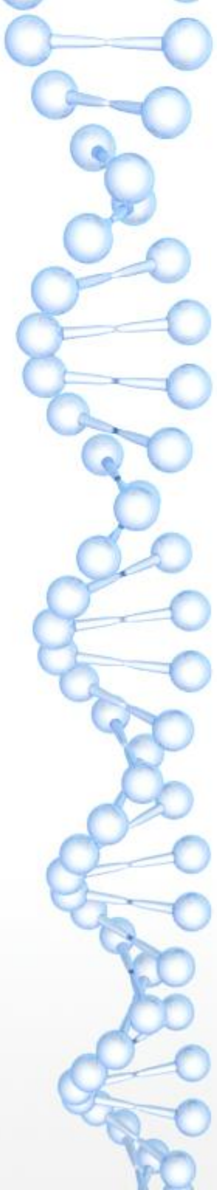


DATA FUSION
FOR BETTER DECISION MAKING

PhD student JAROSLAVA ARSENJEVA

Review for 2018/2019 study year

Thesis advisor: Viktor Medvedev, Dr.

Thesis consultant: Gintautas Dzemyda, Prof. Habil. Dr.

Informatics engineering field (T007)

The time of doctorate studies: 2018 – 2022 years

- **Object of reserch:**

  Data fusion methods.

- **Research goal:**

  To propose a data fusion algorithm (framework) to improve decision making in medicine using data mining methods.

Research tasks

- Perform an analytical review of data fusion existing methods;
- Select appropriate data for further research and data fusion method application;
- Identify main scientific problems (for example, in the medical field) in data fusion tasks (where data from different sensors is aggregated into one framework);
- Select appropriate data fusion methods for previously selected data;
- Develop a data fusion algorithm to improve decision making that would be applied to previously selected data;
- Evaluate the results of the proposed algorithm, make necessary changes.

Planned results

- Identification of proper data fusion methods to improve decision making
- Proposition and further application of a data fusion algorithm (framework), that would improve decision making using data obtained from mining techniques.

Work plan tasks for 2018/2019

- Perform a primary analytical review of data fusion methods;
- Identify scientific problems in data fusion tasks.
- Thesis prepartion stage: create a review of the topic.
- Participate in a summer/ winter PhD school.

- Pass 2 exams:
  - Decision Making Strategies (7 ECTS)
  - Research Methods and Methodology in Informatics and Computer Engineering (8 ECTS)

Completed tasks 2018/2019

- Perform a primary analytical review of data fusion methods;
- Identify scientific problems in data fusion tasks.
- Thesis prepartion stage: create a review of the topic.

- Pass 2 exams:
    - Decision Making Strategies (7 ECTS) – 10 (excellent)
    - Research Methods and Methodology in Informatics and Computer Engineering (8 ECTS) – 7 (satisfactory)

Participate in a summer/ winter school

COST action TD1403 PhD winter school „Big Data in Simulations and Observations“ in Turku, Finland from 26 November 2018 to 1 December 2018.

	Monday Nov26	Tuesday Nov27	Wednesday Nov28	Thursday Nov29	Friday Nov30
09:00 - 10:30	Introduction to Machine Learning - Julien	Deep Learning Workshop - Nima		Antti Penttilä: Data science methods in Solar System small body observations	Victor Debattista and Samuel Earp: machine learning, finding bars in big simulations
10:30 - 11:00	Introduction to Machine Learning - Julien	Deep Learning Workshop - Nima	Talk (Särkkä Simo): Bayesian filtering and smoothing	Antti Penttilä: Data science methods in Solar System small body observations	Victor Debattista and Samuel Earp: machine learning, finding bars in big simulations
11:00 - 11:20	coffee break	coffee break	Talk (Särkkä Simo): Bayesian filtering and smoothing	coffee break	coffee break
11:20 - 1300	Introduction to Machine Learning - Julien	Deep Learning Workshop - Nima	Talk (Särkkä Simo): Bayesian filtering and smoothing	Janne: workflow from photographic meteor data to the reconstruction of	Victor Debattista and Samuel Earp: machine learning, finding bars in big simulations
13:00 - 14:00	lunch	lunch	lunch	lunch	lunch
14:00 - 15:20	Introduction to Deep Learning - Nima	Nima - TBD	exercises (Särkkä Simo): Bayesian filtering and smoothing	Maria: Classification of meteor events in large databases	Victor Debattista and Samuel Earp: machine learning, finding bars in big simulations
15:20 - 15:40	coffee break	coffee break	coffee break	coffee break	coffee break
15:40 - 17:00	Deep Learning Use Case in Astro 1&2: TransiNet - Nima	Nima - TBD	Maria, Simo, Janne, case study: analysis of meteor trajectories	Janne: illustrated cases by MeTra Meteor Trajectory Calculator	Victor Debattista and Samuel Earp: machine learning, finding bars in big simulations

