Multi-Modal Data Fusion

Biomimetics and Intelligent Systems Group
Jaakko Suutala, Markus Harju

Lecture 1: Introduction

Outline

- Data fusion concepts and motivation
- Data fusion strategies
- Data fusion framework
- Machine learning primer
- Probability primer (statistical modelling)

Data Fusion

- data fusion is about combining (sensor) data for more comprehensive and accurate information
- the data usually originates from some physical sensors (or derived from sensors): thermometer, camera, microphone, magnetometer, radar, lidar,...
- we prefer "data fusion" over "sensor fusion" as data might come from other sources too...
- multi-modal means that the data is of different variety (e.g. video, image, sound, tabular data, numbers, text,...)

UNIVERSITY OF OULU

Inspiration

- humans and animals sense their environment by touch, smell, taste, sound and sight
- their brain is able to fuse this information and infer about our surrounding world and better decisions are made
- in computer systems fused data is used in machine vision, AI, machine learning,...

Practical Applications

- autonomous vehicles, driver assists (cameras, radars, lidars,...)
- speech recognition (audio+video)
- human identification (fingerprint photo, facial photo, voice recording)
- video surveillance, object tracking (optical, infrared sensors)

Fusion strategies

Fusion type refers to how fusion is performed.

- Across sensors: several sensors measure the same property. E.g. temperature of an object is measured by several different devices
- Across attributes: sensors measure different properties. E.g. temperature, pressure and humidity of air.
- Across domains: sensors measure same attribute over different ranges or domains.
- Across time: past and present measurements are fused together. E.g. calibration of a device

Sensor configuration

Another characterization of fusion process takes a look at the sensor configuration:

- Complementary: sensors do not direcly depend on each other, but they can be combined into giving a more complete picture of the phenomenon. E.g. multiple cameras observing different parts of a scene, or audio and video used in speech recognition
- Competitive: each sensor provides an independent measurement of the same property. Or one sensor measures at different time instants.
- Co-operative: two or more sensors contribute information not available from a single sensor. E.g. combine two images from different viewpoints to form 3D image. Most diffcult, less accurate, less reliable.

Input/output characteristics

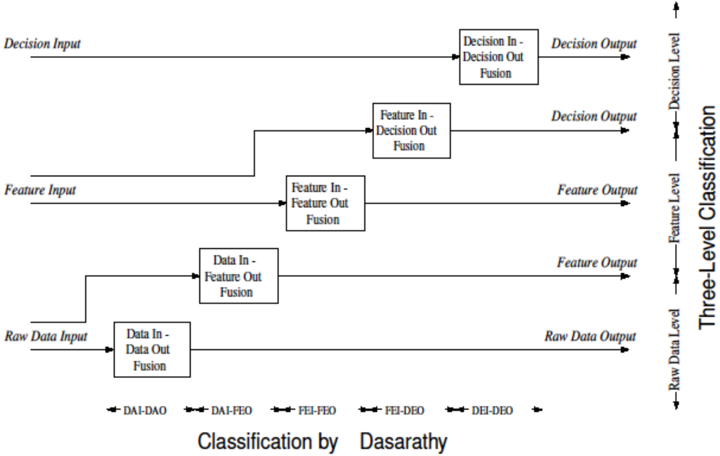


Figure 2.2: Alternative fusion characterizations based on input/output

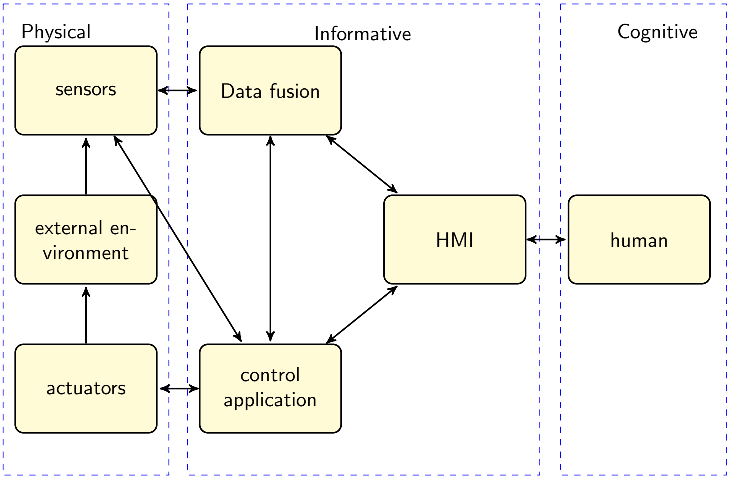
(Image by Wilfried Elmenreich, dissertation, 2002)

Data fusion benefits

- increased robustness and reliability
- greater spatial and temporal coverage
- higher confidence
- reduced uncertainty

