Data Fusion Reference Models: Basic Concept, Architecture and Comparison

[1]Shruti Sharma, [2]C.K.Nagpal

[1]Assistant Professor, [2]Professor
[1,2]Department of Computer Engineering ,
JCBOSEUST, YMCA, Faridabad.

Abstract : Data fusion is the process of amalgamating multiple data sets to produce accurate, and useful information than that provided by any individual data set. World Wide Web is the gigantic, unstructured and richest data source. With the technical advancements, there is a huge increase in the data set generated from the sensors. The amalgamation of data sets generated from multiple sensors can be used accurately to retrieve better information. Various data fusion models have been proposed over the years to fuse data from different sources. In this paper a survey of various data fusion models along with their architectures has been discussed that takes into account various important features like the objective for applying each model, by comparing their advantages and disadvantages.

Index Terms - Data fusion; JDL Model; Dasarthy's classification; Omnibus Model; Object Oriented Model;

I. INTRODUCTION

In the big data era, data have been growing very rapidly in unstructured and semi-structured format. Each dataset belongs to different domain and accessed from different sources. For example, log files are generated. Now considering different datasets equally and applying conventional data fusion methods on them will not be appropriate. An advanced technique named, Cross-Domain data fusion[1],[10], is used to integrate datasets generated from different domains. Cross-Domain data fusion methods focus on knowledge extraction and knowledge fusion as shown in Fig. 1.

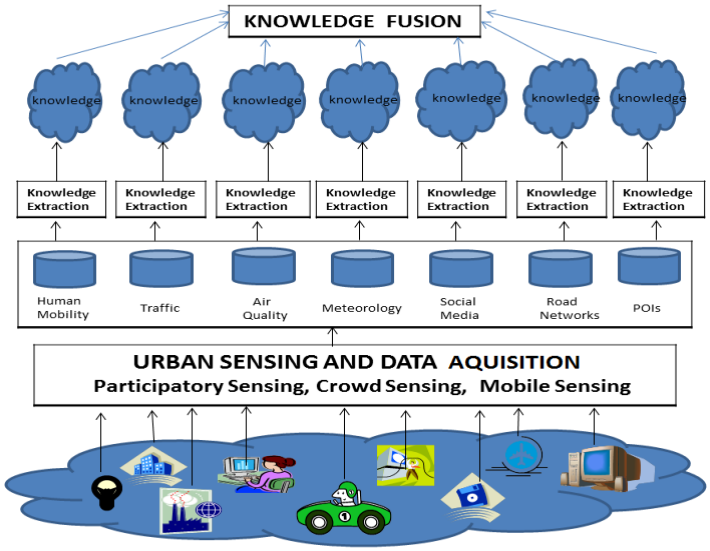

Fig. 1 Cross-Domain Data Fusion

Conventional Data Fusion is considered as a phase of Data Integration Process [2], [9]. It is defined as a process of combining data gathered from different sources in simple, consistent and clean representation for extracting knowledge. Fig. 2, shows the general framework of conventional data fusion. Initially, the correlated attributes which describe the data sources are identified. Then schema mapping techniques are applied for renaming or restructuring, which transform the data present in the data sources into a common representation. There after inconsistent representations of the data are identified by applying duplicate detection techniques over the data sources. Duplicate representations are combined and fused into a single representation by applying data fusion while inconsistencies in the representation are also resolved.

II. DATA FUSION MODELS

Data fusion models are mainly classified into three[3]: data-based model, activity-based model and role-based model as shown in the Fig 3. In this paper various data fusion models are discussed with their architecture and design

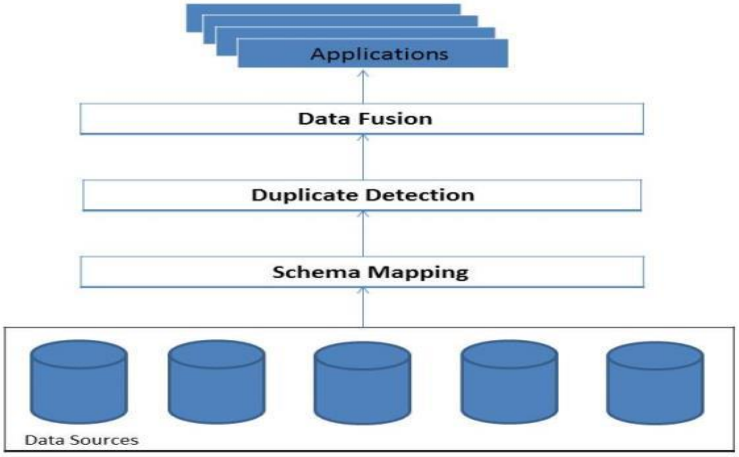

Fig. 2 Conventional Data Fusion

2.1 A. JDL(joint directors of laboratory) Data-based Model:

According To JDL model [4,5,7], "*Data Fusion is the process of combining data to refine state estimates and predictions*". This model is one of the data based model that focuses on the abstraction level of manipulated data by a fusion system.