Certificate

This is to certify that

Jaroslava Arsenjeva

has succesfully completed

The BigSkyEarth Training School, organized by the
Big Data Era in Sky and Earth Observations COST Action TD 1403
https://bigskyearth.eu/category/training-school/
held in Kaarina, Finland, Nov. 26 – Dec 1. 2018

Kirsi Lehto

Work plan for 2019/2020

- Create a deep overview of data fusion problems and their solutions in the medical field
- Develop a research methodology (Selection of research methodology for solving the chosen task)
- Investigate data fusion methods used in data research
- Identify additional scientific issues arising
- Dissertation preparation stage: conducting research

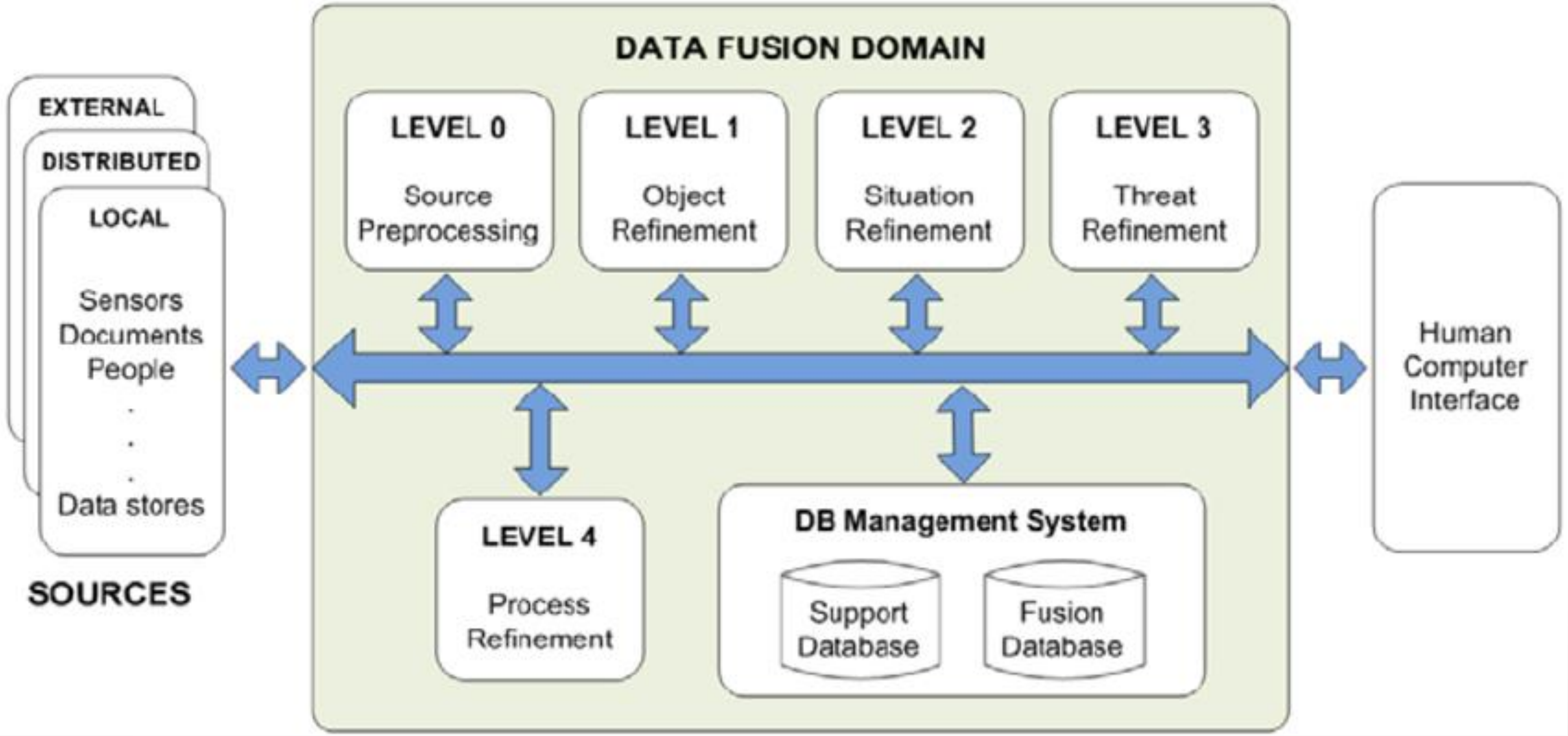
Work plan for 2019/2020

- Participate and present research results at an international scientific conference in Lithuania or abroad (
- Pass two exams:
    - Machine learning (7 ECTS credits);
    - Fundamental Methods in Informatics and Computer Engineering (8 ECTS credits)
- Publish a paper in a peer-reviewed periodical journal

Data Fusion

Data fusion is the process of getting data from multiple sources in order to build more sophisticated models and understand more about a project. It often means getting combined data on a single subject and combining it for central analysis. (Technopedia)

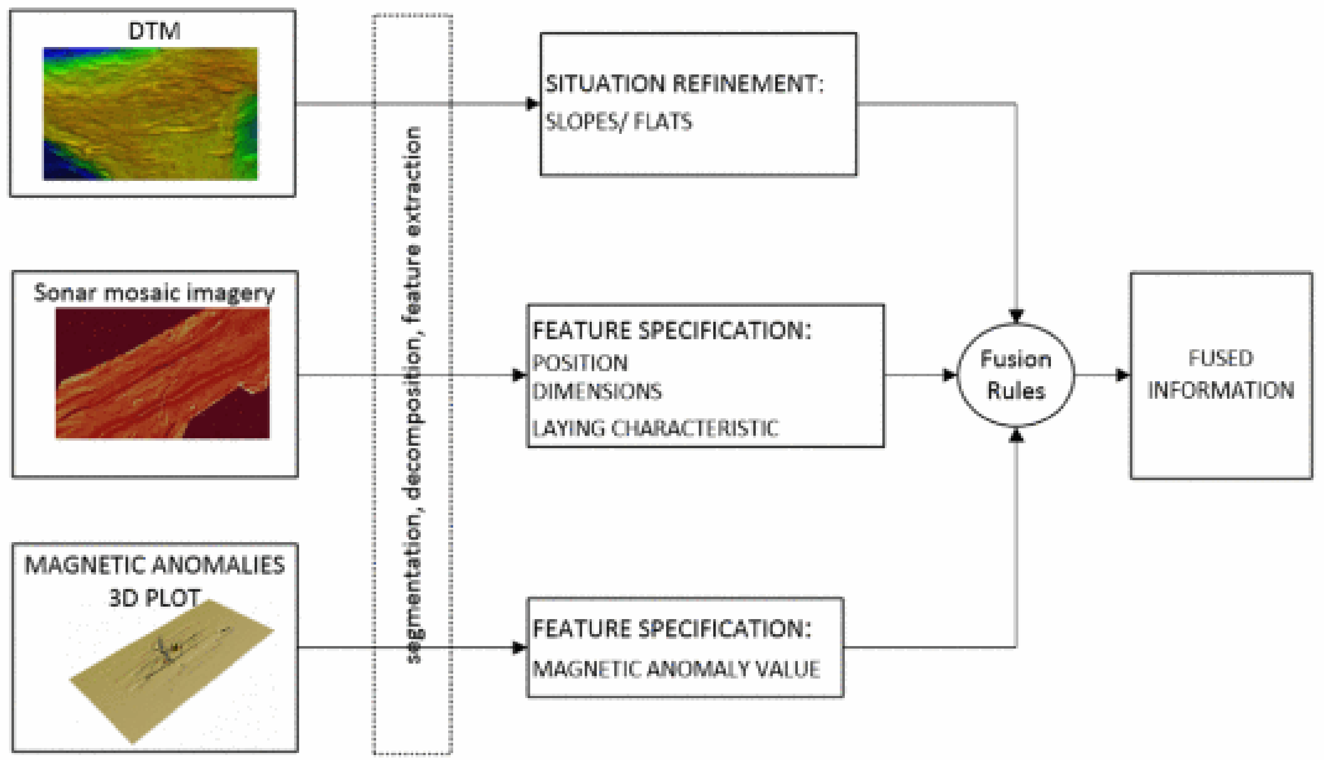
JDL model (1985)