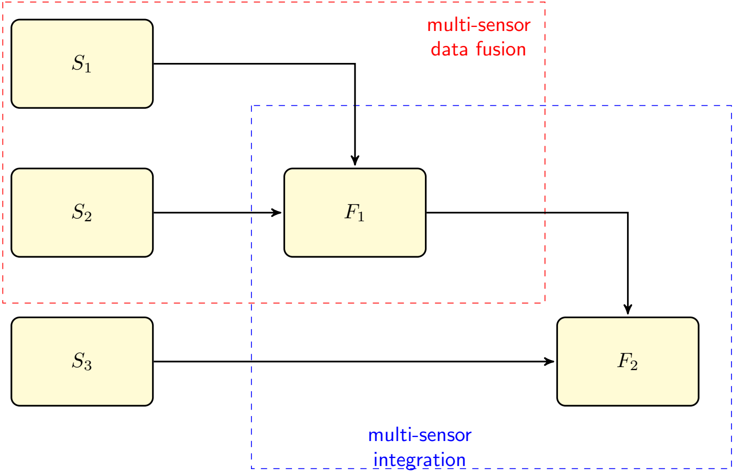
But beware of catastrophic fusion: if one sensor does not operate in an environment where it is designed to work then that sensor might dominate results of the fusion system and cause terrible results.

Domains of data fusion system



(adaptation from text book)

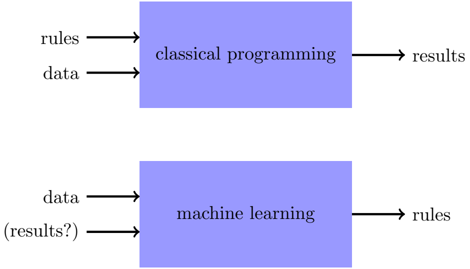
Multi-sensor integration



(adaptation from text book)

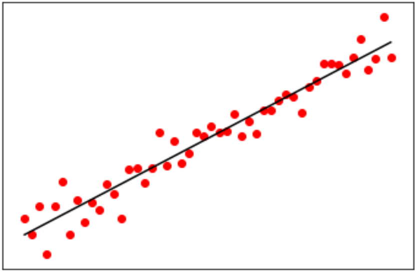
Machine learning

- subset of AI where computers learn from past data (almost) automatically
- uses statistical models which enable computer to make predictions based on data and auxiliary problem specific information
- ML system is trained with data, not programmed explicitly
- ML system creates rules that best (somehow) match data and results. Rules can be used to make predictions:



Supervised learning

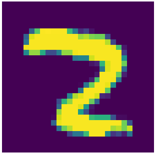
- "learning with a teacher"; system is trained with labeled data
- training = estimating parameters of the model
- system aims to predict future outcomes
- Regression (fit curve to data), predict continuous variable



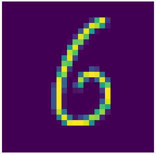
- regression line has two parameters; slope and y-intercept, those are estimatd from (training) data

Classification

- predict a value of categorical variable



(a) label=2



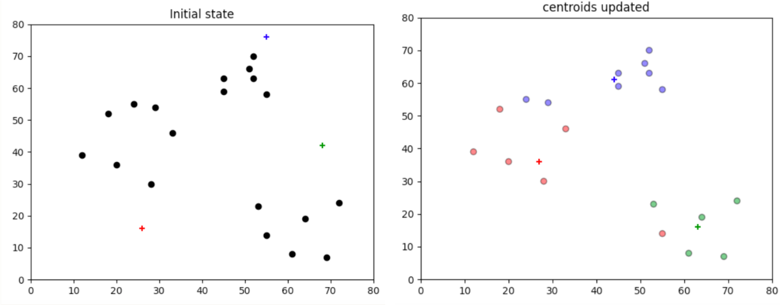
(b) label=6

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.3	3.7	1.5	0.2	setosa
5.0	3.3	1.4	0.2	setosa
5.1	2.5	3.0	1.1	versicolor
5.7	2.8	4.1	1.3	versicolor
6.2	3.4	5.4	2.3	virginica
5.9	3.0	5.1	1.8	virginica

- logistic regression, k nearest neighbors, support vector machine, decision tree, neural networks,...

Unsupervised learning

- "learning on your own"; no labels given
- identify hidden patterns
- clustering, association, anomaly detection,...



Probabilistic framework

We take a statistical approach: to describe relationships between sensors, sources of information are incorporated with inherent uncertainties.

