Business Case Studies by GK

Gopalakrishna Palem

About the Speaker

- In a single line:
  - Management & Strategy consultant specialized in Big-data, IOT, Predictive Analytics and M2M Telematics.

About the Speaker (contd.)

- Research Scholar (IIT-Bombay)
  - Artificial Intelligence (AI)
  - Systems Theory, Simulation & Modeling
  - Game Programming & Visualization
  - Text Mining, Natural Language Processing (NLP)
- Product/Service/Executive Strategy (Microsoft, Oracle, Symphony Teleca)
  - Escalation Engineer for Distributed Services N.A. Division, Microsoft
    - Single-Point-of-Contact accountability for all North American Premier Accounts
    - Technology leadership and Solution Delivery management for 25+ frontline engineers
  - Passlogix Post-Acquisition Integration Manager, Oracle
    - Incubated the eSSO-IDC India Development Team
    - Cloud SSO integration for Oracle Access & Security Fusion Middleware
  - Big data Analytics Solution Strategy, Symphony Teleca
    - Laid out the M2M Big data Executive Vision and Strategy roadmap for Big-data Analytics COE.
    - Executive lead for the Big data Retail and Automotive Pre-Sales operations.

Publications

- Conference Proceedings
  - G. Palem. Data-dependencies and Learning in Artificial Systems. In *Proc. of Approaches and Applications of Inductive Programming*, page 69, 2005.
  - G. Palem. On Solving the System. In *Proc. of the International Conference on Systemic, Cybernetics and Informatics*, 2005.
- Book Chapters
  - G. Palem. Sequence Indexing for Game Development. In M. Dickheiser, editor, *Game Programming Gems 6*, pages 161–174. Charles River Media, 2006.
  - G. Palem. Creating Distributed Enterprise CRM System for Service Management. In T*he ITIL Experience*. Emereo Pty Ltd, 2010.
- Journal Articles
  - G. Palem. Condition-Based Maintenance using Sensor Arrays and Telematics. *International Journal of Mobile Network Communications & Telematics*, 3(3):19–28, June 2013.
  - G. Palem. Medicare Healthcare Charge Disparity Analysis. *Health Informatics International Journal*, 2(3), 9—15, 2013.
  - G. Palem, Formulating an Executive Strategy for Big data, *The Technology Innovation Management Review*, Mar 2014.
- Whitepapers
  - G. Palem. *M2M Telematics & Predictive Analytics*. Big-data Technical Report, Symphony Teleca Corp., 2013.
  - G. Palem. *The Practice of Predictive Analytics in Healthcare*. Big-data Technical Report, Symphony Teleca Corp., 2013.
  - G. Palem, Designing Condition-based Maintenance Management Systems for High-speed Fleet, 2014.

Technology Hands-on

- Areas of Expertise:
  - M2M, Big Data, NOSQL, Predictive Analytics, Scalable distributed Architectures, Machine Learning
- Cloud:
  - Azure, Hadoop, OpenStack, Storm, CouchDB, Pig, Hive, HBase, ZMQ, MQTT
- Programming:
  - R, C++, Java, VB.Net, HTML5, Qt, Python
- Security:
  - LDAP, AD, OID, IDM, SSO
- Visualization:
  - D3.js, Vtk, OpenGL, DirectX
- Designed Big data predictive analytics solutions for Market Uplift modeling and Social Media Sentiment-Analysis based on statistical Text-mining.
- Created real-time sensor data processing M2M Telematics IOT platform for Symphony.
- Delivered predictive maintenance Fleet management solution and realtime Epidemic Outbreak detection solutions based on Hadoop, Storm and HBase M2M IOT technologies.

Recent Big data Projects

Recent Projects

- Retail:
  - Social Media Sentiment Analysis
- Automotive:
  - Predictive Maintenance System for Heavy-Duty Machinery
- Healthcare
  - Food Borne Disease & Epidemic Outbreak Detection
- BFSI
  - Usage based Insurance System

Social Media Sentiment Analysis

- Customer: Large Retailer from US
- Business Problem:
  - Capture the market sentiment and monetize Social media trends
- Solution:
  - Big-data analytics with Text Mining
  - Large unstructured data streams from top social media firms were processed for trend, loyalty and intent
- Results:
  - Conversion rate increased by 4%
  - Product reach by 23%
  - Product sales by 8%

Predictive Maintenance for High-speed Fleet

- Customer: High-speed fleet owner from Europe
- Business Problem:
  - Reduce the maintenance costs and improve asset usage
- Solution Developed:
  - M2M Big data PaaS platform for predictive maintenance
  - Predictive maintenance schedules were planned based on equipment condition and RUL of assets
- Results Reported:
  - 32% decrease in maintenance costs
  - 75% reduction in downtime
  - 20% increase in productivity

Epidemic Outbreak Detection

- Customer: A Healthcare Organization, East Asia
- Business Problem: Needed a system to find causes for food-borne diseases and warn about potential epidemic outbreaks
- Solution devised:
  - Big data SaaS solution with Farm to Fork analysis
  - Pathogens observed in outbreak data were classified into Etiology groups and hierarchical clusters of regions were monitored for outbreak incidents
- Results:
  - 86% accuracy in predicting food-borne illness in realtime

Usage based Insurance System

- Customer: Large Financial Institution in US
- Business Problem: Needed a system to customize the insurance offerings
- Solution:
  - M2M Analytical system that monitored driver behavioral profile in real-time
  - Each driver gets computed a score in the range of 1 to 100 based on his fuel efficiency and energy savings that get linked to his insurance incentives
- Results:
  - The system eliminated fraudulent claims by 78%
  - Reducing settlement times and overhead costs by 28%.

High-volume transactional system

- Background
  - Large customer support vendor with cross-geographical presence
- Problem: Customer wanted to
  - Reduce maintenance costs on existing CRM
  - Build an in-house CRM system with follow-the-sun model
  - Minimize the lead-time during cross-geographical hand-offs
  - Improve the throughput and reduce transactional latency
- Solution: A distributed transactional system with Complex Event Processing (CEP) engine
  - Distributed data-storage for improved accessibility with auto-sync
  - SLA monitoring with work-flow enforcements
  - Rich UI with in-built collaboration features
- Value Add:
  - CSAT ratings increased by 4% due to IRT conformance
  - Improved collaboration features reduced the MPI, resulting in lower per-incident costs and improved hand-offs

Distributed Rendering Pipeline

- Background:
  - A fast-growing render-farm service provider in N.A.
- Problem: Customer wanted to
  - Offer rendering services in SaaS model
  - Setup a distributed render farm with commodity hardware and open-source tools
  - Offer competitive pricing with strict *availability* and *performance* SLA guidelines
  - Start with few nodes and scale as business grows
- Solution: A distributed rendering pipeline system with parallel job execution
  - Openstack based Dev-ops management and easy deployment solution
  - Map-Reduce based distributed pipeline for rendering
- Value Add:
  - Open-source based cost-effective solution provided competitive cost-advantage
  - Scalable node architecture enabled small clusters to grow
  - Virtualized solution enabled dynamic load-balancing and scaling for high-performance

Real-time data-flow modeling

- Background
  - A large CAE visualization and modeling product ISV
- Problem:
  - Existing product is decade old, at the brink of becoming obsolete
  - Unable to process large models - Occasionally large models took days to process
  - ISV Wanted to Redesign existing product with SaaS based distributed architecture
  - Real-time interactive modeling with high throughput is sought
- Solution:
  - Redesigned call-lists architecture with scene-graph API and K-d tree polygon subdivisions in real-time
  - Completely revised control-flow design with data-flow based dependency graph architecture
- Value Add:
  - Client-server visualization enabled large models be operated from thin clients
  - Scene graph implementation improved the response time and model loading time by more than 300x
  - Data-flow dependency graph architecture enabled the complete process to complete in hours as opposed to days

Data security for M2M Telematics Servers

- Background
  - Leading telecom M2M Telematics service provider with cross-geographical presence
- Problem: Customer wanted to consolidate multi-domain telematics data center operations into single interoperable multi-tenant model
- Solution: A federated-identity SSO security model with ReST-based credential management system
  - Claims based authentication with federated identities enabled interoperation between multiple existing back-end id-provisioning servers as LDAP, AD and OID
- Value Add:
  - SSO implementation reduced the id-provisioning management costs by 13% in the first year alone
  - Interoperable solution enabled the operator to consolidate and add diverging solutions with low cost of training and maintenance