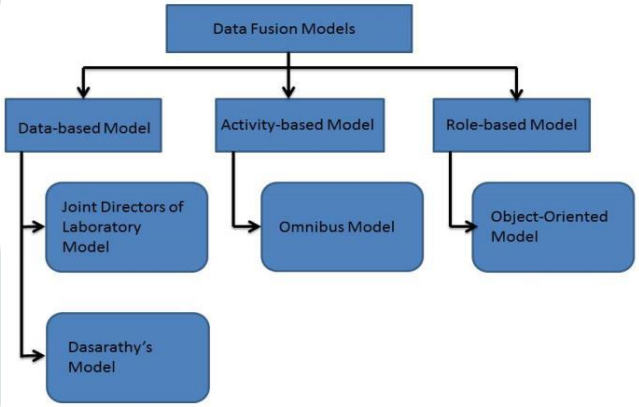

Fig. 3 Classification of various Data fusion Models

This model has been widely used for US data fusion community. According To JDL model, data fusion process contain five different levels from level 0 to level 4, input sources and a database for storage as shown in Fig 4.

Input Sources: provide input to the data fusion process. Different types of sources exist like sensors, priori information (summarized or referenced data), databases and human inputs. For example in traffic management system, all the traffic sensing devices which are monitoring vehicles are considered as the input sources.

Human/ Computer Interface: is an interface that provides input to the system through operator and reverts output to the operator. HCI includes commands, alarms, applied query, and information obtained from result. Desktop applications,internet browsers, handheld computers, and computer kiosks make use of the prevalent graphical user interfaces (GUI) [11]. Voice user interfaces (VUI) are used for speech recognition and synthesizing systems, and the emerging multi-modal and gestalt User Interfaces (GUI) allow humans to engage with embodied character agents in a way that cannot be achieved with other interface paradigms.

Database Management: stores the provided information, intermediate generated data, and fused results. It is a very critical component of framework as it stores highly diverse information. Databases is used to support internal operations of organizations and to underpin online interactions with customers and suppliers .They are also used to hold administrative information and more specialized data, such as engineering data or economic models. Examples of database applications include computerized library systems, flight reservation systems, computerized parts inventory systems, and many content management systems that store websites as collections of webpages in a database.

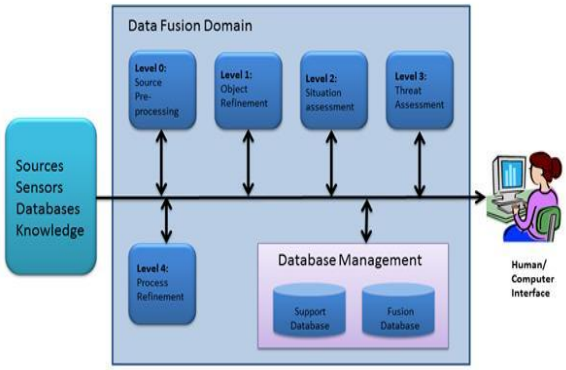

Fig 4: JDL Data Fusion Model

The five levels of data processing are explained as follows:

Level 0 - Source pre-processing: It is the lowest level of the data fusion processing. It is associated with pre-detection activity like pixel and signal processing, spatial and temporal registration. It reduces the amount of data and maintains useful information.

Level 1 - Object Refinement: Estimates and predicts the continuous or discrete state of objects and include procedures like identity fusion, association, clustering, and alignment, removal of false values and combining features of objects. The result of the level is object discrimination and object tracking. It transforms the input into consistent data structures.

Level 2 – Situation assessment: Estimation of relationship among the entities or objects. This level includes high level Inferences and identifying significant activities and events. The result of this level is set of high level inferences.

Level 3 – Impact Assessment/Threat Assessment: It includes identification of possible risks, vulnerabilities, and operational opportunities. Impact assessment focuses on specific themes, such as social impact assessments and gender impact assessments.

