A BAYESIAN APPROACH FOR THE MAINTENANCE ACTION RECOMMENDATION 2013 PHM DATA CHALLENGE

PHM 2013 – 5th Annual Conference of the PHM Society

Vassilis Katsouros, Vassilis Papavassiliou and Christos Emmanouilidis

ATHENA Research & Innovation Centre, Greece
www.athena-innovation.gr
www.ceti.athena-innovation.gr/compsys
e-mail: christosem AT ieee.org

Outline

- Problem definition
- Modeling
- Results
- Discussion – conclusion

ATHENA Research & Innovation Centre

Athena Research Center

Research and Innovation Center in Information, Communication and Knowledge Technologies

Applied research, prototypes and applications development	Dissemination and R&D results valorization
Innovation support	Training and knowledge transfer

Research

ATHENA Research & Innovation Centre

Industrial Systems Institute

- Industrial Information & Communication Systems
- Embedded Systems
- Enterprise Integration
- Security & Privacy
- Industrial Automation & Control
- ICT in Manufacturing

Institute for the Management of Information Systems

- Large-scale information systems
- Data management
- Big data
- Geo-informatics
- Databases
- Software engineering
- Digital 'curation'
- Bio-informatics
- E-Government

Institute for Language and Speech Processing

- Natural & Embodied Language Processing
- Multimedia, 2D/3D imaging & VR
- Intelligent Systems
- Speech and Music Technology
- Technology-Enhanced Learning
- Cultural Informatics

Corallia Innovation Clusters

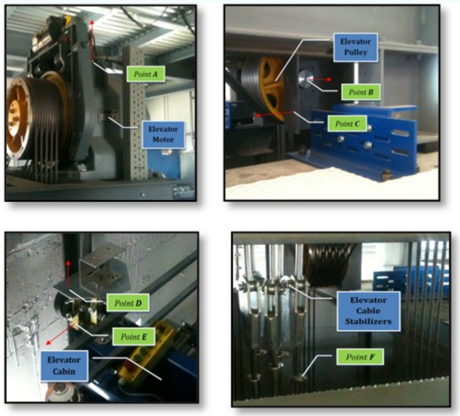
- Develop innovative ecosystems in specific sectors and regions of the country, and where a competitive advantage and export orientation exists
  - Nano/Microelectronics-based Systems and Applications (mi-Cluster)
  - Innovative Gaming Technologies and Creative Content (gi-Cluster)
  - Space Technologies and Applications (si-Cluster)
- Initiation of R&D Projects

