Meng, Tong; Jing, Xuyang; Yan, Zheng; Pedrycz, Witold

A survey on machine learning for data fusion

Published in:
Information Fusion

DOI:
10.1016/j.inffus.2019.12.001

Published: 01/05/2020

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

A Survey on Machine Learning for Data Fusion

Tong Meng , Xuyang Jing , Zheng Yan , Witold Pedrycz

PII: S1566-2535(19)30390-2
DOI: https://doi.org/10.1016/j.inffus.2019.12.001
Reference: INFFUS 1175

To appear in: *Information Fusion*

Received date: 24 May 2019
Revised date: 30 September 2019
Accepted date: 9 December 2019

Please cite this article as: Tong Meng , Xuyang Jing , Zheng Yan , Witold Pedrycz , A Survey on Machine Learning for Data Fusion, *Information Fusion* (2019), doi: https://doi.org/10.1016/j.inffus.2019.12.001

Highlights

- We sum up a group of main challenges that data fusion might face.
- We propose a thorough list of requirements to evaluate data fusion methods.
- We review the literature of data fusion based on machine learning.
- We comment on how a machine learning method can ameliorate fusion performance.
- We present significant open issues and valuable future research directions.

A Survey on Machine Learning for Data Fusion

Tong Meng[1], Xuyang Jing[1], Zheng Yan[1, 2], Witold Pedrycz[3]

[1]*State Key Laboratory on Integrated Services Networks, School of Cyber Engineering, Xidian University, China*

[2]*Department of Communications and Networking, Aalto University, Finland*

[3]*Department of Electrical & Computer Engineering, University of Alberta, Canada*

Abstract Data fusion is a prevalent way to deal with imperfect raw data for capturing reliable, valuable and accurate information. Comparing with a range of classical probabilistic data fusion techniques, machine learning method that automatically learns from past experiences without explicitly programming, remarkably renovates fusion techniques by offering the strong ability of computing and predicting. Nevertheless, the literature still lacks a thorough review of the recent advances of machine learning for data fusion. Therefore, it is beneficial to review and summarize the state of the art in order to gain a deep insight on how machine learning can benefit and optimize data fusion. In this paper, we provide a comprehensive survey on data fusion methods based on machine learning. We first offer a detailed introduction to the background of data fusion and machine learning in terms of definitions, applications, architectures, processes, and typical techniques. Then, we propose a number of requirements and employ them as criteria to review and evaluate the performance of existing fusion methods based on machine learning. Through the literature review, analysis and comparison, we finally come up with a number of open issues and propose future research directions in this field.

Keywords: Data fusion, machine learning, fusion methods, fusion criteria

1. Introduction

In the era of information explosion, huge volumes of data are created, collected and processed. We can extract and gain valuable information from data to look for the rules of the world and to discover the nature of things. Instead of believing in experiences or intuition, we are more likely and feel more confidence to draw a conclusion or make a decision on the basis of real-world data. However, big data also accompany with difficulties and challenges in data driven service provision because of its "5V" characteristics: Volume, Variety, Velocity, Veracity and Value. Obviously, traditional data processing techniques in the literature are hard to meet the demand in the new era of big data. How to capture reliable, valuable and accurate information in massive data is one of the most significant research topics nowadays.

The cyber world brings us overmuch data to dispose. However, raw data captured from various environments are heterogeneous, complex, imperfect, and of a huge scale, which brings us many challenges to transform them into useful information. All kinds of data processing technologies, including but not limited to data preprocessing, data storage, data transfer, data fusion, data analysis, information retrieval and so on, are major in solving these problems and stemming from diverse processing theories. In this paper, we focus on data fusion. It is a technology that merges data to obtain more consistent, informative and accurate information than the original raw data that are mostly uncertain, imprecise, inconsistent, conflicting and alike. Varieties of data fusion

methods have been designed in different application fields. Generally, data fusion is widely used in wireless sensor networks, image processing, radar systems, object tracking, target detection and identification, intrusion detection, situation assessment, etc. [1].

