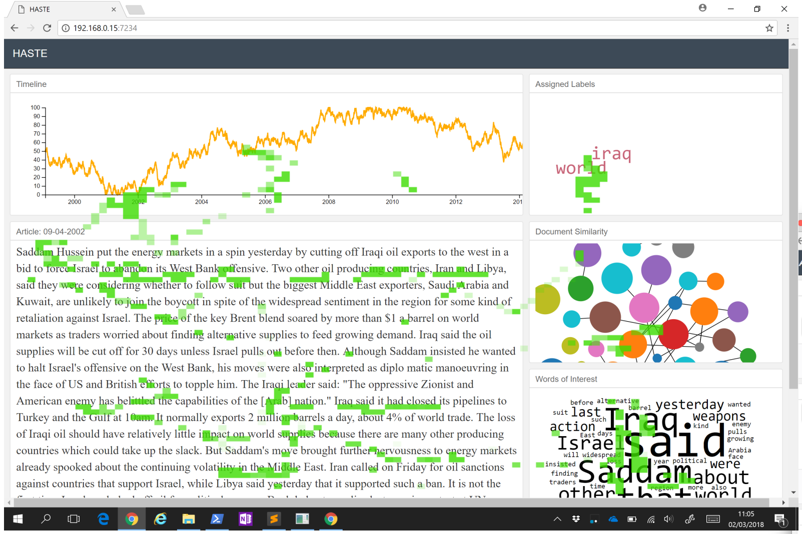
User can observe tree for each objects as evolved by a Learning Classifier System (LCS) over time that describes best matched rule for threat class (i.e., ***why*** machine believes this is threat).

Sample Selection

Size indicates number of detected objects.
Colour border indicates potential threats.
Filtering / retrieval based on interactions in other views.

UWE Bristol University of the West of England

Additional HASTE Case Studies



Text analysis – understanding complex concepts in crisis news articles

Understanding (mis) classification in machine learning applications

Outcomes and Benefits

- Proof-of-Concept demonstrator tool
  - Interface maps to process of how threats are identified and analysed
  - User can explore threats to inform machine of threat classifications
  - Machine can iteratively learn from each user interaction as new samples are observed to contribute towards model
    - ***why*** a threat is posed
  - Machine can recognize human interaction patterns for what may constitute a threat, and can model semantic relationships between objects in scene
    - ***how*** user identifies threat

- Currently piloting user studies on decision / classifier explainability through the use of the evolved threat trees

Future Requirements

- We wish to explore richer datasets with more tailored challenges for defence and security needs.
  - *How can the HASTE concept be deployed with existing 'dashboard' tools to better integrate user analytics and machine collaboration in current practice?*

- We wish to further explore how human observation data can be integrated to inform decisions (using eye tracking and/or EEG).
  - *Currently, eye tracking serves as a 'filter' of weak indicators. More to be done on how best to learn about the sequence of eye-tracking, and how this becomes generalizable for future observation tasks*

Thank you

Dr. Phil Legg
Phil.Legg@uwe.ac.uk
http://go.uwe.ac.uk/phil
@dr_plegg

UWE Bristol
University of the West of England

HASTE
Supplementary Material

Object Threat Detection

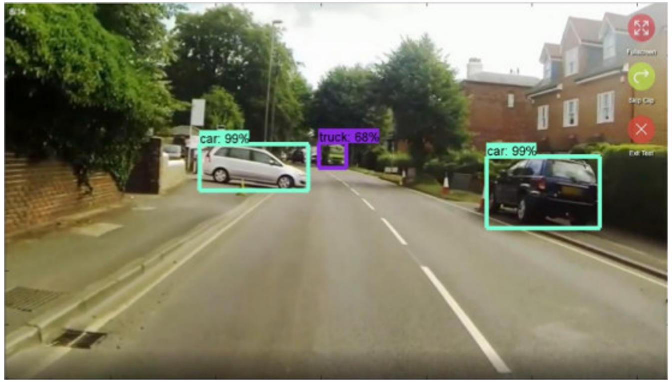
- Given an image with multiple objects
  - Which ones are threats? Why?
- Road hazard perception example.

- **Object detection** – using a combination of detection models.
- **Relationship detection** – spatial / temporal / behavioural.
- **Semantic graph** – descriptive model of the image: objects and relationships.
- **Threat classifier** – receiving a unique description of each object in the image.
- **Human-in-the-loop** – selecting, labeling, filtering, creating – understanding.

Object Detection

- A combination of detection models.

