UNIVERSITY OF THESSALY
1988 — 2018 creative years

iPRISM
Intelligent Pervasive Systems Research Group
Dept. of Informatics and Telecommunications

A Proactive Uncertainty Driven Model for Data Synopses Management in Pervasive Applications

Kostas Kolomvatsos

6th IEEE International Conference on Data Science and Systems (DSS)
December 14-16, 2020

OUTLINE

- Introduction
- Problem Description
- Fuzzy reasoning Process
- Time Series Forecasting
- Decision Making
- Experimental Evaluation
- Conclusions & Future Work

INTRODUCTION

Numerous Devices	The increased adoption of Internet of Things (IoT). The development of IoT application and usage IoT devices.
Huge Volumes of Data	IoT devices and applications produce or collect data. The data should have appropriate form to draw conclusions
Interaction with the Edge of the Network	The Edge Computing (EC) nodes are placed close to the data sources. Edge minimize the latency in the provision of responses. Edge Computing perform analytics over distributed data streams.
Synopses	Synopses represent the statistical information of distributed datasets. They are exchanged to depict the statistical information to the ecosystem of nodes.

CONTRIBUTION

- We argue for the dissemination of the synopsis of each dataset to have the insight on data present in peer nodes
- We provide a monitoring mechanism for detecting the magnitude of the updated synopses
- We deliver a forecasting scheme for estimating the future realizations of data synopses
- We describe and analyse an uncertainty driven model for detecting the appropriate time to distribute data synopses to peer nodes
- Uncertainty if managed through the use of the principles of Fuzzy Logic

PROBLEM DESCRIPTION

- We consider a set of edge nodes that are owners of distributed datasets
- Contextual data vectors are reported by IoT devices that capture them through interaction with their environment and users
- We consider the online knowledge extraction model as the statistical synopsis for each dimension of the multivariate data
- Synopses can be useful when we want to have a view on the collected data in a remote location
- Instead of migrating huge volumes of data, we can distribute the discussed synopses with a positive impact on the performance of the network
- Edge nodes try to act in a cooperative manner and decide to exchange their synopses regularly

PROBLEM DESCRIPTION

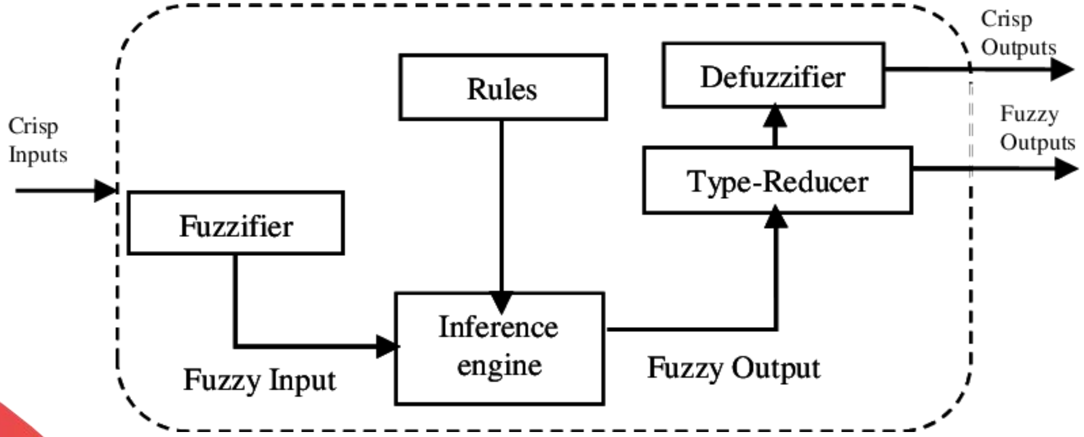
- We propose a dynamic approach for sharing synopses with peer nodes
- The idea is to let EC nodes to decide the 'magnitude' of the collected statistical synopsis before they decide a dissemination action
- Obviously, there is uncertainty around the amount of magnitude that should be realized before we fire a dissemination action
- For calculating the discussed magnitude, we rely on the error/difference between two consecutive synopses realizations
- We define the update quantum, i.e., the magnitude of the difference between the last calculated synopsis circulated to peer nodes and the current synopsis
- We monitor the update quanta during functioning while receiving more data from the IoT devices
- We apply forecasting techniques and Fuzzy Logic for solving the problem of finding the appropriate time for circulating synopses