Traditional data fusion techniques include probabilistic fusion (e.g., Bayesian fusion), evidential belief reasoning fusion (e.g., Dempster-Shafer theory), and rough set-based fusion, etc. [2]. In recent years, the development of sensors, processing hardware and many other data processing technologies bring a new development opportunity to data fusion. As a technique with strong abilities to compute and classify data, machine learning is highly expected to improve the overall performance of data fusion algorithms.

Machine learning is a technique that lets the computer "learn" with provided data without thoroughly and explicitly programming of every problem. It aims at modeling profound relationships in data inputs and reconstructs a knowledge scheme. The result of learning can be used for estimation, prediction, and classification. The name of "machine learning" was first proposed in 1959 [3]. After decades, the advance of computation ability of digital computers notably improves the performance of machine learning. Machine learning enables classification and prediction based on known data and can achieve high accuracy and reliability, which makes it more likely to inform a correct decision. In recent years, machine learning has been applied into data fusion to improve its performance and offer satisfactory fusion results.

There are some surveys about data fusion published in recent years with different emphases. Alam et al. [4] completed a literature review on data fusion in IoT, which contains mathematical fusion methods such as probabilistic methods, artificial intelligence, and theory of belief in the domain of IoT. Focusing on IoT narrows down the review, while data fusion with machine learning covers a wide area. Gite and Agrawal [5] focused on data fusion models used in context-aware systems. Pires et al. [6] summarized the state of the art of data fusion techniques about sensors embedded in mobile devices. Navarro-Arribas and Torra [7] reviewed the approaches of information fusion for achieving data privacy. Faouzi et al. [8] concentrated on the application of data fusion models in intelligent transportation systems. Corona et al. [9] studied information fusion methods for computer security. Yao et al. [10] made an overview on web information fusion and integration. Ding et al. [76] reviewed data fusion methods in Internet of Things, mainly focusing on secure and privacy-preserving fusion. We can see that the above surveys hold different concentrations from our survey presented in this paper.

On the other hand, some works provide an overview on machine learning in some specific application scenarios, especially in big data processing related environments. For example, Liao et al. [11] surveyed machine learning applications and achievements in the past decade (2000-2011). Rudin and Wagstaff [12] reviewed the advances of machine learning in real-world problems of science and society. Qiu et al. [13] studied on machine learning for big data processing. They pointed out five significant issues in the learning of big data through a literature review. Zhang et al. [14] reviewed representative works of deep learning in big data.

In summary, we can find many existing surveys about data fusion and machine learning from various views. However, in the context of fast growth of artificial intelligence-based fusion models and their excellent properties, a survey specific to data fusion based on machine learning is still lacking. Although Alam et al. [4] provided a review on data fusion techniques with artificial intelligence, they only paid attention to

the literature about data fusion in Internet of Things. Their review is limited with regard to the scope of models. A horizontal comparison with detailed analysis is still missed. Considering the recent advance of machine learning, it becomes essential to comprehend elementary knowledge, current application state and future trends of this field with the help of a thorough survey.

In this paper, we perform a serious survey on data fusion techniques with machine learning. We first comprehensively introduce basic definitions and background knowledge about machine learning and data fusion. Then, we indicate critical challenges of data fusion and propose a number of criteria of data fusion. We make a deep-insight overview on data fusion techniques based on machine learning by commenting the performance of each reviewed work with the help of and by employing the criteria. Through analysis and discussion, as well as comparison, we find some open problems, which further allow us to indicate several research directions to motivate future research in this promising research field. In particular, the main contributions of this paper are described below:

- We sum up a group of main challenges that data fusion might face. Then, we propose a thorough list of requirements as uniform criteria that can serve as a measure to evaluate the performance of data fusion methods based on machine learning.
- We review the literature of data fusion based on machine learning in various application scenarios, discuss their advantages and weakness in detail according to the proposed criteria. In each literature review, how a machine learning method can ameliorate fusion performance is especially commented.
- Based on the completed review and in-depth analysis, some significant open issues and valuable future research directions are presented, which are useful and referable for the researchers and practitioners in this field.