PERSONAL Info

Education

University/College	Degree	Period	Marks	Major
Nagarjuna University - AP	B.Tech (CSE)	1997-2001	Distinction with 80.01%	Computer Science & Engineering
Board of Intermediate Education - AP	10+2	1995-1997	86.10%	Maths, Physics, Chemistry
Board of Secondary Education - AP	10th	1994-1995	85.50%	-

Open source contributions

- Creator of
  - CFugue Runtime
  - CarMusTy Typesetting System
  - CVMaker Latex Package
  - PhTranslator Library
  - GePhi exporter for R
  - RZMQ for Windows
- Contributions to
  - Qt
  - Vtk
  - OpenSceneGraph
  - ALSA driver for JDKMidi

Telematics & Predictive Analytics

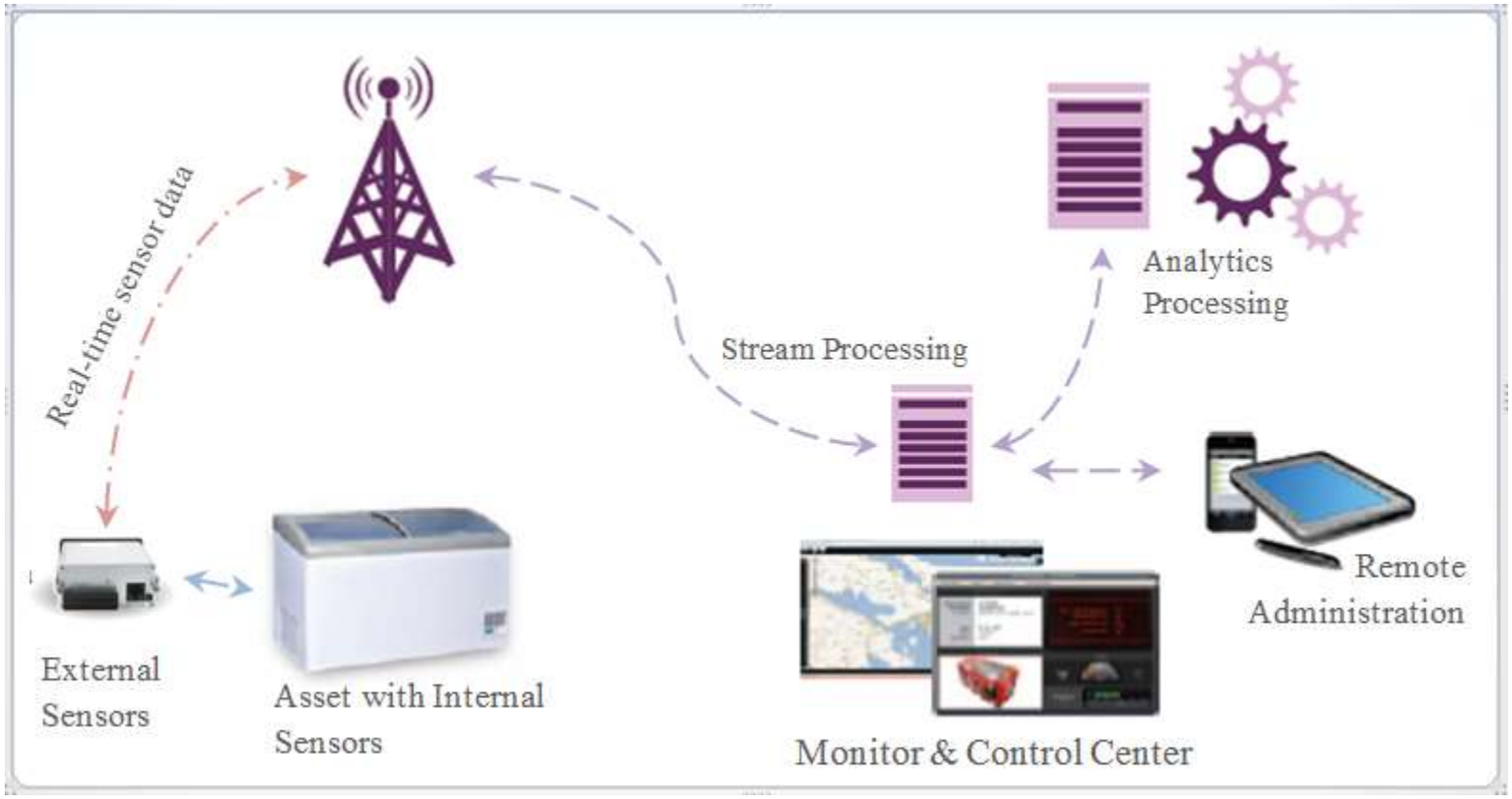
IP Opportunities

Machine-to-Machine

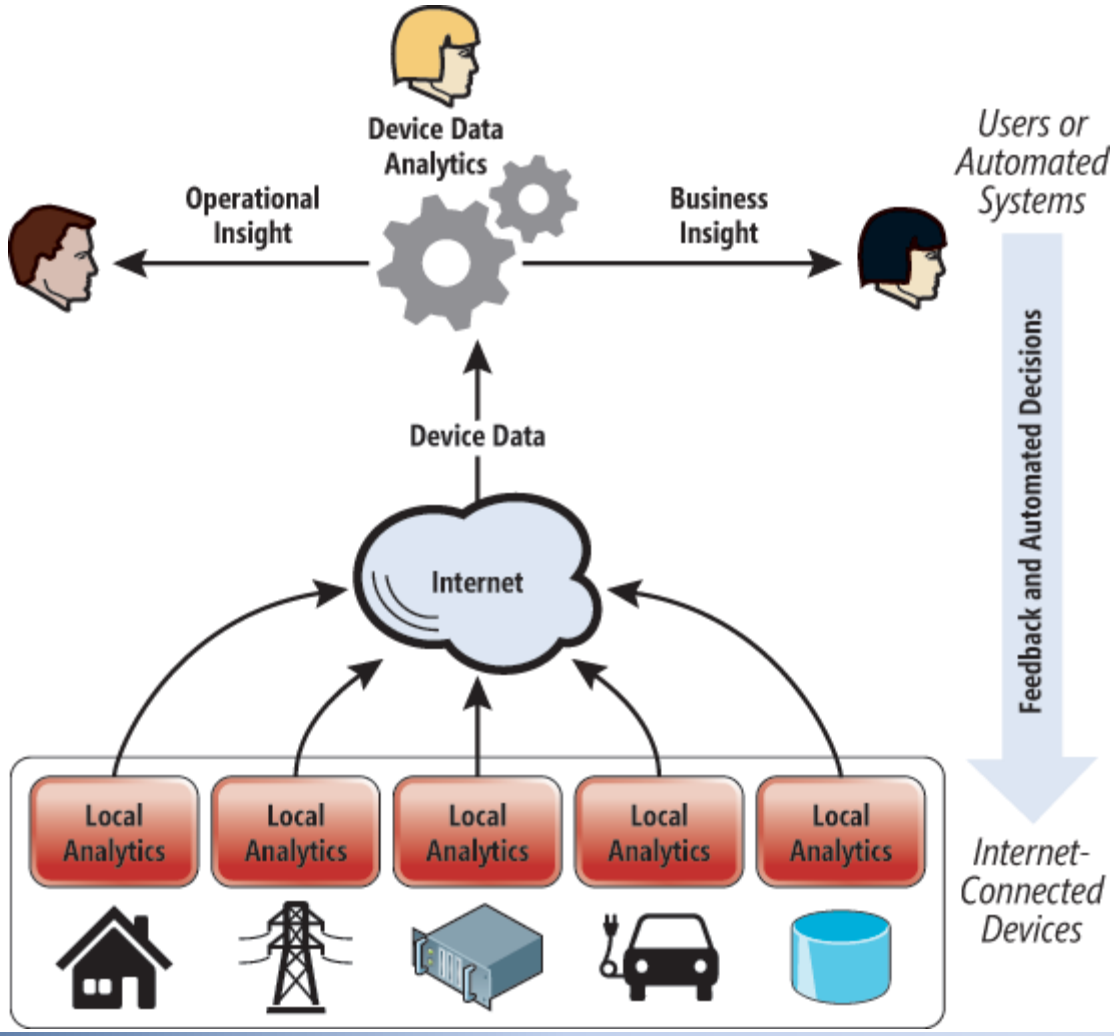
- M2M = connectivity
- Telemetry = remote measurement
- Telematics = remote control
- Predictive analytics = decision assistance
  - Helps decide control what, when…

- M2M+Telematics+Predictive Analytics ==
  - Smart homes
  - Smart cities
  - Smart “things”