Recall these basic concepts:

- Sample space S consists of all possible outcomes of an experiment.
- Any subset E of S is called an event.
- For any two events E and F we define
  - the union $E \cup F$ to consist of all outcomes that are in E or in F or in both.
  - The intersection $E \cap F$ consists of all outcomes that in E and in F.
  - The complement E^c of E consists of all outcomes (in sample space S) that are not in E.

Probability

Let $P(E)$ denote the probability of an event E. It satisfies the following axioms:

$$P(S) = 1$$

$$0 \leq P(E) \leq 1$$

$$P(\bigcup E_j) = \sum P(E_j)$$

if E_j mutually exclusive i.e. $E_i \cap E_j = \emptyset, i \neq j$
These imply:

$$P(E^c) = 1 - P(E)$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Conditional probability ("of E given F")

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \quad \text{if} P(F) > 0.$$

Here event F is called condition or evidence.

Bayesian approach

Bayes formula reads:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

Bayes rule is useful because in many cases it allows us to compute probabilities which cannot be found directly but it becomes straightforward if we know some other event has occurred (or has not).
Hence it allows one to incorporate new information to obtain an updated probability.

Random variables

A random variable is a quantity which is determined by the result of an experiment.

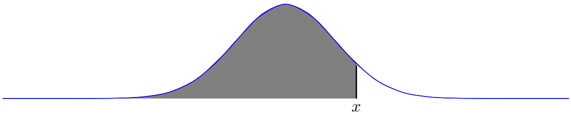
Random variables can be discrete or continuous.

Examples:

- sum of two (fair) dice; number between 2 and 12.
- value of the first die; number between 1 and 6.
- people's height in population

Cumulative distribution function (cdf) of random variable is

$$F(x) = P(X \leq x)$$



Expectation and variance

Expectation (mean, expected value) of a discrete random variable is defined as

$$E[X] = \sum_i x_i P(X = x_i).$$

For continuous random variable it is

$$E[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x,$$

where $f(x)$ is the probability density function (pdf). We denote also $\mu = E[X]$.

Variance is

$$\mathrm{Var}(X) = E[(X - \mu)^2] = E[X^2] - \mu^2$$

while standard deviation is $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ (same unit as X). It holds:

$$E[aX + b] = a\mu + b, \quad \mathrm{Var}(aX + b) = a^2\sigma^2$$

Covariance

Covariance of two random variables is

$$\sum(X, Y) = \mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

It describes the relationship of the random variables as follows: If $\mathrm{Cov}(X, Y)$ is positive (negative) then Y tends to increase (decrease) when X increases.
The strength of the relationship is measured by correlation (dimensionless coefficient between -1 and 1)

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Normal distribution (Gaussian)

We denote $X \sim \mathcal{N}(\mu, \sigma^2)$, when

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

It can be shown that

$$E[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2$$

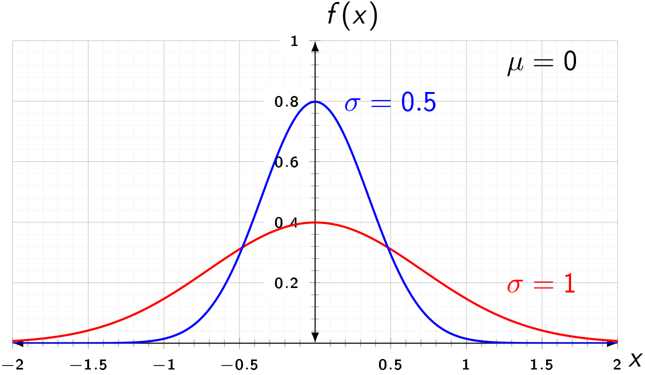
so that

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

(standard normal distribution).

Normal distribution (Gaussian)

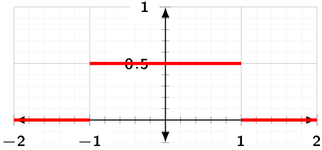
Expectation μ indicates the middle point of the graph.
Smaller σ means smaller variance or less spread (after square rooting).
Area under the curve always 1.



Uniform distribution

A random variable X is said to be uniformly distributed over the interval $[a, b]$ if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}$$



The mean and variance are

$$E(X) = (a+b)/2, \quad \mathrm{Var}(X) = (b-a)^2/12.$$

Value of uniform random variable over the interval $(0, 1)$ is called a random number.

Some linear algebra

Matrix A is invertible if $\det A \neq 0$. In such case inverse of A is A^{-1} such that

$$AA^{-1} = A^{-1}A = I$$

(recall the non-commutative matrix multiplication here).

Properties:

$(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}, \quad (AB)^{-1} = B^{-1}A^{-1}, \quad \det A^{-1} = (\det A)^{-1}$

Eigenvalue λ and eigenvector $v \neq 0$ of given matrix A satisfy

$$Av = \lambda v.$$