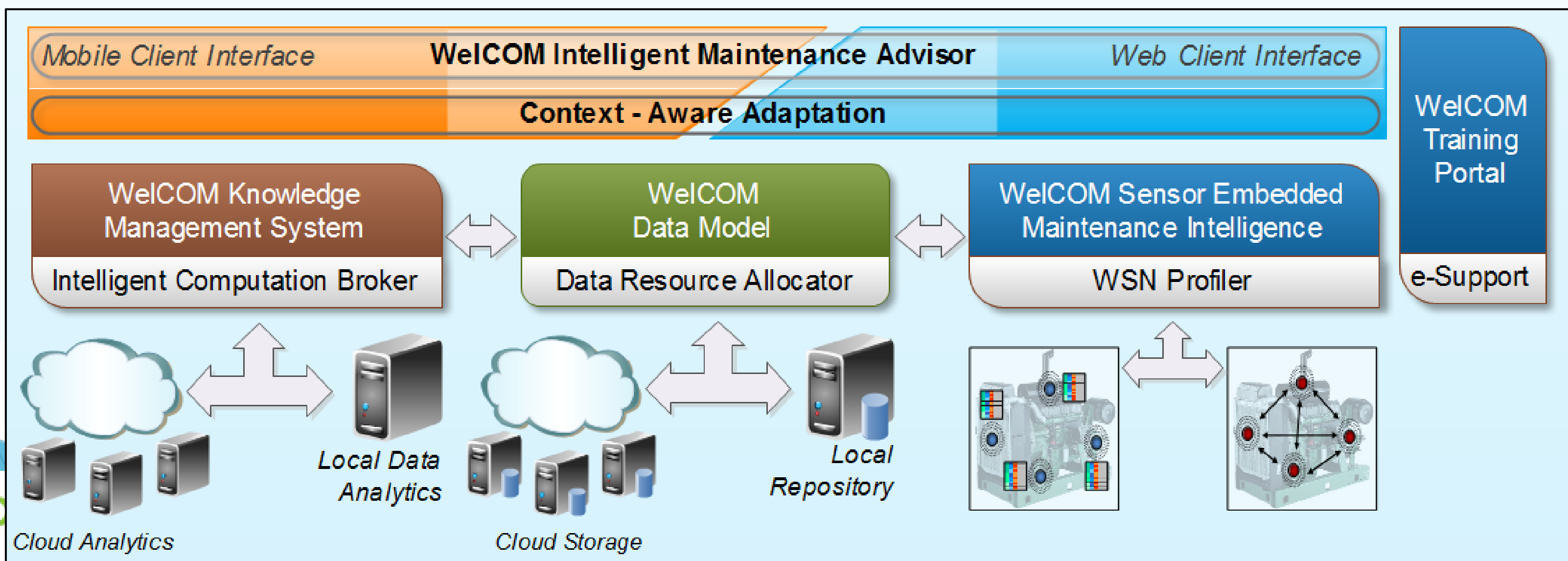
ρο Αθηνά
ch Center

Research Context: the WelCOM e-Maintenance Platform

welcom-project.ceti.gr

- Wireless sensor networks for engineering asset lifecycle optimal management





Our PHM 2013 Data Challenge Approach

- Worked with different modeling approaches

- Best results obtained with a Bayesian classification approach

  - define distinct problem classes one for each problem type
  - calculate the posterior probability of a test case for each problem class
  - recommend the problem class that corresponds to the maximum a posteriori (MAP) probability

Some characteristics of the PHM 2013 Data Challenge 1/3

- **case**: consists of a collection of **event codes**, each of which corresponds to a number of **parameters**
- **a record** of a case can be defined as a single event code along with the respective measurement parameters
- **30 parameters** of onboard measurements, **recorded each time an event code was generated**
- 1 dataset of cases with their **event codes** and **respective parameters** for training and another 1 for testing/evaluation
- The training dataset included the classification of the cases into nuisance or problem
  - for the cases classified as problems, the corresponding problem label/identifier was also provided.

Some characteristics of the PHM 2013 Data Challenge 2/3

- **Training dataset:** 1.316.653 records that correspond to 10.459 cases of which 10.295 were characterized as nuisance and 164 as problem (13 distinct problem IDs)
- **Testing dataset:** 1.893.882 records of event codes – measured parameters that correspond to 9.358 distinct cases.
- **Ground truth** of the testing dataset involved 174 problem cases, with the remaining 9.184 being nuisance cases
- A recommender should **identify the problem cases** from the 9.358 testing cases and for each one of them provide the respective **problem identifier**.
- **Evaluation metric**: calculated on a set of cases that involved all 174 ground truth problem cases and a random selection of 174 from the total of 9.184 nuisance cases.

Some characteristics of the PHM 2013 Data Challenge 3/3

- maximum performance: 348, i.e. sum of 174 ground truth problem cases and 174 nuisance
- that the training dataset includes one case with a problem type identifier (P7940) that is not found in the test dataset
- test dataset includes cases that have been categorized in two problem types (P0932 and P6880) for which there is no available training data

Problem ID	Training dataset	Test dataset
	Number of Cases	
P0159	19	15
P0898	4	6
P0932	–	2
P1737	2	2
P2584	53	26
P2651	13	13
P3600	17	20
P6559	3	1
P6880	–	15
P7547	6	4
P7695	17	37
P7940	1	–
P9766	14	12
P9965	2	5
P9975	13	16
Total	164	174

phmsociety

Posterior Probability of a Problem given a Case

$$\Pr[P_j \mid C_k] = \frac{\Pr[C_k \mid P_j] \cdot \Pr[P_j]}{\Pr[C_k]}$$

Bayes rule

phmsociety

Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Decision by selecting Maximum Posterior Probability

C_k

P_1

P_2

P_3

⋮

P_M

$$P_k = \arg\max_{j=1..M} \left\{ \log \Pr\left[C_k \mid P_j\right] + \log \Pr\left[P_j\right] \right\}$$

phmsociety

Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Conditional Probability of a Case given a Problem

$$\Pr\left[C_k \mid P_j\right] = \Pr\left[E_1^{C_k} E_2^{C_k} \ldots E_{N_k}^{C_k} \mid P_j\right]$$

$$= \Pr\left[E_1^{C_k} \mid P_j\right] \cdot \Pr\left[E_2^{C_k} \mid P_j\right] \ldots \Pr\left[E_{N_k}^{C_k} \mid P_j\right]$$

Assuming independence among events

Cases–Events representation

- Consider the caseIDs as columns of a matrix
- Consider the event IDs as rows of a matrix
- The elements e_{ij} of the matrix is the number of occurrences of event i being observed in case j

$$\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1K} \\ e_{21} & e_{22} & \cdots & e_{2K} \\ e_{31} & e_{32} & \cdots & e_{3K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{N1} & e_{N2} & \cdots & e_{NK} \end{bmatrix}$$

phmsociety
Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Cases–Problems representation