The reminder of the paper is organized below. We provide an overview of background knowledge of data fusion and machine learning in Section 2. To review the literature comprehensively with a uniform measure, we propose a number of criteria on data fusion in Section 3. Section 4 reviews the recent literature about data fusion with machine learning that are categorized into three classes: signal level data fusion, feature level data fusion and decision level data fusion. All the literatures are reviewed with respect to their model structures, application background and technical advantages. Besides, we discuss their performance with the help of the proposed criteria. We also summarize the overall comparison of all the reviewed models/methods in this Section. In Section 5, we point out open issues and propose future research directions in this research field based on the result of literature review. Finally, conclusions are provided in the last section.

2. Overview of Data Fusion and Machine Learning

This section provides background information and concepts related to data fusion. It also specifies the challenges of data fusion and makes a brief introduction of machine learning and its common models.

2.1 Data Fusion

White [15] defined data fusion in the book “Data Fusion Lexicon” as “a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and

complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results." Hall et al. [1] thought that "information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making."

For easy understanding, we introduce the most important elements of data fusion:

- Data sources: Single or multiple data sources from different positions and at different points of time are involved in data fusion.
- Operation: One needs an operation of combination of data and refinement of information, which can be described as "transforming".
- Purpose: Gaining improved information with less error possibility in detection or prediction and superior reliability as the goal of fusion. Example purposes of actual applications are decision making, entity identification, situation estimation, and so on.

The superiority brought by fusion of multi-source data is quite obvious. Even in a static single source system, the fusion of sampling with replication can result in a more accurate observation. On the other hand, especially in wireless sensor networks, distributed data fusion reduces the redundancy of data, which reduce time and resource consumption and the frequency of data collision in the process of data transportation. What's more, in all data fusion applications, data is transformed into a modality with more value and higher quality, which makes a data fusion system able to reemerge the full view of an observed phenomenon. For instance, data enlarges its cover a lot in both time dimension and spatial dimension. In other models, appropriate handling on redundant data can help acquire improved, accurate and reliable information, with little imperfection.

Researchers began working in this field since 1960s, as a part of data processing firstly. Later, in 1970s [1], US Department of Defense (DoD) utilized this technology into military usage for defense and monitor. So far, in the military domain, there have been many applications including entity target identification and tracking, land, ocean and airspace surveillance, radar tracking, remote sensing, and so on. More than that, data fusion models are nowadays widely used in nonmilitary applications, e.g., fault detection in varieties of machines, intrusion detection, malware detection, review ranking, vehicle monitoring and prediction in traffic systems, environmental monitoring, pattern recognition, face identification, and so on [16, 75, 77-82].

Along with multiple application scenarios that data fusion used in, the term data fusion also has many extend forms with its own practical meaning. For example, multi-source/multi-sensor data fusion relates to the data from multiple sources compared with the data from a single source. Image fusion focuses on fusion of images. Information fusion concentrates on the data that has been processed, which is different from raw data fusion. Decision fusion is specialized to describe information in a high semantic level for making a decision. These terms might be used interchangeably with "data fusion" in some particular situations.

2.2 Architecture and Classification of Data Fusion

Apparently, raw data collected by collectors is usually not applicable for prediction

or other applications due to many reasons, such as data incompleteness, data confliction and data inconsistency. Therefore, methods are requested to deal with data imperfection. Furthermore, raw data cannot be extracted as information with high value by once. We need a hierarchical transformation to manipulate data systematically. Since data fusion is a complex system constituting of a number of parts to process data, we need to unify expressions or terminologies to describe each part's functionalities and characters. An excellent and concise architecture can also help researchers and developers communicate easily, which will promote the development of the research field. Herein, we introduce some widely-spreading data fusion architectures as below, which include Joint Directors of Laboratories (JDL) [15], the Luo and Kay architecture [19] and the Dasarathy's architecture [20].

2.2.1 Joint Directors of Laboratories (JDL)

JDL was first proposed by US Department of Defense (DoD) in 1986, which mainly aims at military usage. However, it can also easily adapt into nonmilitary use. In order to utilize the architecture extensively, there appeared many revised or intended versions of JDL data fusion models later on, which makes it fit into many application scenarios. In this paper, we only introduce the original JDL for easy understanding.