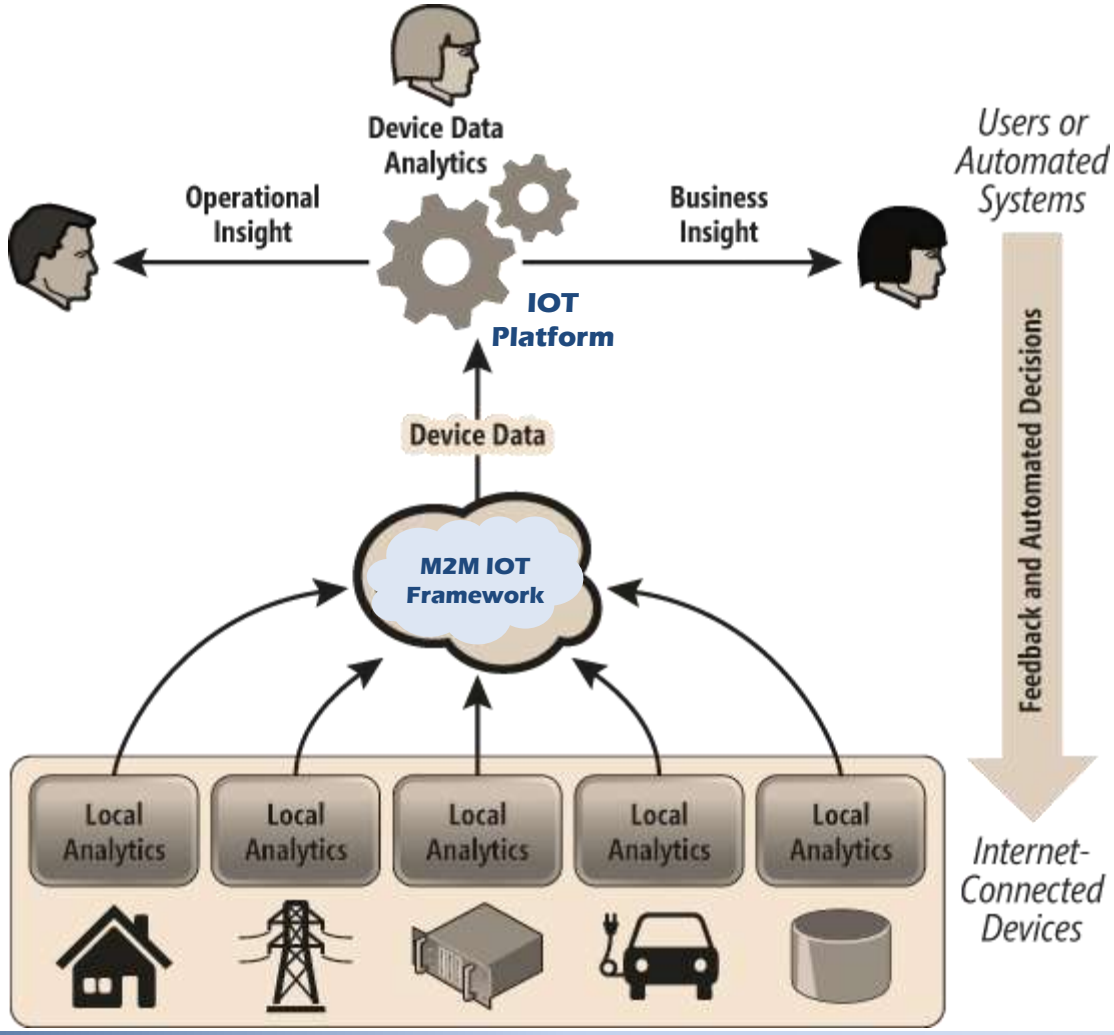
System Components



Internet of Things (IOT)



Internet of Things (IOT)



M2M Telematics Framework

Demo

Solutions Landscape

1. **End-user** (Individual) Level Solutions
   - Goal: Safety, Convenience , Cost reduction
   - Example: Driver behavioral profiling
2. **Business Management** Level Solutions
   - Goal: Cost reduction, waste elimination, increased Asset usage
   - Example: Fleet management
3. **Public** (Society) Level Solutions
   - Goal: Demand prediction, Increased resource availability
   - Example: Traffic management

Solutions Landscape Examples

- Traffic management
  - Peak hour traffic prediction and alternate routing
  - Demand prediction and schedule enhancements
    - If introducing a new service how many will use it
    - Using secondary data (social data, past travel history) to estimate how many people will probably use a route or service for special events
  - Emergency alternate routing, schedule delay updates
- Fleet management
  - Optimal workforce schedules, operational costs
  - Optimize maintenance costs
    - Component lifetime prediction
    - Inventory management for spare parts (ordering parts in time)
    - Workforce management (having the workmen scheduled readily)

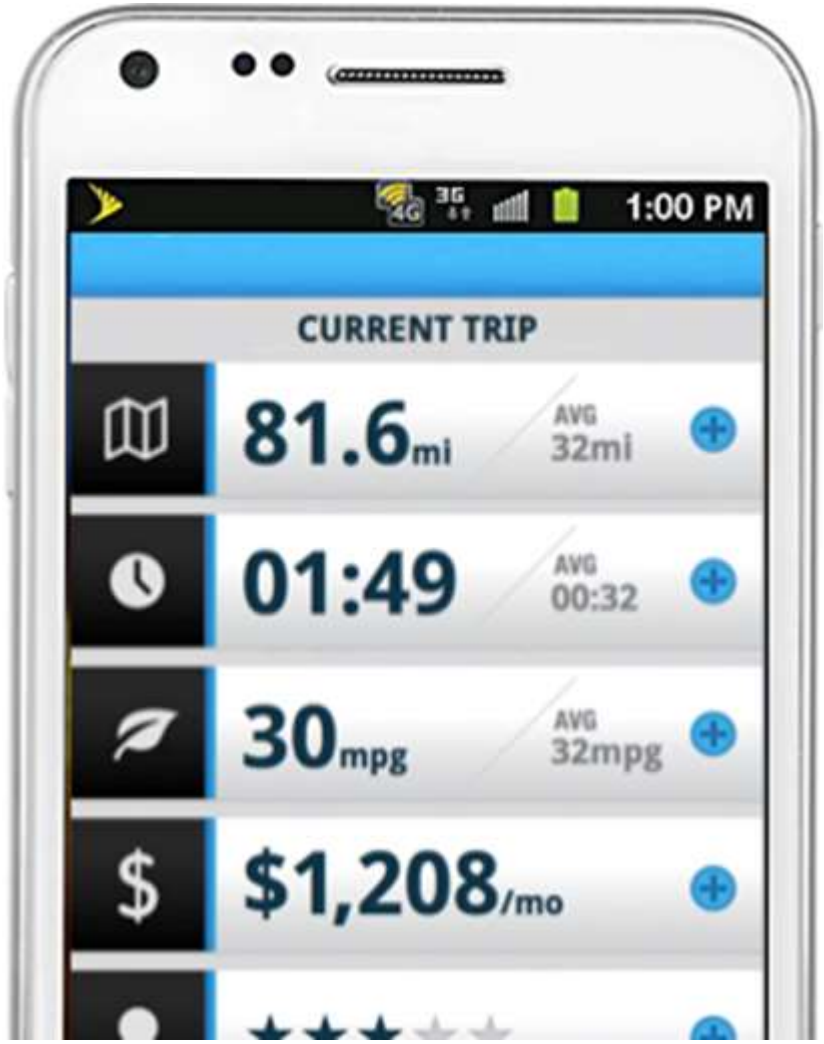
Example: Usage based Insurance

Driver Behavioral Profiling

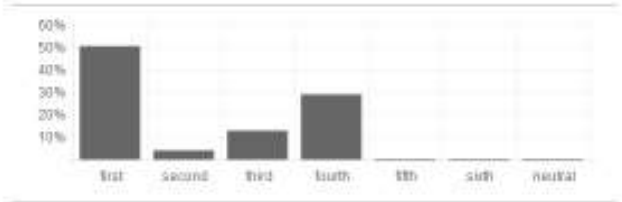
Driver Behavioral Profiling

- Drivers real-time performance calculated on the fly
- Goal is to arrive at a safety raking for the driver
- Sub-goal is to measure the fuel efficiency, energy savings of driver and give incentives
- During a trip driver gets measured on:
  - Acceleration patterns
  - Breaking patterns
  - Speed patterns
  - Idle timings …
- Gets assigned a score in the range of 0 to 100

Driver Behavioral Profiling (contd.)