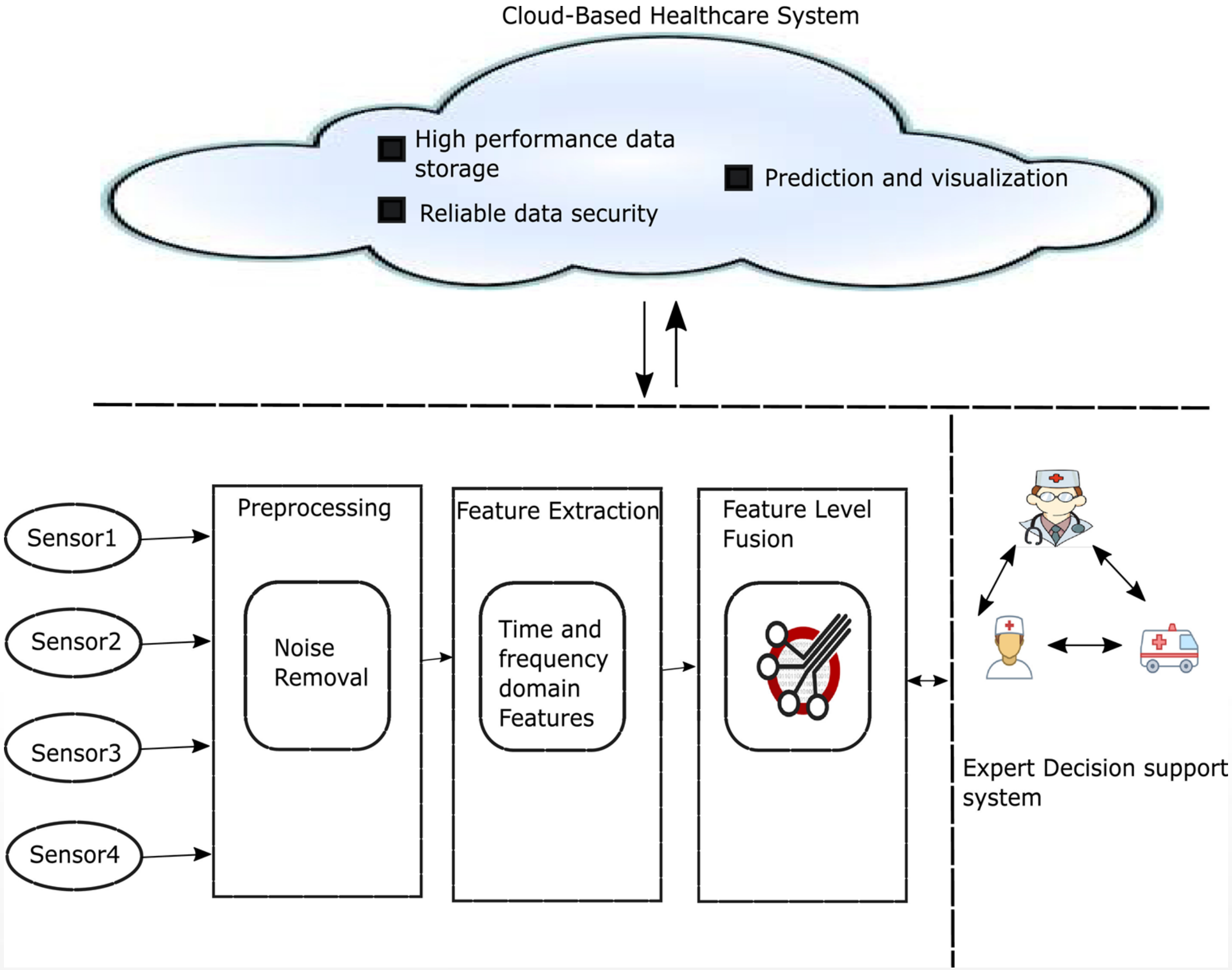
Application fields

- Agriculture
- Medicine
- Industry (automotive)
- Military
- others

Agriculture

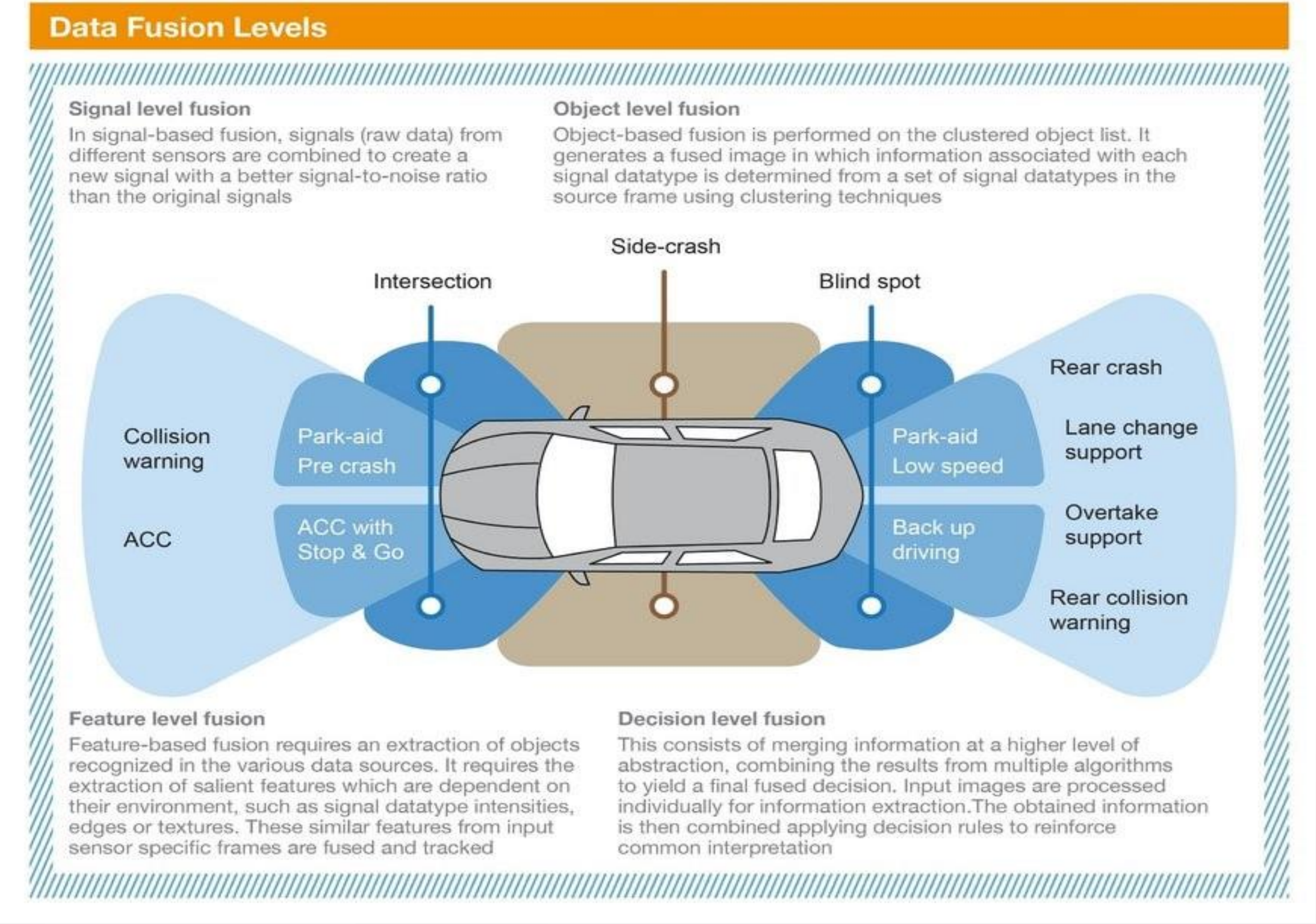


Medicine



Cloud-Based Healthcare System
High performance data storage
Reliable data security
Prediction and visualization
Sensor1
Sensor2
Sensor3
Sensor4
Preprocessing
Noise Removal
Feature Extraction
Time and frequency domain Features
Feature Level Fusion
Expert Decision support system

Industry


Data Fusion Levels
Signal level fusion
In signal-based fusion, signals (raw data) from different sensors are combined to create a new signal with a better signal-to-noise ratio than the original signals
Object level fusion