JDL data fusion model is a functional model, which describes a series of concepts and functions to identify each process in a data fusion system. Figure 1 shows the JDL data fusion architecture. There are five levels of data processing (level 0 – source preprocessing, level 1 – object refinement, level 2 – situation refinement, level 3 – threat refinement, and level 4 – process refinement) and three supporting components (sources, human-computer interaction (HCI), and database management) in the JDL architecture, as follows.

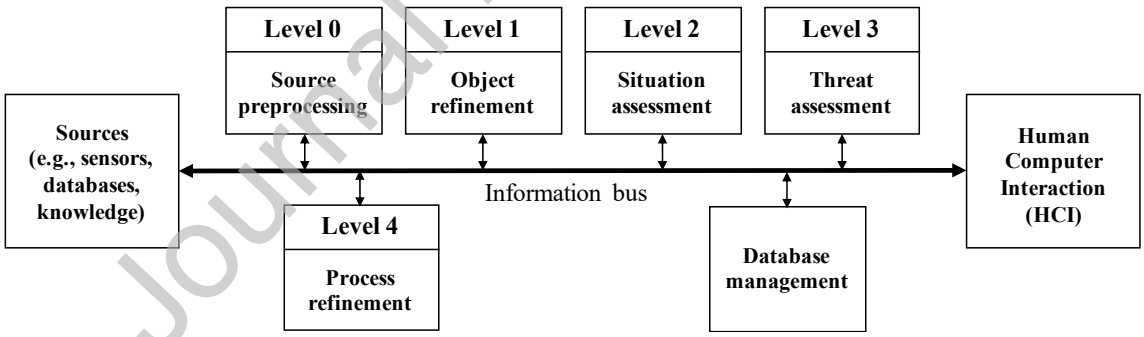

Figure 1. Joint Directors of Laboratories (JDL) architecture [15]

- Level 0 – source preprocessing: It is the lowest data processing level, which mainly deals with raw data in signal or pixel levels. Level 0 needs to prepare data well for next steps. Thus, its primary mission is to transform and assign data to a proper level for further processing. This step of data processing can obviously reduce system load and makes Level 1-3 pay more attention to the data corresponding to their own responsibilities without disturbing.
- Level 1 – object refinement: this step is responsible for outputting the identification information of individual objects. It focuses on identifying a particular entity. In this level, all static information about an entity's location, direction, state and other

attributes are collected and combined into a consistent pattern. Then, the system can get a comprehensive view of it from both time dimension and spatial dimension for a further estimation.

- Level 2 – situation refinement: Based on the individual entity‟s information gained from the previous level, this level broadens the horizon of investigation into the environment of the entity. The relationships between various entities form an environment and a situation, which is the main concern of Level 2. The relations between entities are defined based on communications and tightly connected with the environment.
- Level 3 – threat refinement: The current situation assessment from Level 2 helps Level 3 concern about threat and impact. Level 3 predicts risks, vulnerabilities and operational probability. Because judgement is based on much uncertainty information, process in Level 3 becomes quite difficult.
- Level 4 – process refinement: This level is the management part of the whole processing levels. It monitors other levels in real time, records performance of the system and makes decisions to improve system efficiency. For example, in this level, the system can find out what kind of information is currently scarce, approve each level‟s work in terms of getting source data or satisfying other particular needs, and direct the whole system.
- Sources: This component is the base of the whole system. It can be in many forms such as sensors (local sensors or distributed sensors), databases, priori knowledge, and so on.
- Human-computer interaction (HCI): This component is indispensable for smooth system execution. It allows human operations on the system, including commands, information inquires, messages about system results and decisions, and so on. In fact, HCI realizes assistance between human and computer reciprocally.
- Data management: This component stores data in different forms containing raw data and information. Different processing levels interact with data management frequently. Its responsibilities include but not limited to data retrieval, data storage, data security, and data compression. The big amount of data involved and the need for rapid interaction make data management a tough task.