Driver Behavioral Profiling (contd.)



Gear Position

Vehicle Speed

Engine Speed

Torque

Accel. Pedal (%)

Fuel Consumed

Odometer

Fuel Efficiency

Driver Behavioral Profiling

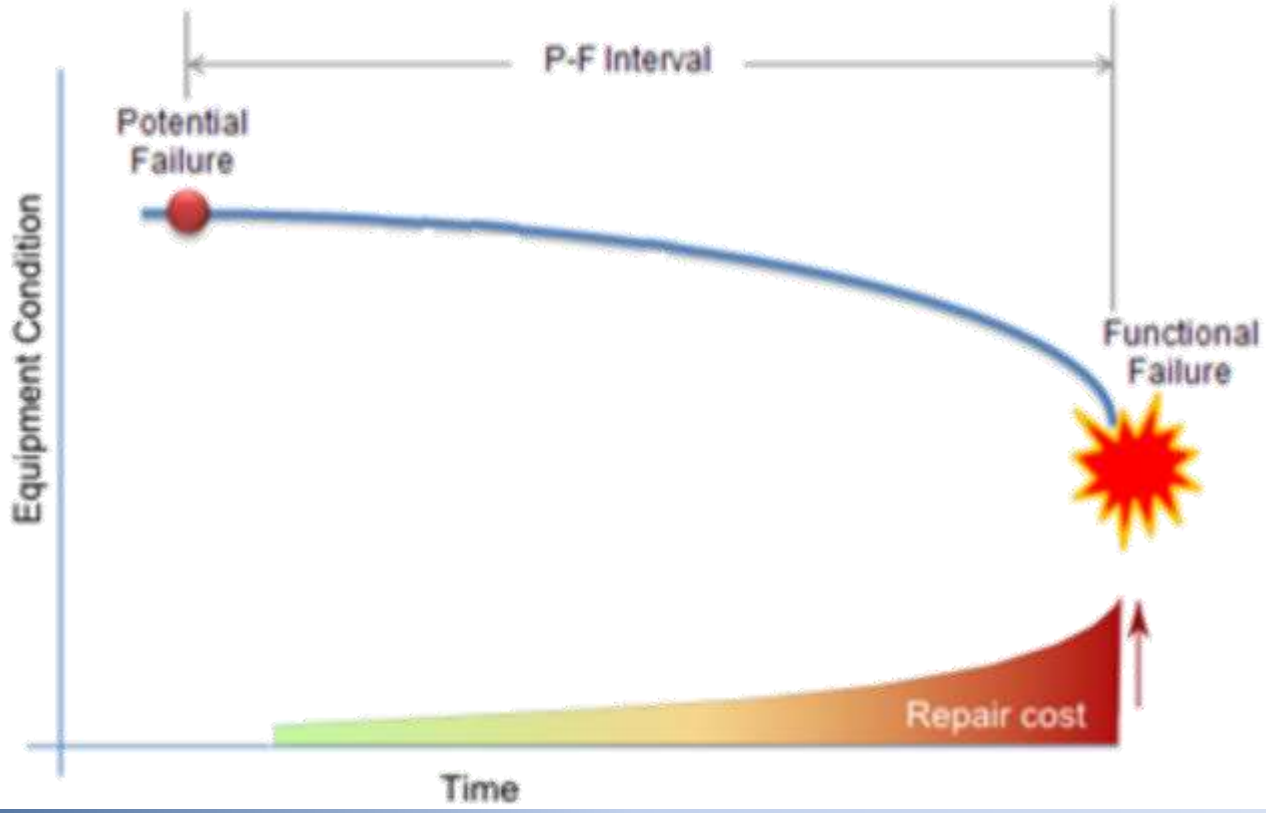
DEMO

Example: Fleet management

Condition-based Maintenance

Condition-based Maintenance

- Detecting failures in the early stages and preventing them

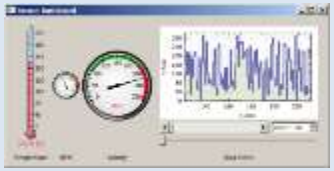


The Goal

- Estimating the *Failure Rate* for assets
  - Find the ‘Remaining Useful Life’ of assets
  - Schedule ‘Predictive Maintenance’
  - Maintain right levels of ‘Inventory’ of spare parts
  - Schedule right skilled and sized ‘workforce’
  - Optimize ‘Inspection’ routines
  - Evaluate ‘What If’ alternate scenarios
  - Decide right ‘Warranty’ period at design time
  - Compare different designs for reliability goals