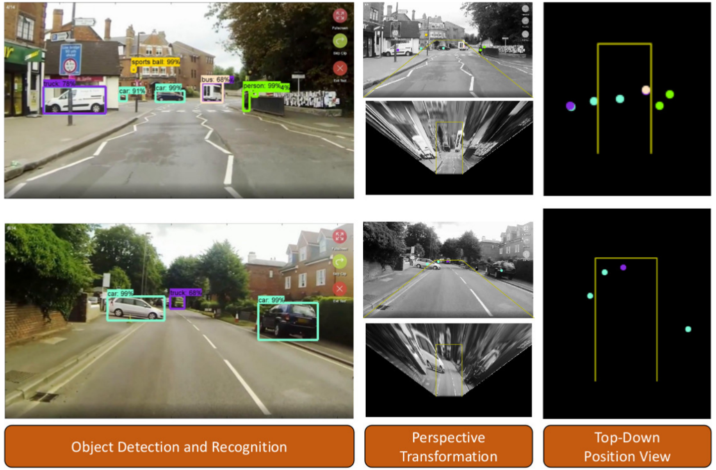
- Big data: pre-trained **offline** models, where large pre-existing data available.
  - (Re)use of general models: e.g., MS COCO, pre-trained on 90 common objects.
  - Leverage existing training data of domain-specific object types.
    - E.g., convolutional neural network trained on labeled crossing patrol officers.
  - Accurate detection of previously seen objects that are uniform in appearance.



- Small data: **online** learning, where little or no data available.
  - Leverage human generated labeling at runtime.
  - Less accurate, but enables the detection of previously unseen or frequently changing object types.

Relationship Detection

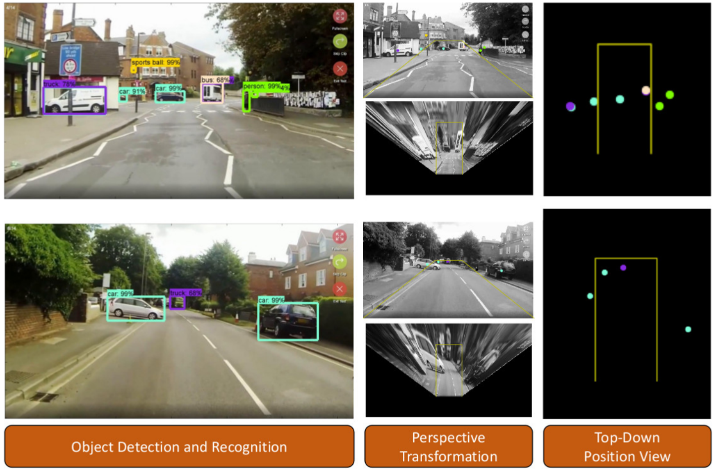
- Detection of spatial / temporal / behavioural relationships **between** objects.
- Perspective transformation – e.g., aerial view to restore the lost depth dimension.



Relationship Detection

- Detection of spatial / temporal / behavioural relationships **between** objects.

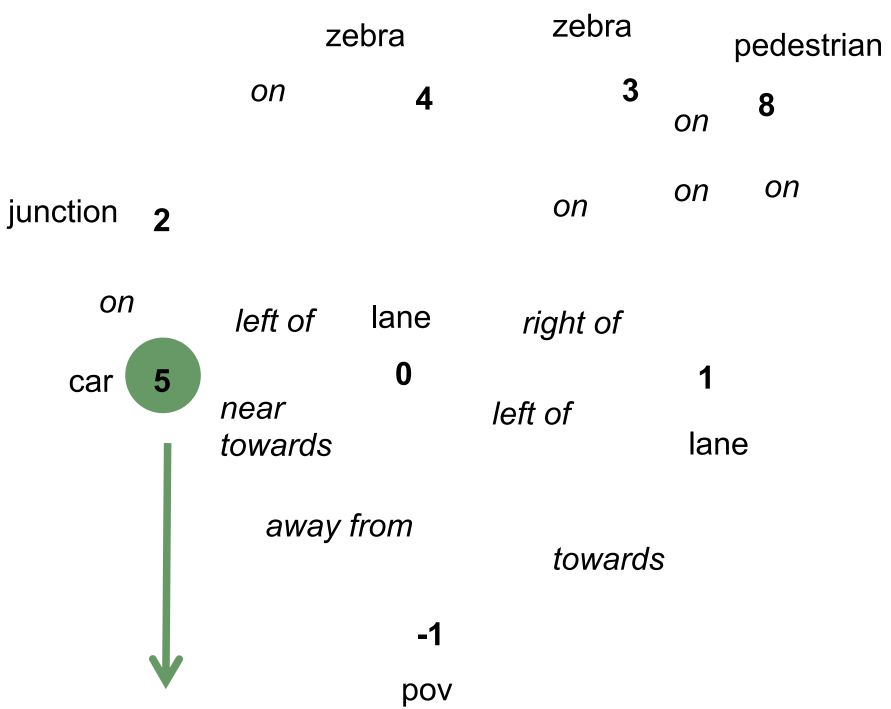
- Conversion of precision numbers to human-interpretable fuzzy relation sets:
    - ***x*-axis position:** *left of, right of*
    - ***y*-axis position:** *behind, in front of*
    - ***z*-axis position:** *above, below*
    - **Overall distance:** *near, far from, on*
    - **Direction:** *towards, away from*