2.2.2 Luo and Kay's Architecture

Luo and Kay studied multi-sensor integration and fusion [19, 69]. They provided a new general architecture of multi-sensor integration based on the abstract level of used integrated data, as illustrated in Figure 2.

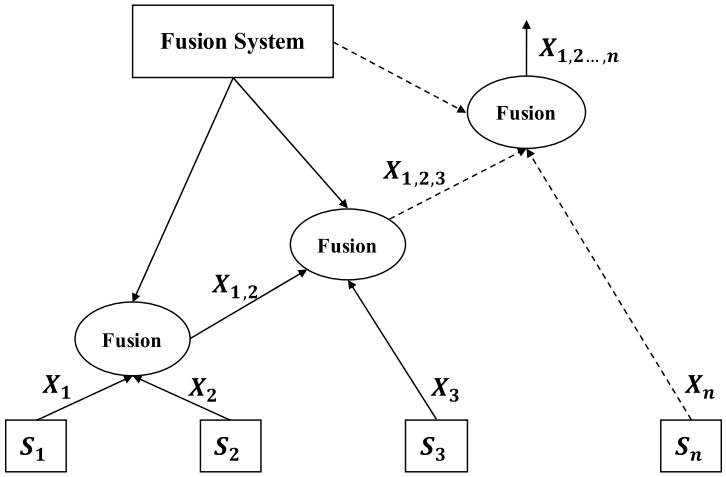

Figure 2. Luo and Kay's architecture [19]

In the Luo and Kay's architecture, raw data come from sensors, and are fused in the nodes of an information system. For example, data from sensor 1 and 2 can be fused as data $x_{1,2}$. After that, the output data $x_{1,2}$ will be further fused in the next fusion node with data from sensor 3, turning into data $x_{1,2,3}$. Similarly, data $x_{1,2,\dots n}$ from the last fusion node is the highest fusion result. The authors summarized four levels from low to high to represent data in different fusion process including signal level, pixel level, feature level and symbol level. The different levels deal with different input data patterns, are applied in various systems for a variety of purposes and also provide distinct degrees of promotion of information quality.

- Signal level: Raw data captured from sensors are as input into fusion models to be combined directly. The fusion models corresponding to this process belong to the category of signal level data fusion. Data will be turned out with higher accuracy, less noise or refined features after this fusion process. If raw data are commensurate or in the same pattern, they can be fused in this level. Signal level fusion sometimes occurs in real-time fusion scenarios or may be an additional step in preprocessing of signals. Sometimes researchers also called these models as "low level fusion" or "raw data fusion".
- Pixel level: It is a special case of signal level fusion for image processing especially. Fusion in pixel level promotes some image processing applications like segmentation.
- Feature level: In feature level, not raw data but features or characteristics take part in the fusion process. Sensor data are often preprocessed into certain necessary features first before fusion is conducted. As an output, we can obtain refined characteristics or features in other patterns for achieving other targets, or data in a higher level – decision level. Feature level data fusion is also known as "medium level fusion" or "characteristic level fusion".
- Symbol level: Symbol level data fusion has a more common name – decision level data fusion, referring to dealing with some information that is refined from sensor data and has already been generated to represent some determinations of a task. Usually, a global and accurate decision is highly required through data fusion. Apart from decision level data fusion, the symbol level data fusion is also known as "high level fusion". Compared to low-level fusion, symbol level fusion methods often

generate preliminary classification and can fuse different types of data to obtain accurate fusion results.

The Luo and Kay‟ architecture intends a hierarchical fusion scheme to transform data from a raw state to a form of high quality. Data sets captured from sensors follow the order of processing to become useful information for the purpose of assisting decision making or estimation.

2.2.3 Dasarathy's Architecture

Based on the Luo & Kay‟s three-layer (data-feature-decision) fusion architecture, Dasarathy extended it into five fusion processes regarded of I/O characterization in 1997 [69]. Dasarathy thought that some ambiguous conditions in the three-layer architecture lead to the demand of a more precise definition. Thus, he reformed the old architecture from I/O perspective and classified data fusion models into five categories: Data In-Data Out (DAI-DAO) Fusion, Data In-Feature Out (DAI-FEO) Fusion, Feature In-Feature Out (FEI-FEO) Fusion, Feature In-Decision Out (FEI-DEO) Fusion and Decision In-Decision Out (DEI-DEO) Fusion, shown in Figure 3. The new classification defined in the Dasarathy architecture considers the nature of input data and output data, which reduces uncertainty in the three-layer architecture.