Level 4 – Process Refinement: This level provides efficient resource management while accounting for task priorities and scheduling and provides sensor management.

The JDL Model is often criticized for its implication that the levels necessarily happen in order and also for its lack of adequate representation of the potential for a human-in-the-loop. Despite these shortcomings, the JDL model is useful for visualizing the data fusion process, facilitating discussion and common understanding [13], and important for systems-level information fusion design [12].

2.2 Dasarathy's Classification[14]:

Dasarathy[5,14] classifies data fusion process into five categories on the basis of nature of input and output components as follows:

Data In – Data Out (DAI - DAO): As the name suggest, the input and output of data fusion process is data. Data is collected from sensors as input and data fusion process is applied on them and the resultant output is more accurate and reliable data. The manner in which human eyes fuse different colour information and generate new information is an example of this classification.

Data In – Feature Out (DAI - FEO): Data fusion system with input as data and output as features falls under this classification. It takes raw data collected from different sources and fuses them to produce features or characteristics of the object in the environment.

Feature In – Feature Out (FEI - FEO): Input and output of data fusion process are features. Instead of operating on collected information, it operates on set of derived features and produces some new improved and refined features. It is also known as feature fusion.

Feature In – Decision Out (FEI - DEO): It takes set of features as input of the data fusion process and decisions like pattern recognition as output. Classification systems that provide decisions based on features of sensor's input fall under this categorisation. For example, fusing the features extracted from data to generate a decision of finding the target class to which it belongs.

Decision In – Decision Out (DEI - DEO): Also known as Decision fusion. Local decisions made on the each sensor based on the derived data are passed to the fusion system that generates new and improved decisions

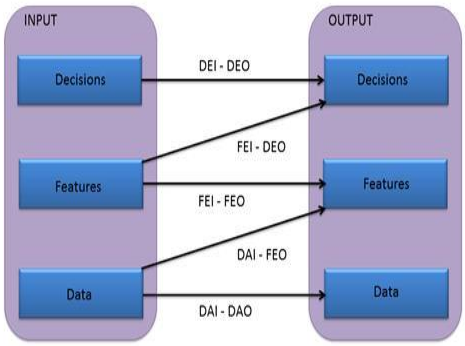

Fig. 5 Dasarathy's Model

2.3 Omnibus model [3,6]

It is an activity-based model, which describes the various states of data fusion process in a cyclic sequence. This model should be applied in design phase of data fusion system. It is composed of four stages shown in figure. These stages are explained as follows:

1. Sensing and Signal Processing: for collecting and pre-processing data.
2. Feature extraction: patterns or features are extracted from the gathered data.
3. Decision making: processed patterns or features are fused and set of decisions are made and threats are tracked.
4. Act and control: a best plan is chosen and executed to give some actions.

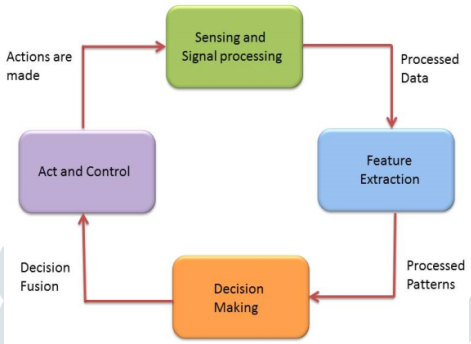

Fig. 6 Omnibus Model

2.4 Object-Oriented model [3,8]

It is proposed by Koker et. al. It is a role-based data fusion model having cyclic architecture representing roles. These roles are specified in fig. 7 and are described as follows:

1. Actor: role of interacting with the world,
2. Perceiver: role of analysing and evaluating the collected data,
3. Director: role of making plans to achieve goals set by system,
4. Manager: role of executing the plan made by director.

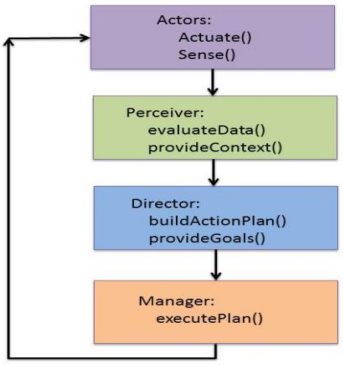

Fig. 7 Object-Oriented Model

III. COMPARISON OF DATA FUSION MODELS:

The table-I given below summaries various factors to compare and contrast the data fusion models discussed in this paper. JDL and Omnibus model faces difficulty when previous results are used to enhance the data fusion process. Other data fusion models as Dasarathy and Object-oriented model are simple and well executed.