CBM System Overview

Requirement Specifications

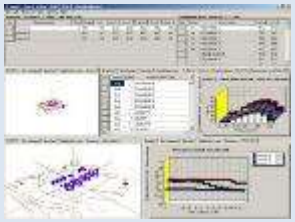


Real-time Monitoring

Real-time Alerts

OTMR / Sensors

M2M Framework



Expert system

Control Center

Alert reaction

- Design Specs
- Historical Data
- Experts Opinions
- Journals, Articles

Domain Knowledge

Early failure Signals

CBM System Overview

Requirement Specifications

Real-time Monitoring

Real-time Alerts

STAGE 2

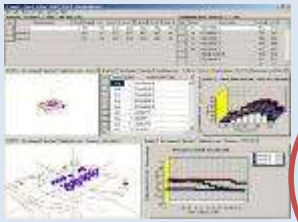
OTMR / Sensors

STAGE 1

M2M Framework

Control Center

Alert reaction



Expert system

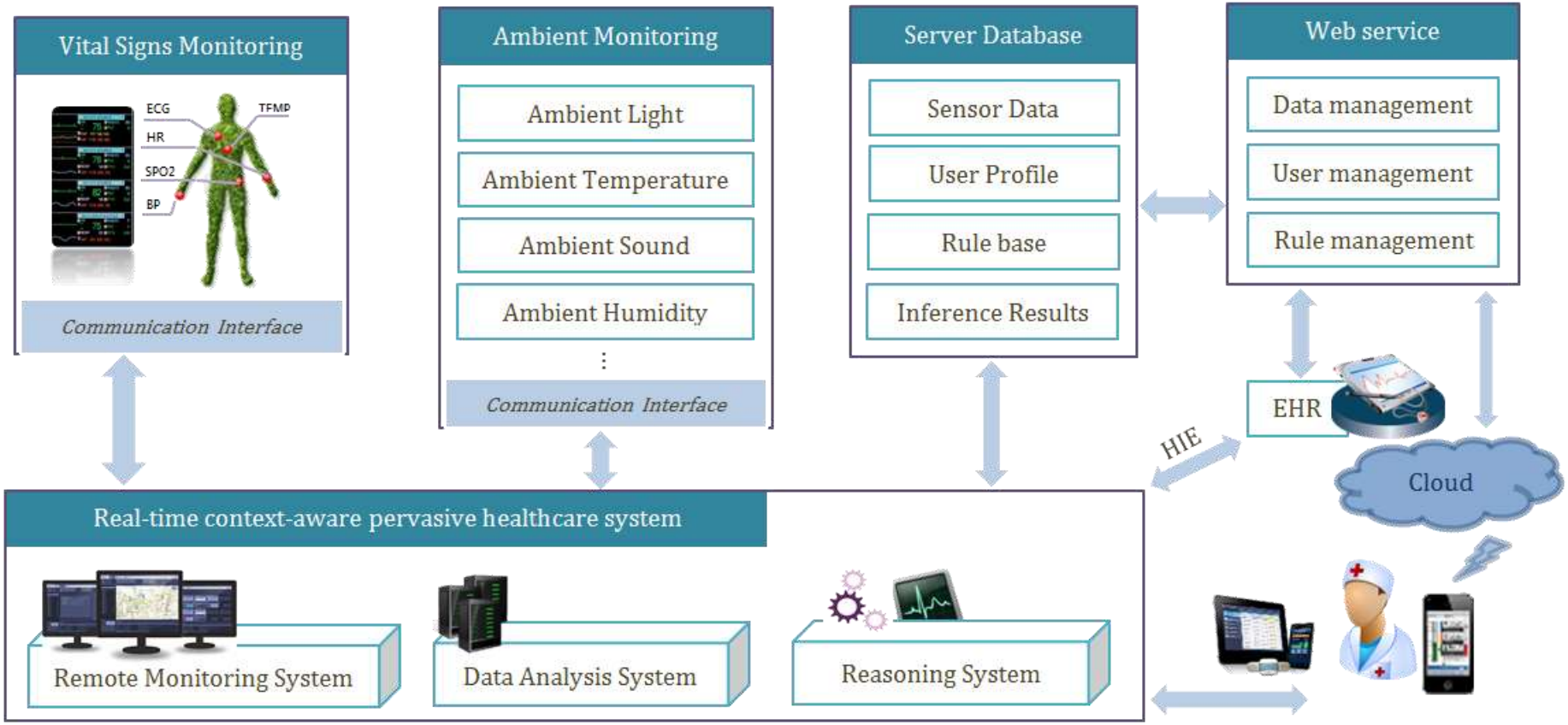
STAGE 3

rly failure Signals

- Design Specs
- Historical Data
- Experts Opinions
- Journals, Articles

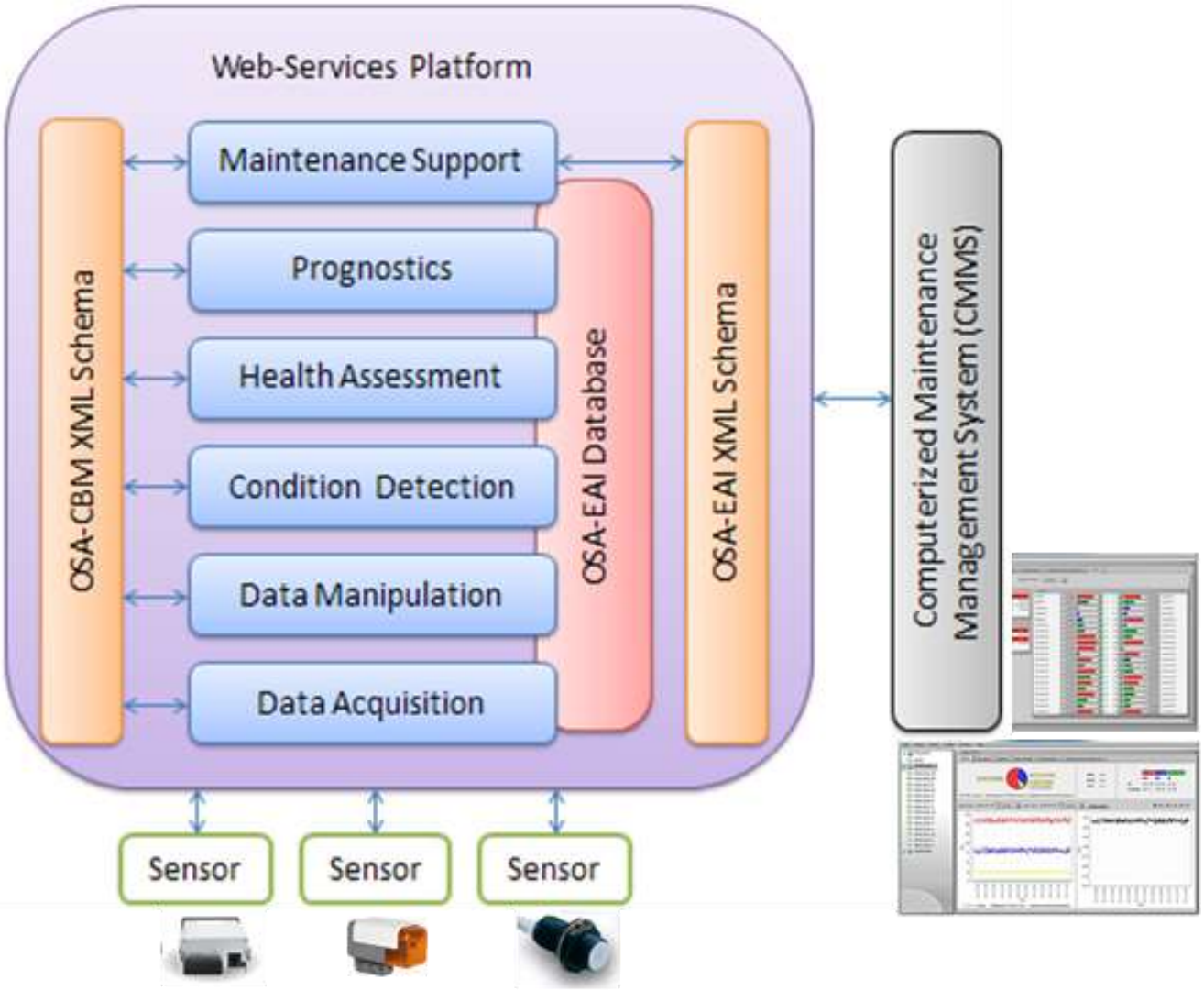
Domain Knowledge

CBM System Overview (Healthcare)

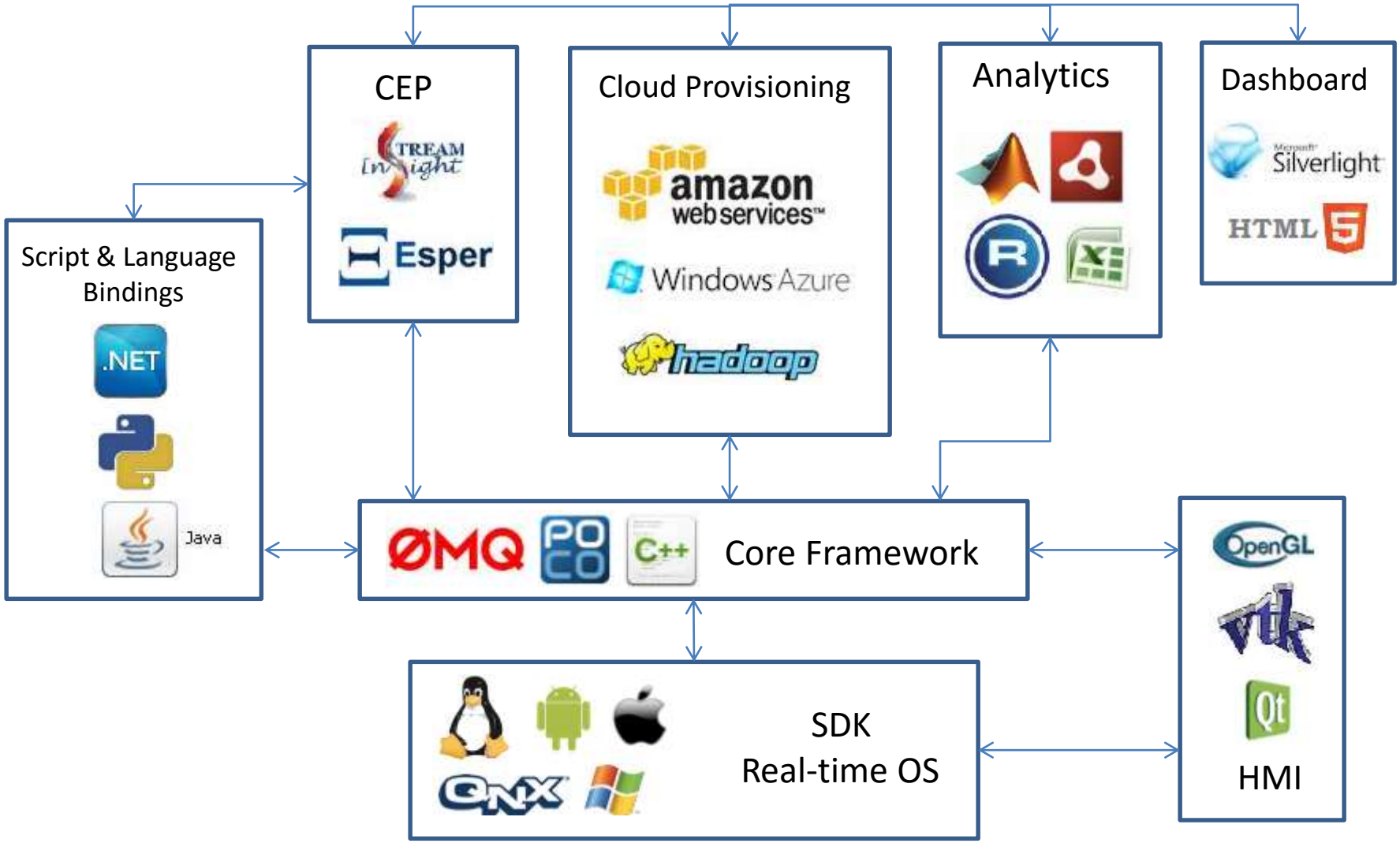


Source: *The Practice of Predictive Analytics in Healthcare*, Gopalakrishna Palem, 2013

CBMMS Architecture



Technology Stack



Real-time Failure Monitoring & Alerts

Demo

Failure Prediction from Historic Data

- Field failure of an asset occurred at times: 8204, 14688, 16460, 16578, 19479 and 20373 hours
- With an unfailed population of 1500:
  - Is the demonstrated B10 Life ≥ 10000 hours?
  - If not:
    - How many failures will occur by 5000 hours? 10000 hours? (with no inspection)
    - How many failures will occur by 10000 hours if we initiated a 5000 hour inspection?
    - With a utilization rate of 25 hours per month, how many failures can be expected in the next year?