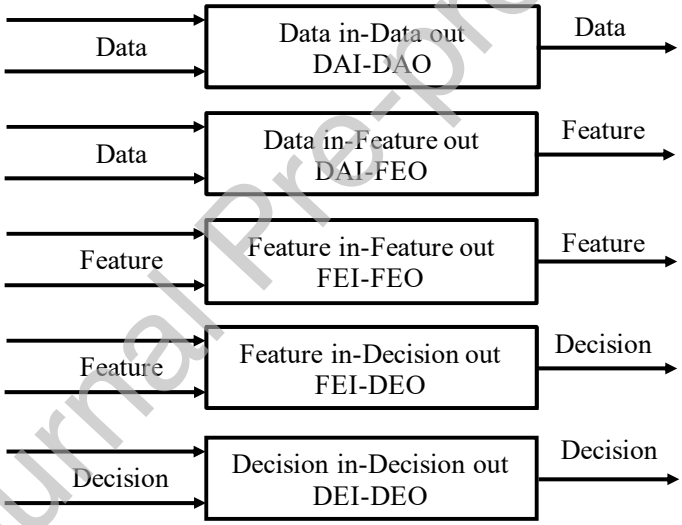

Figure 3. Dasarathy's architecture [69]

- Data In-Data Out Fusion: This type of fusion processes input data to make them more accurate or polished. It is the most elementary and basic layer in fusion family. DAI-DAO fusion instantly appears after raw data captured from an environment. Its typical applications include signal processing and image processing.
- Data In-Feature Out Fusion: In this type of fusion, data sets are first integrated and extracted into some abstract information, called feature. Some simple and intuitive results can be gained from raw data by applying DAI-FEO fusion.
- Feature In-Feature Out Fusion: Apparently, most feature fusion algorithms belong to this category, with feature inputs and also feature outputs. Different from data inputs, feature inputs often show some refined characteristics, which have been extracted preliminarily already.
- Feature In-Decision Out Fusion: The majority of fusion algorithms fall into this

category. And most of them are for the purpose of classification, which is a typical case of decision. With feature inputs, a sequence of decisions can be obtained. Another example of this type of fusion is pattern recognition. Features transmitted from multi-sensors are recognized with priori knowledge to form a decision.

- Decision In-Decision Out Fusion: As the highest fusion level in the Dasarathy's architecture, DEI-DEO fusion transfers some decisions in low-level or local fusion nodes to a global decision, which comprehensively consider information of all low-level or local level decisions.

There are also many other data fusion architectures, such as Bowman Df&Rm architecture [17, 18], Durrant-Whyte architecture [20], Pau Architecture [67], Laas architecture [68], and so on. They specify data fusion process from different views and each proposed architecture has its own advantages and characteristics in comprehending or modeling particular applications. Steinberg et al. [21] and Ayed et al. [22] compared these architectures in detail.

2.3 Data Fusion Challenges

Data fusion is still confronting a number of challenges in order to maximize its advantages [24] although various data fusion models were proposed to address specific demands in many concrete applications. Most of these challenges are resulted from the complexity of application environments where sensors are located, the variety of data that should be combined, and so on. In this subsection, we list some of them as below.

(1) Data imperfection: It is a common problem and a main issue that all data fusion methods are expected to settle. The data captured by sensors are often imprecise, uncertain, ambiguous, vague, and incomplete. Usually, we can improve data quality by modeling its imperfection and making use of other available information and powerful mathematical tools. Data imperfection will seriously affect fusion quality if precise and useful data cannot be extracted by data fusion.