Object-based fusion is performed on the clustered object list. It generates a fused image in which information associated with each signal datatype is determined from a set of signal datatypes in the source frame using clustering techniques
Side-crash
Intersection
Blind spot
Collision warning
Park-aid
Pre crash
ACC
ACC with Stop & Go
Rear crash
Lane change support
Park-aid
Low speed
Overtake support
Back up driving
Rear collision warning
Feature level fusion
Feature-based fusion requires an extraction of objects recognized in the various data sources. It requires the extraction of salient features which are dependent on their environment, such as signal datatype intensities, edges or textures. These similar features from input sensor specific frames are fused and tracked
Decision level fusion
This consists of merging information at a higher level of abstraction, combining the results from multiple algorithms to yield a final fused decision. Input images are processed individually for information extraction.The obtained information is then combined applying decision rules to reinforce common interpretation

Military


Data Link
Command and Control
AESA Radar
Active Electronically
Scanned Array
IR Missile
Seekers
Optronics
> Identification
> Targeting
Intra-Flight
Data Link
Data Link
Ground Station
SPECTRA
Electronic
Warfare Suite
> Digital / Internal
> Multi Threats
NETWORK CENTRIC OPERATIONS

Issues

Can be fixed

- Heterogeneous data
- Different measuring rates
- Noisy data
- Contradictive data
- Incomparable dimensionality

Cannot be fixed

- No good sensor substitute
- Downstream cannot makeup for upstream failures
- No ideal algorithm
- Sensors must have weights/ accuracy estimators
- The inference accuracy is hard to quantify