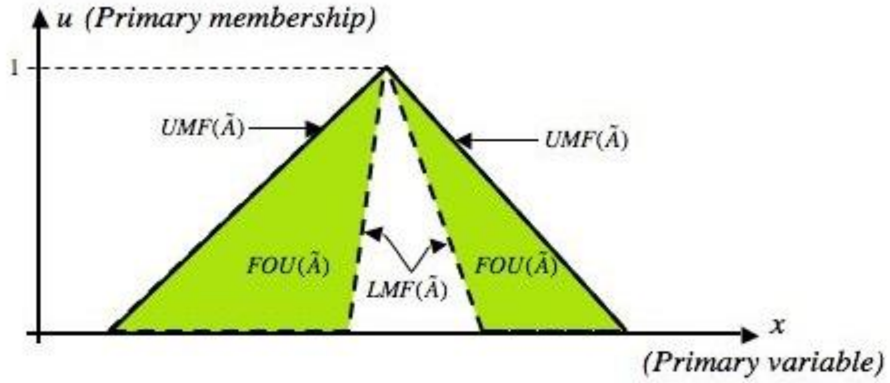
THE FUZY REASONING PROCESS

- We propose a Type-2 Fuzzy Logic System (T2FLS)
- The envisioned fusion of update quanta is achieved through a finite set of Fuzzy Inference Rules (FIRs)
- FIRs incorporate and 'combine' past quanta or future estimations (two different processes) to reflect the Potential of Distribution (PoD)
- The T2FLS is invoked two times, i.e., for delivering the PoD upon historical values and the PoD upon future estimations
- The T2FLS is invoked upon the last three (3) quanta realizations, and the future three (3) quanta estimations
- The outputs are PoD_p & PoD_f
- For the fuzzy sets, we characterize their values through the terms: low, medium, and high
- We incorporate 27 Fuzzy Rules for handling all the combinations between inputs