TABLE I – COMPARISON OF DATA FUSION MODELS

Data Fusion Model	JDL Model [4,5,7]	Dasarathy 's Model[14]	Omnibus Model[3,6]	Object-Oriented Model[3, 8]
Type	Data-based Model	Data-based Model	Activity-based Model	Role-based Model
Advantages	Popular and general model.	Focus on the abstraction level of processed data and its output.	Describes the processing steps of system in cyclic loop.	Explains roles of the system.
Disadvantages	Difficult to reapply on an application due to its complex techniques.	-	Not good for specific applications.	System tasks are not separated

IV. CONCLUSION

With the growth of the web, its content is becoming more diverse, thus the role of data fusion becomes even more important. It is believed that data fusion techniques have interesting challenges with many advantages. To build an effective data fusion model, various issues need to be addressed as the size of the web is tremendous, with all sort of unstructured and disoriented content. In order to fuse data, we need to devise a technique to amalgamate datasets generated from different sources. In this paper various data fusion models have been studied. With exponentially growing data over the WWW, it will be progressively more important to fuse these datasets effectively for knowledge extraction.

V. REFERENCES

[1]Yu Zheng,"Methodologies of Cross-Domain Data Fusion: An Overview", IEEE VOL.1 NO. 1, January- March, 2015.
[2] JensBleiholder and FelixNaumann,"Data Fusion", Hasso-Plattner-Institute, Dec-2008.
[3] Marwah M Almasri, and Khaled M Elleithy,"Data Fusion Models in WSMs: Comparison and Analysis",IEEE, 2014.
[4] Dale A. Lambert, Information Technology Division, Defence Science and Technology Organisation, "An Exegesis of Data Fusion", 2003.
[5] FedericoCastanedo, "A Review of Data Fusion Techniques", 2013.
[6] Mark Bedworth and Jane O'Brien, "The Omnibus Model: A New Model of Data Fusion?"IEEE,VOL.15 Issue: 4, April-2000.
[7] Alan N. Steinberg, C. L. Bowman, Franklin E. White,"Revision in the JDL Data Fusion Model", 2000.
[8] Carl B. Frankel, Mark D. Bedworth, "Control, Estimation and Abstraction in Lessons from Human Information Processing Fusion Architectures", 2000.
[9] "Data Fusion for Cyber Intelligence", Ponemon Institute LLC, Big Data Analytics in Cyber Defense, Feb-2013.
[10] "Urban Computing: Concepts, Methodologies, and Applications", ACM Transactions on Intelligent Systems and Technology, Vol. 5, No. 3, Article 38, September 2014.
[11] Hewett, Baecker, Card, Carey, Gasen, Mantei, Perlman, Strong, Verplank,"ACM SIGCHI Curricula for Human-Computer Interaction", ACM SIGCHI, Retrieved 15 July 2014.
[12]Blasch, Erik P. Bossé, Éloi Lambert, Dale A., "High-Level Information Fusion Management and System Design", Norwood, MA, Artech House Publishers, 2012.
[13] Liggins, Martin E. Hall, David L. Llinas, James,"Multisensor Data Fusion", Second Edition: Theory and Practice (Multisensor Data Fusion), 2008.
[14] Belur V. Dasarathy, "Sensor Fusion Potential Exploitation – Innovative architectures and illustrative Applications", IEEE Vol. 85 No.1, January-1997.
[15] UN, 68% of the world population projected to live in urban areas by 2050, says UN. UN, 2018. [Online]. Available: https://www.un.org/development/desa/en/news/population/ 2018-revision-of-world-urbanization-prospects.html
[16] S. Nation and D. G. Office, "Smart nation singapore," 2018. [Online]. Available: https://www.smartnation.sg/
[17] X. Luo, D. Zhang, L. T. Yang, J. Liu, X. Chang, and H. Ning, "A kernel machine-based secure data sensing and fusion scheme in wireless sensor networks for the cyber-physical systems," Future Generation Computer Systems, vol. 61, pp. 85–96, 2016.
[18] N. Dawar and N. Kehtarnavaz, "A convolutional neural network-based sensor fusion system for monitoring transition movements in healthcare applications," in 2018 IEEE 14th International Conference on Control and Automation (ICCA). IEEE, 2018, pp. 482–485.