Algorithms I

Raw data fusion

K-Nearest neighbors

PDA

Joint PDA

Feature fusion

Maximum likelihood

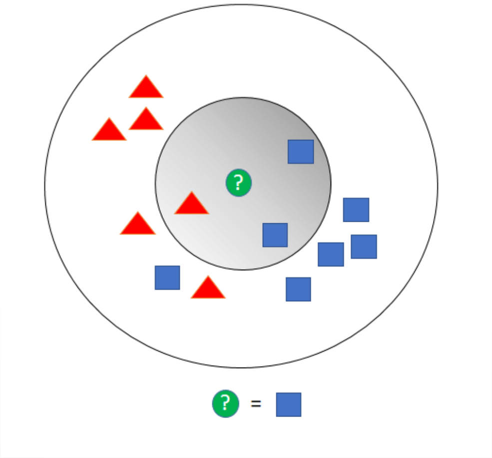
Maximum a posteriori

Kalman filter

Particle filter

K-Nearest Neighbors

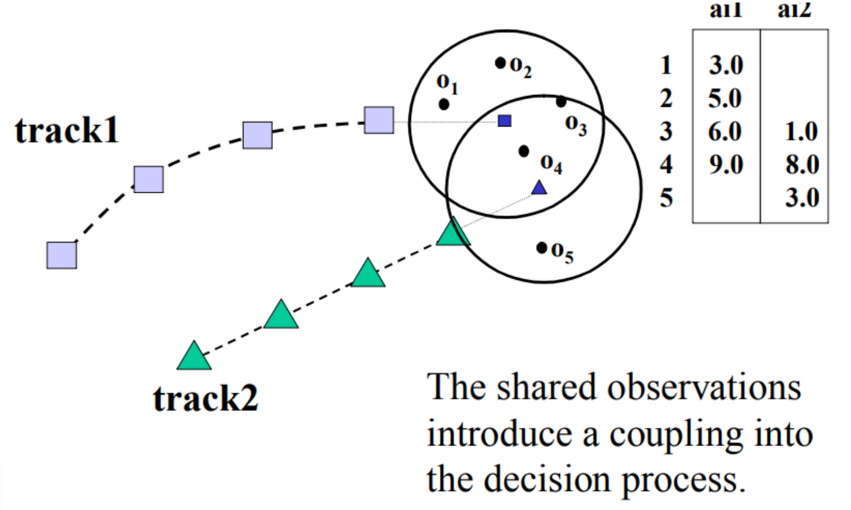
(KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems



PDA

Used for target tracking;

Updates based on all observations, weighted by their likelihoods.



Maximum likelihood

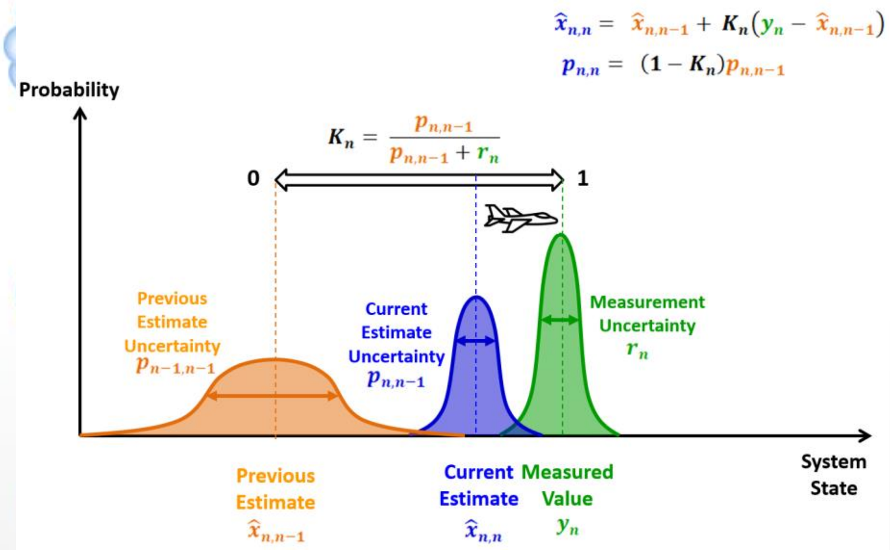
Choose value that maximizes the probability of observed data