THE FUZY REASONING PROCESS

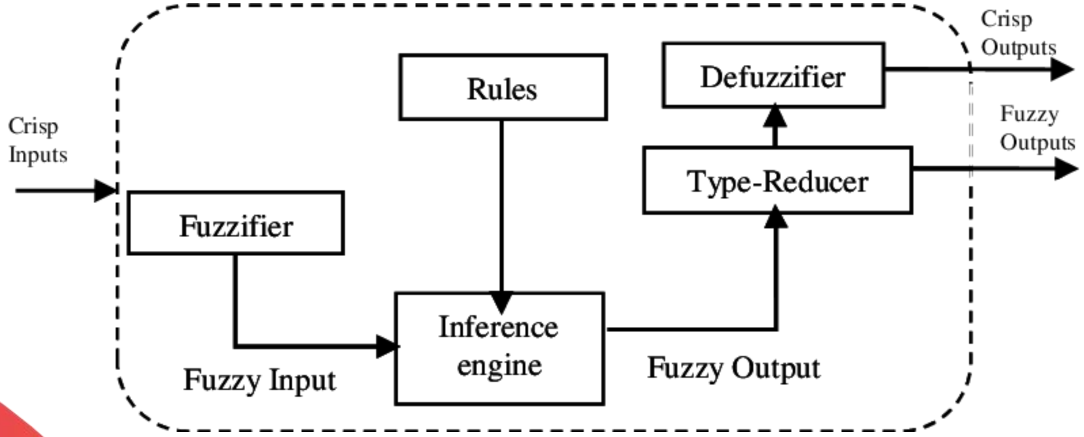
Type-2 Fuzzy Sets

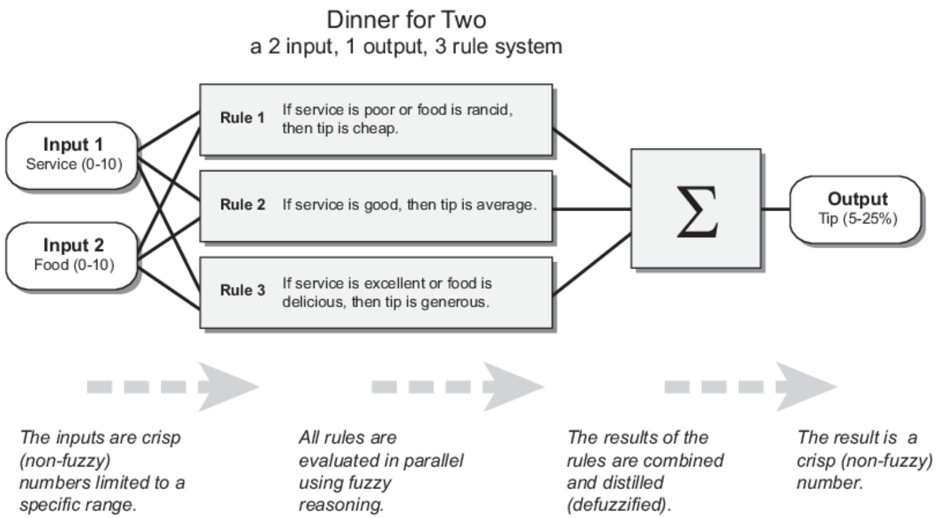




THE FUZY REASONING PROCESS

Type-2 Fuzzy Sets





TIME SERIES FORECASTING

- We adopt the double exponential smoothing for estimating the future three (3) update quanta e_j based on all the already available/calculated synopsis quanta

- The following equations hold true

$$v_j = \alpha \mathrm{e}_j + (1-\alpha)(v_{j-1} + b_{j-1})$$

$$b_j = \beta(v_j - v_{j-1}) + (1-\beta)b_{j-1}$$

$$v_j = \alpha \mathrm{e}_j + (1-\alpha)(v_{j-1} + b_{j-1})$$

$$b_j = \beta(v_j - v_{j-1}) + (1-\beta)b_{j-1}$$

$$v_1 = \mathrm{e}_1$$

$$b_1 = \mathrm{e}_2 - \mathrm{e}_1$$

- α: the data smoothing factor, β: the trend smoothing factor

DECISION MAKING

1^{st} Step

Invoke the T2FLS and get the PoD_p based on the paste three quanta

2^{nd} Step

Invoke the T2FLS and get the PoD_f based on the three quanta estimations

3^{rd} Step

Combine theT2FLS outcomes with the geometric mean and get the aggregated result

- The aggregated value is calculated as follows:

$$G(PoD_p, PoD_f) = \left(\prod_{i=1}^{2} PoD_i\right)^{1/2}$$

- When G is over a pre-defined threshold θ, the current synopses is circulated in the network

EXPERIMENTAL EVALUATION

- We perform a large set of simulations to evaluate the proposed Uncertainty Driven Dissemination Model (UDDM)

- We compare it with:
  - a baseline model (BM) that disseminates synopses when any change is observed over the available data
  - a Prediction based Model (PM) that proceeds with the stopping decision only when the estimation of the future update quanta violates a pre-defined threshold

- Performance metrics:
  - ϕ: the percentage of the available interval for disseminating the synopsis (compared to an upper deadline T)
  - δ: the average magnitude of the difference between the current and the new synopses
  - ψ: the ability of the proposed system to avoid the overloading of the network and limiting the required number of messages (how many times the model stops in an interval [1,T])

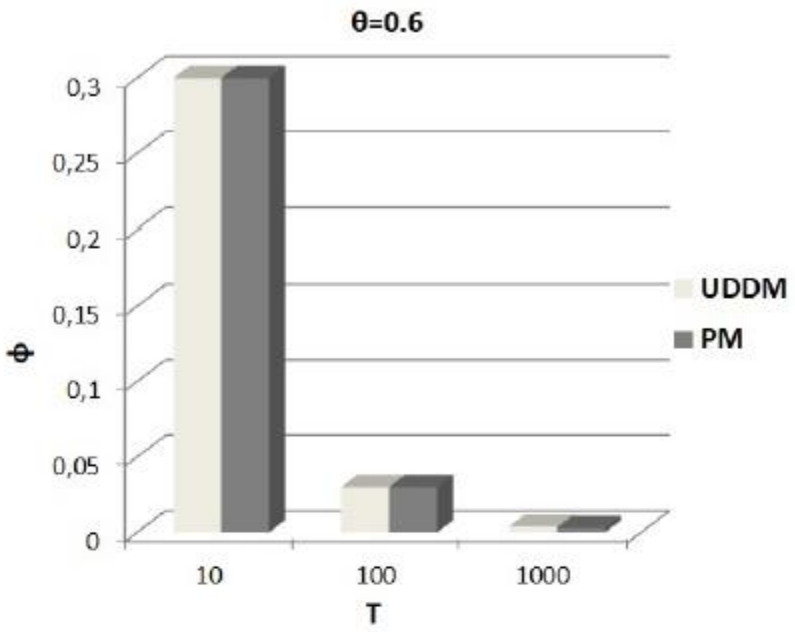
- Dataset: we adopt the dataset being related with measurements from 54 sensors deployed in the Intel Berkeley Research Lab (the reference is available in the paper)