UWE Bristol University of the West of England

Semantic Graph

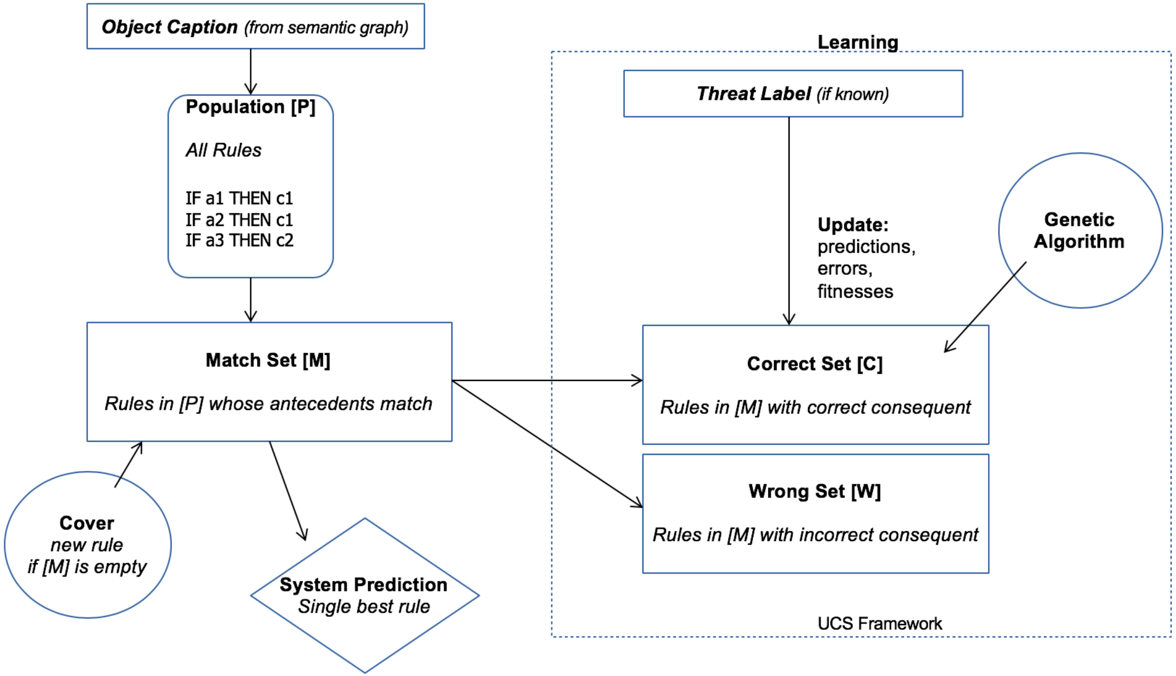
- The semantic graph generates unique descriptions of each object in the image.
- *n*-level graph expansion.
  - Performed for each desired object.
  - More levels = longer and detailed.
- Object descriptions / captions become the *inputs to the threat classifier*.



<u>2-level expansion of object 5:</u>
is **car**,
on [*is* **junction**, *left of* **lane**],
near [*is* **lane**, *away from* **pov,** *left of* **lane**],
towards [*is* **lane**, *away from* **pov,** *left of* **lane**]

Threat Classifier

- Learning Classifier System – Evolves an ensemble of rules



Classifier Rules

- Rule antecedents encoded as trees:
  - Each rule has a match TYPE (car, pedestrian, etc.)
  - Each rule has its own set of (abstract) object types [A, B, C, ...]
    - Referenceable by the main tree: e.g., *near* A **AND** *towards* B
    - Also encoded as trees with a match type.
    - Evaluates True if a matching (concrete) object found within the image.
  - Can be viewed as a search pattern.

  - BOOLEAN OPERATORS = [AND, OR, NAND, NOR, TRUE]
  - PRIMITIVES composed of FUZZY SET and TYPE SET
    - FUZZY SET = [on, near, far from, away from, towards, ...]
    - TYPE SET = [pov, agent, vehicle, car, truck, pedestrian, …]
- Rule consequents: [no, low, high]

Example Classifier Rule

- Fitness = 0.2088
- Prediction = 1000.0
- Error = 0.0
- Numerosity = 11
- Experience = 116
- Correct = 116
- Set size = 45.99
- Time = 199
- Human = False