Failure Prediction from Historic Data

Demo

Failure Prediction from Historic Data (Answers)

- Is the demonstrated B10 Life ≥ 10000 hours?
  - Answer: No, the B10 Life demonstrated is ~9222 hours
- How many failures will occur by 5000 hours? 10000 hours? (with no inspection)
  - Answer: (Assuming failed units are not replaced)

    No. of units failed by 5000 hours = failure prob. by 5000 hours * population = 0.01278 * 1500 = 19

    No. of units failed by 10000 hours = 0.12993 * 1500 = 195
- How many failures will occur by 10000 hours if we initiated a 5000 hour inspection?
  - Answer: Each unit will be cycled through 2 times, so, by 10000 hours (0.01278+ 0.01278)*1500 =38 failures expected

Failure Prediction (Answers)

- With a utilization rate of 25 hours per month, how many failures can be expected in the next year?

Failure Prediction for Next Year

No. of Units (n)	Current time on each unit (t)	Time on each unit at the end of year (t+hpy)	F(t)	F(t+hpy)	Single unit risk: (F(t+hpy)-F(t))/(1-F(t))	Total risk: n*(F(t+hpy)-F(t))/(1-F(t))
350	5500.00	**5800.00**	**0.01769**	**0.02119**	**0.00357**	**1.24818**
450	7800.00	**8100.00**	**0.05755**	**0.06525**	**0.00817**	**3.67757**
200	9800.00	**10100.00**	**0.12177**	**0.13413**	**0.01407**	**2.81454**
500	12000.00	**12300.00**	**0.22924**	**0.24681**	**0.02279**	**11.39450**
1500						**19.1348**

Shape Parameter	3.435511
Characteristic Life	17,753.48
Hours per Year	300

Topics not covered

- Considerations for Censored data
- Calculating system failure from individual components
- Identifying the ‘failure modes’

Example: Sentiment Analysis of Social Media with Text Mining

Natural Language Processing

Text Analytics

- Applying Text-analytics for polarity detection
  - Sources: Blog posts, Social media, Reviews

- Classifying the *polarity* at the document, sentence, or feature/aspect level — whether the expressed opinion is positive, negative, or neutral

- *Beyond polarity* sentiment classification looks, for instance, at emotional states such as *angry sad, happy* etc..

Text Analytics (contd.)

- Goal is not to validate each sentence for correct rating
  - Approx. human raters typically agree only at 79%
- Polarity is highly biased by algorithms and training sets
  - What is the right training set size?
  - Are the samples right?
- Polarity at an instant is also not that useful
  - Gives information, but not action-plan
- Goal is to get the 'rate of change of polarity'
  - Trend Analysis
    - Action plan is given by identifying the controller events that triggered the 'change'
  - Not feasible/useful to predict the trend
    - Influence of external events (unrelated news), product life-cycles
  - But gives insight for predicting the demand
    - How many IPhone5 should I manufacture next month to meet the demand?

Approaches

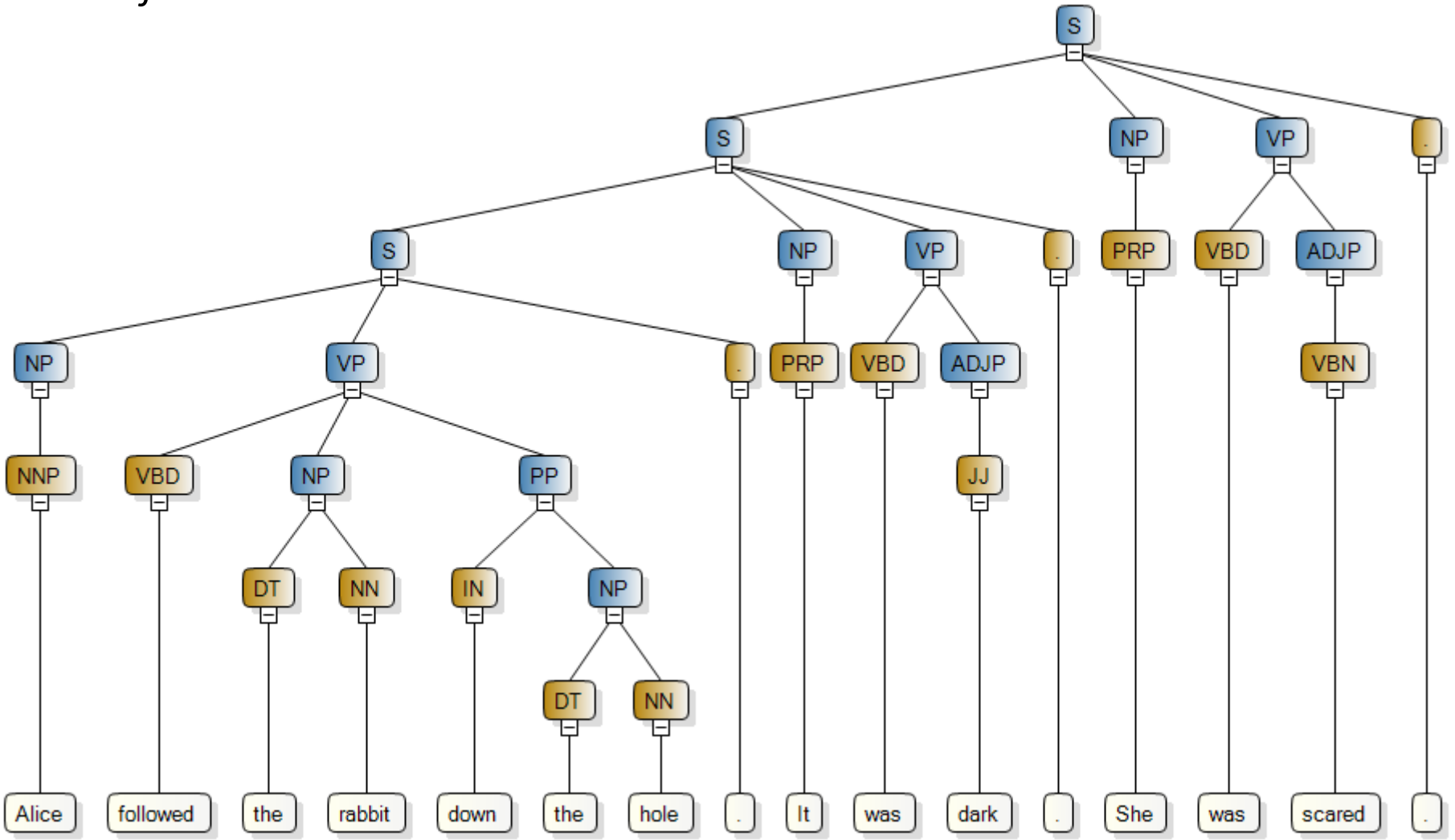
1. Bag of Words (BOW)
   - Syntactic
     - E.g.: Positive, negative lists
   - Ignores the relations among words and may miss enforcement adjectives
     - E.g.: Not that much happy
   - Fails to detect sarcasm, irony
     - E.g.: *As if its going to help*
2. Text Mining (TM)
   - Semantic
     - E.g.: Wordnet, Ontology
   - Represents complete parts of speech and summarizes the text
     - E.g.: Rama went to bank and saw the river. It was deep.
   - Slightly complex but highly accurate

Text Analytics: Semantics

Alice followed the rabbit down the hole. It was dark. She was scared.

Text Analytics: Semantics

Alice followed the rabbit down the hole. It was dark. She was scared.