- Again case IDs are the columns of the matrix
- And the rows being the problem IDs
- With the elements 0 or 1

$$\begin{bmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix}$$

Conditional Probability of a Case given a Problem

C_k

P_1

P_2

P_3

$\vdots$

P_M

$$\log \Pr[C_k \mid P_j] = \sum_{i=1}^{N_k} \log \Pr\left[E_i^{C_k} \mid P_j\right]$$

phmsociety

Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Problems–Events representation

- Consider the event IDs as rows of a matrix
- Consider the problems as columns of a matrix
- The element n_{ij} of the matrix is the number of occurrences of event i is observed in a case that is classified in problem j

$$
\begin{bmatrix}
n_{11} & n_{12} & \cdots & n_{1M} \\
n_{21} & n_{22} & \cdots & n_{2M} \\
n_{31} & n_{32} & \cdots & n_{3M} \\
\vdots & \vdots & \ddots & \vdots \\
n_{N1} & n_{N2} & \cdots & n_{NM}
\end{bmatrix}
$$

phmsociety

Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Conditional probabilities Events given Problems

Add an ε term when $n_{ij} = 0$ to avoid zero probabilities

$$\Pr[E_i \mid P_j] = \frac{n_{ij}}{\sum_{i=1}^{N} n_{ij}}$$

	P_1	P_2	…	P_M
E_1	p_{11}	p_{12}	$\cdots$	p_{1M}
E_2	p_{21}	p_{22}	$\cdots$	p_{2M}
E_3	p_{31}	p_{32}	$\cdots$	p_{3M}
⋮	$\vdots$	$\vdots$	$\ddots$	$\vdots$
E_N	p_{N1}	p_{N2}	$\cdots$	p_{NM}

phmsociety

Ερευνητικό Κέντρο Αθηνά
Athena Research Center

Prior probabilities for Problems

$P_1 \quad P_2 \quad \ldots \quad P_M$

$$[p_1 \; p_2 \; \ldots \; p_M]$$

$$\Pr[P_j] = \frac{\sum_{i=1}^{N} n_{ij}}{\sum_{j=1}^{M} \sum_{i=1}^{N} n_{ij}}$$

Training Set – Correctly Classified Cases

Problem ID	Number of Cases	Top-1	Top-3	Top-5
P0159	19	15	16	18
P0898	4	4	4	4
P1737	2	1	1	1
P2584	53	49	53	53
P2651	13	13	13	13
P3600	17	15	17	17
P6559	3	1	2	2
P7547	6	2	5	5
P7695	17	15	17	17
P7940	1	1	1	1
P9766	14	14	14	14
P9965	2	2	2	2
P9975	13	10	11	12
Total	164	142	156	159
Overall (%)		86.59	95.12	96.95

Test Set – Correctly Classified Cases

Problem ID	Number of Cases	Top-1	Top-2	Top-3	Top-4	Top-5
P0159	15	1	2	4	7	10
P0898	6	0	1	1	5	6
P1737	2	0	0	0	0	0
P2584	26	12	19	21	24	25
P2651	13	6	7	7	7	11
P3600	20	9	15	18	19	19
P6559	1	0	0	0	0	0
P7547	4	1	1	1	1	1
P7695	37	24	30	32	35	37
P9766	12	5	5	5	8	9
P9965	5	0	0	0	1	1
P9975	16	2	3	4	7	7
Total	157	60	83	93	114	126
Overall (%)		38.22	52.87	59.24	72.61	80.25

Other techniques we tried...

- Gaussian Mixture Models (GMMs) for each problem ID trained on the case to events parameters
  - Diagonal covariance type
  - No. of Gaussian mixtures 10
  - Best score: 18
- Support Vector Machines (SVM) classifiers with features the case to events parameters
  - Radial Basis Kernel
  - Best score: 35

What could work better ?

- A hybrid technique comprising:

  (a) A recommender for problem type ID

  (b) A classifier for nuisance / problem

  - Tried the proposed (a) approach with an SVM-based (b) approach but the gain in nuisance rate detection was balanced by the loss in problem ID identification and there was no time left for further improvements

Conclusion

A Bayesian Approach for the Maintenance Action Recommendation 2013 PHM Data Challenge was proposed

A hybrid technique could achieve improved results

The definition of evaluation metrics for different PHM problems is a key issue

How about data challenges with multiple objectives ?

A BAYESIAN APPROACH FOR THE MAINTENANCE ACTION RECOMMENDATION 2013 PHM DATA CHALLENGE

PHM 2013 – 5th Annual Conference of the PHM Society

Vassilis Katsouros, Vassilis Papavassiliou and Christos Emmanouilidis

ATHENA Research & Innovation Centre, Greece
www.athena-innovation.gr
www.ceti.athena-innovation.gr/compsys
e-mail: christosem AT ieee.org

Ερευνητικό Κέντρο Αθηνά
Athena Research Center

phmsociety