Maximum a posteriori

Choose value that is most probable given observed data and prior belief

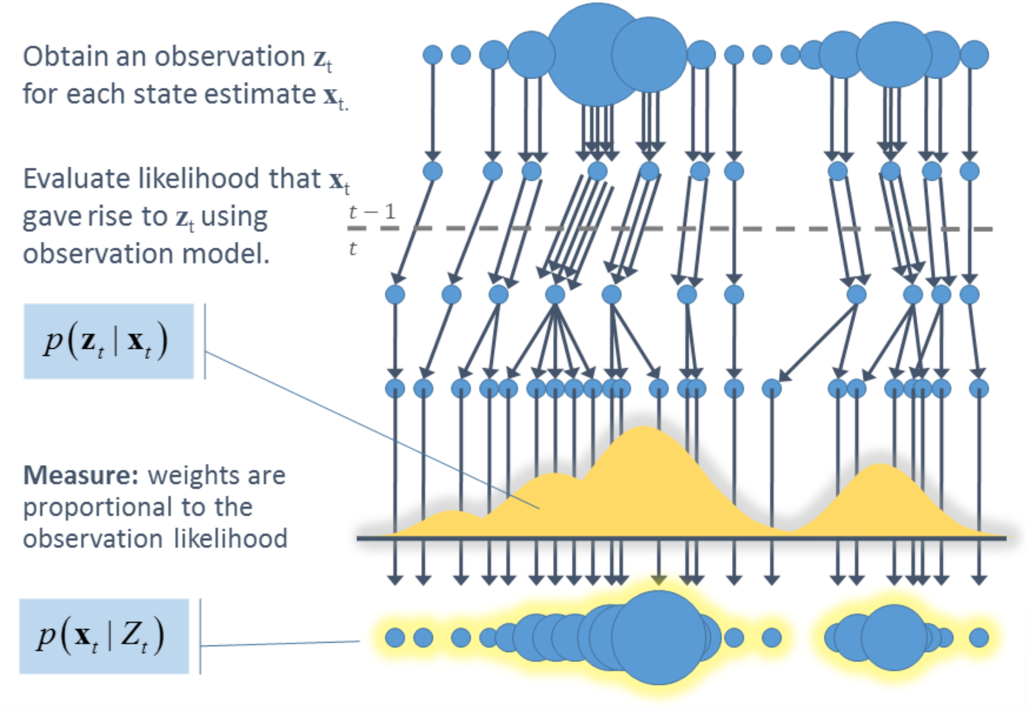
Kalman filter

Is used when you have uncertain information about some dynamic system, and you can make an educated guess about what the system is going to do next



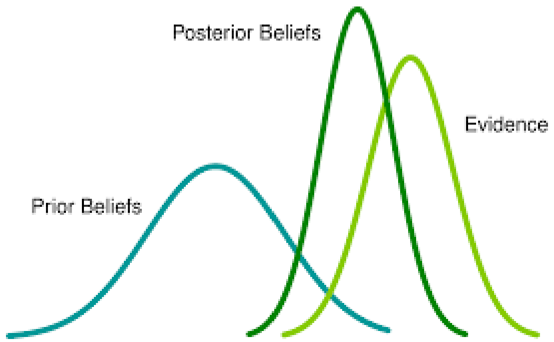
Particle filter

Non-linear Kalman



Algorithms II

- Bayesian theorem
- P(θ|D)=(P(D|θ) X P(θ))/P(D)



- Demster-Shafter Inference
- is a generalization of probability theory that allows for incomplete knowledge

References

- https://www.researchgate.net/figure/Joint-Directors-of-Laboratories-JDL-model-Level-0-Preprocessing-of-sensor_fig1_221787812
- https://www.techopedia.com/definition/32735/data-fusion
- https://ieeexplore.ieee.org/abstract/document/7226263
- https://www.sciencedirect.com/science/article/pii/S1566253519302118
- https://www.visteon.com/current-sensor-data-fusion-architectures-visteons-approach/
- https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/
- https://www.researchgate.net/figure/k-Nearest-Neighbor-Algorithm_fig2_328738427
- http://www.cse.psu.edu/~rtc12/CSE598C/datassocPart2.pdf
- https://www.sciencedirect.com/topics/engineering/particle-filter
- https://www.stats.ox.ac.uk/~doucet/doucet_johansen_tutorialPF2011.pdf
- https://towardsdatascience.com/particle-filter-a-hero-in-the-world-of-non-linearity-and-non-gaussian-6d8947f4a3dc