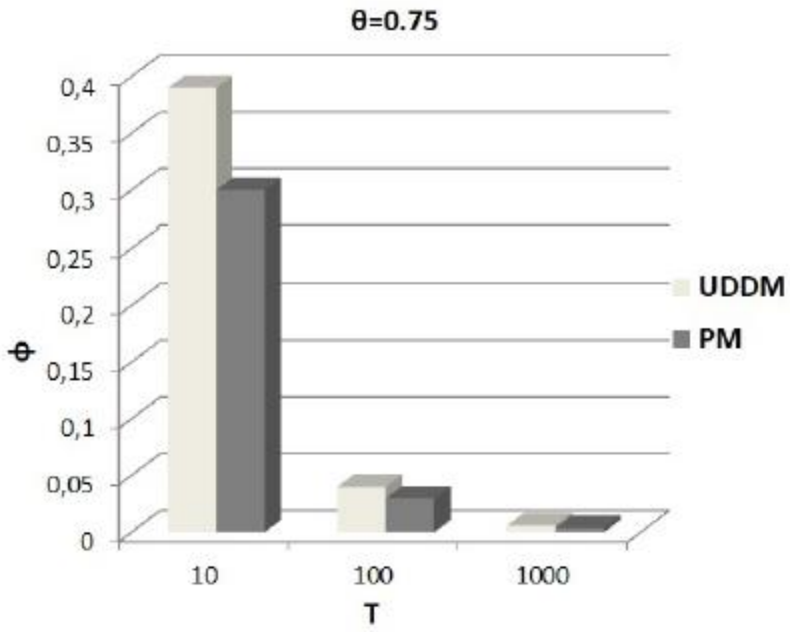
EXPERIMENTAL EVALUATION

- A low θ (threshold for deciding the dissemination action) and a low T (deadline to conclude the distribution of synopses) lead to an increased time for the final decision

- When θ and T are high, the percentage of T devoted to conclude the dissemination decision is very low

- The proposed system manages to deal with the final decision as soon as it detects that update quanta are aggregated over time even in small amounts





EXPERIMENTAL EVALUATION

- We observe that the UDDM requires a higher magnitude than the BM and the PM before it concludes the dissemination action (δ metric - left)
- In general, there is an increment in δ as T increases
- We observe that the UDDM demands for less dissemination messages compared with the BM & PM (ψ metric - right)
- The proposed model decides the dissemination of messages every 2.5 and 1.5 (approximately) monitoring rounds for θ=0.6 & θ=0.75, respectively

EXPERIMENTAL OUTCOMES FOR THE δ METRIC

	$\theta = 0.6$			$\theta = 0.75$		
T	**UDDM**	**BM**	**PM**	**UDDM**	**BM**	**PM**
10	16.42	13.87	13.87	15.82	12.05	8.61
100	20.95	17.76	17.02	17.60	15.59	13.92
1,000	19.62	17.68	16.34	20.55	17.79	16.63

EXPERIMENTAL OUTCOMES FOR THE ψ METRIC

	$\theta = 0.6$			$\theta = 0.75$		
T	**UDDM**	**BM**	**PM**	**UDDM**	**BM**	**PM**
10	2.5	1.42	1.66	1.66	1.66	1.42
100	2.5	1.2	1.13	1.47	1.29	1.23
1,000	2.09	1.2	1.12	1.46	1.24	1.15

CONCLUSIONS & FUTURE WORK

- ❖We propose an efficient model for decision making under uncertainty
- ❖We reason upon past update quanta and future estimations
- ❖Our aim is to provide a scheme that minimizes the number of messages circulated in the network, however, without jeopardizing the freshness of the exchanged statistical information
- ❖Nodes monitor their data and decide when it is the right time to deliver the current data synopsis
- ❖In the first place of our future research plans, it is to incorporate a deep machine learning model in the uncertainty management scheme

iPRISM

THANK YOU

Kostas Kolomvatsos
kostasks@uth.gr

http://www.iprism.eu