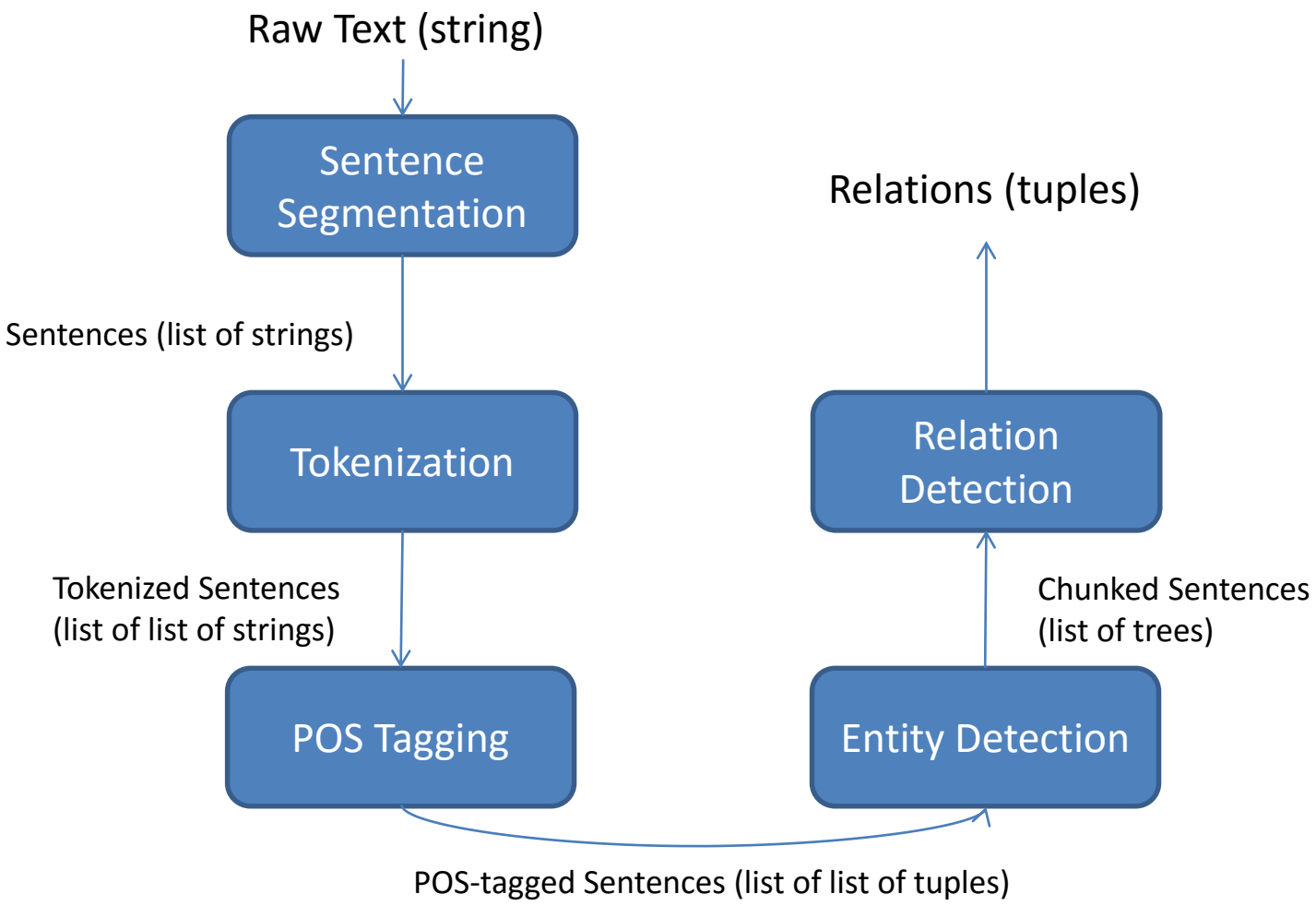


Text Analytics: POS-Tags

CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition/subord. conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or masps
NNP	Proper noun, singular
NNPS	Proper noun plural
NNS	Noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle

SYM	Symbol (mathematical or scientific)
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund/present participle
VBN	Verb, past participle
VBP	Verb, non-3rd ps. sing. present
VBZ	Verb,3rd ps. sing. present
WDT	wh-determiner
WP	wh-pronoun
WP$	Possessive wh-pronoun
WRB	wh-adverb
#	pound sign (currency marker)
$	dollar sign (currency marker)
"	close quote
(	open parenthesis
)	close parenthesis
,	comma
.	period
:	colon
``	open quote

Text Analytics: Information Extraction



TF-IDF

- Numeric indicator: Represents how important a word is to a document set or corpus
- Summarizes the content of document
- Useful for filtering *Stop-Words* (e.g. the, a, I ..)

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

TF-IDF (contd.)

- Term-Frequency: frequency of a term in a document (normalized to prevent bias towards longer documents)

$$TF(t,d) = \frac{f(t,d)}{\max\{f(w,d) : w \in d\}}$$

- Inverse Document Frequency: measure of rarity of a term across all documents

$$IDF(t,D) = \log \frac{|D|}{|\{d \in D \ : \ TF(t,d) \neq 0\}|}$$

- IDF a measure of whether the term is common or rare across all documents

Precision & Recall

- Precision: How many retrieved are relevant?
  - Algorithm returned more relevant than irrelevant

$$precision = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{retrieved\ documents\}|}$$

- Recall: How many relevant are retrieved?
  - Algorithm returned most of the relevant (if not all)

$$recall = \frac{|\{relevant\ documents\} \cap \{retrieved\ documents\}|}{|\{relevant\ documents\}|}$$

Topics not covered

- Sentiment assignment to sentences
- Weight calculations for adjectives and verbs
- Multi-language adaptation of sentiment
- Non-standard text processing (e.g. tweets)

Example: Environment safety with sensors

Behavioral Modeling for Fire

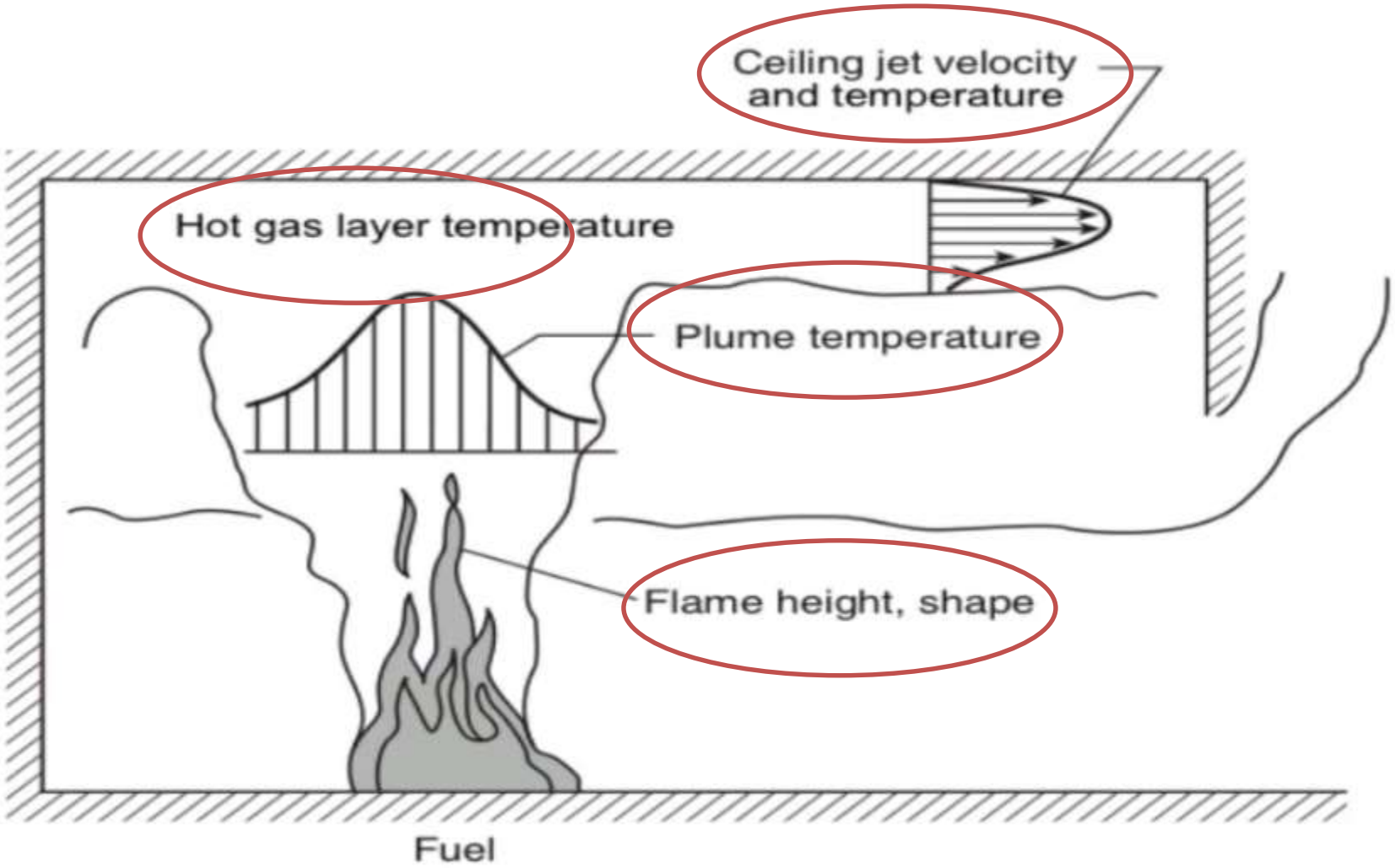
Fire behavioral modeling

- Hazard calculation by estimating heat transfer from flames
- An Age old problem with rich mathematical background
  - Cellular Automata
    - Alternative to current day Von Neumann Bottleneck architectures
  - Grid Propagation methods
    - Natively fits into Cellular Automata
    - Works with existing single-IP architectures through,
      - K-d tree multi-zone methods
      - CFD Voxels (Computational Fluid Dynamics)

Overview

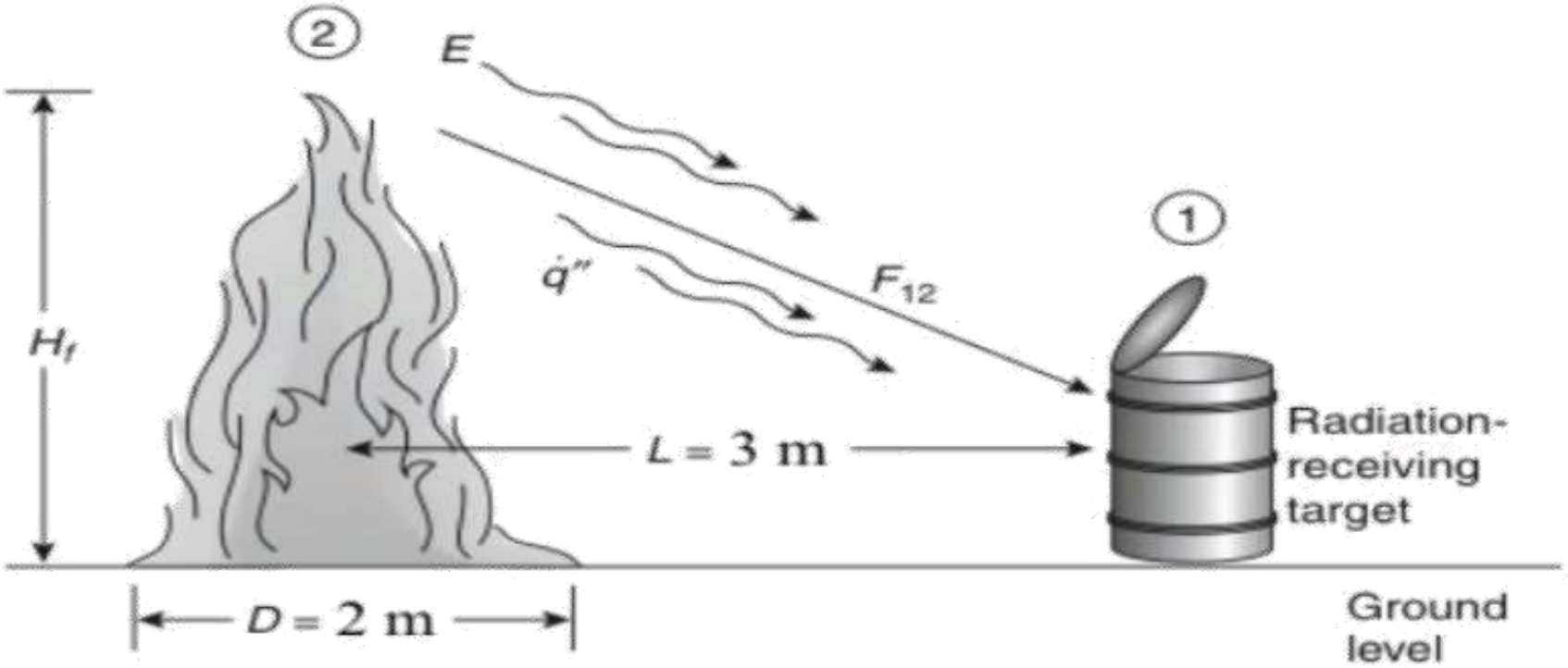
- Heat transfer from flames
  - Point-source, Shokri & Beyler, Cylindrical methods
  - Computation Fluid Dynamics (CFD)
- Fire propagation
  - Physically based modeling
  - Rothermel, Mc Alpine, Albini, McAurther … methods
- *Goal*: Calculate exposure temperature of targets
  - To know what can be saved and when to evacuate

Enclosure Fires



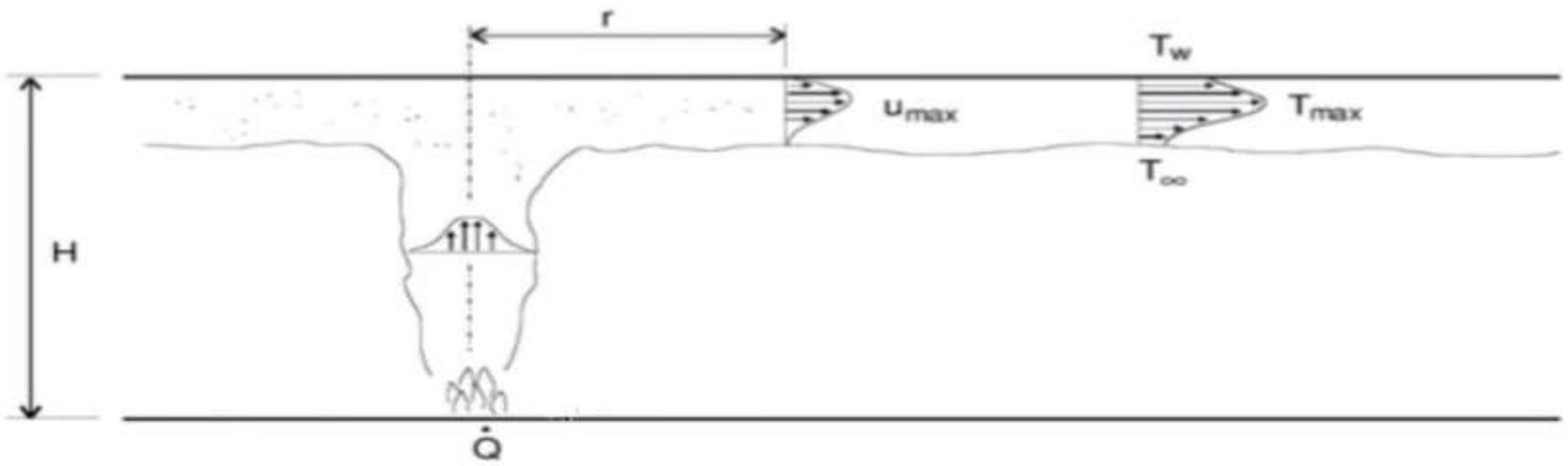
Radiation from Flame to Target

- A 500 KW gasoline pool fire with a diameter of 2m is burning 3m away from a barrel. Estimate the incident heat flux and temperature of barrel



Ceiling Jet Calculation

- Estimate expected gas temperature at a sprinkler that is located at 5m radially from a fire of 1200kw with a ceiling height of 3m and ambient temperature 20^0C



Topics not covered

- Estimating time for Smoke layer to reach certain height
- Calculating Mechanical Ventilation requirement to keep the smoke at certain height
- External Flame propagation behavior
- Effect and presence of carbon gases

References

- Brin, S.; Page, L. (1998). "The anatomy of a large-scale hypertextual Web search engine". *Computer Networks and ISDN Systems* **30**: 107–117
- Bing Liu (2010). "Sentiment Analysis and Subjectivity". Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010
- Gopalakrishna Palem (2013), M2M Telematics & Predictive Analytics
- Gopalakrishna Palem (2013), The Practice of Predictive Analytics in Healthcare
- SFPE Handbook of Fire Protection Engineering
- Gopalakrishna Palem (2014), Formulating an Executive Strategy for Big data
- Gopalakrishna Palem (2005). “Data-dependencies and Learning in Artificial Systems”, Proc. of Approaches and Applications of Inductive Programming, 69--78