(2) Data inconsistency: There are some uncertainties caused by inherent noises in measurements, sensors and also environments. These noises lead to data outliner or disorder, which is collectively known as data inconsistency. Apparently, data inconsistency introduces extremely bad effects to data fusion if a fusion model cannot distinguish the reasons that cause the noises. Data fusion techniques should overcome this problem by eliminating the influence of data inconsistency. In addition, there are some spurious data caused by lasting or dynamic failures, which are difficult to model and predict in usual ways.

(3) Data confliction: This issue often appears in a system applying belief functions or Dempster-Shafer theory. When some problems that should be treated independently are erroneously integrated, a representation error occurs.

(4) Data alignment/registration and correlation: Data captured from different sensors with different frames must be aligned into a common frame before they are fused, which is called data alignment or data registration. An over/under confidence will happen if some errors happen in this process. There are also some other challenges, such as data correlation, which appears mostly in a distributed environment when a same set of data is computed or fused more than once mainly because of cyclic tracks in topology, called data incest phenomenon. Correlated data often markedly affect a fusion system with serious biased estimation if it cannot be eliminated by data fusion algorithms well.

(5) Data type heterogeneity: Data are captured by sensors in different environments. So, they might belong to quite different types. Just like people's eyes,

nose, mouth, sensors are with different purposes, too. Data fusion methods should be able to integrate different types of data to describe the whole status of an object.

(6) Fusion location: This is also an outstanding problem in wireless sensor networks and other distributed fusion environments. Data can be fused in a central node or a local node. The first manner costs more bandwidth and time. With the later manner, we can reduce communication burden, but we have to give up data accuracy certainly because of the information loss of local fusion. How to balance fusion cost and fusion quality is a tough issue.

(7) Dynamic fusion: The complexity of data fusion is caused by not only data type and collection environment, but also its timeliness. To estimate a system state, especially for a time-varying system, data might be significant only in a limited time period. This challenge should be dealt with well in a real-time application environment. Fusion node should be able to distinguish the right order of data and its validation.

2.4 Machine Learning

Machine learning is one of the hottest research topics because of the massive impact of Alpha Go and other artificial intelligence applications. Machine learning is a sub field of computer science and artificial intelligence. It describes a field that utilizes some particular algorithms to make computer systems "learn" by using given data without specific programming. Specifically, it is a process to let computer systems or machines see, know, learn and predict the world like a human being. "Machine learning is the study of making machines acquire new knowledge, new skills, and reorganize existing knowledge" [70, 72-74]. At the beginning of the birth of machine learning, people performed researches to let a machine study, gain skills and build its own knowledge world automatically. After that, Samuel proposed the term "machine learning" explicitly in 1959 [3], which was evolved from some artificial intelligence study fields such as pattern recognition and computational learning theory. The main idea of a machine learning method is to let the computer have the ability to acquire experience and adjust itself accordingly without too much human intervention. It is suitable for solving such problems that are difficult to program or model.

Data plays an important role in machine learning. Data patterns determine learning results and effects. Machine learning need some data inputs firstly, which are also known as samples, training sets and instances. With the help of provided data sets, a machine reconstructs internal relationships of them, which is the result of "learning" (also known as „training"), and presents acquired knowledge by the means of specific output forms like recognition, classification and prediction (known as „testing"). More concretely, regression models produce a mathematical variable; classification models form a categorical variable, and so on.

Machine learning methods are usually divided into three classes based on if a given data set has labels about its attributes for learning: unsupervised learning, supervised learning and semi-supervised learning [23].

If the attributes of input data sets and output data sets are completely labeled, the goal of machine learning algorithms becomes to construct a model to map input to output, which is called supervised learning. Representative applications in supervised learning include classification, regression, and so on. Two typical supervised learning algorithms are introduced below:

- Support Vector Machine (SVM): SVM is a typical supervised learning model to

realize binary classification. With a series of training data sets with labels, each data marked as belonging to one or the other of two categories, an SVM training algorithm constructs and trains a model that can arrange new data into one category or another, making it a non-probabilistic binary linear classifier. An SVM machine works out a hyperplane or a set of hyperplanes in the feature space between classes. SVM models are particularly suitable for classifying inconsistent sensor data with high dimension features.

- Neural Network (NN): Also known as Artificial Neural Network (ANN), NN is a large-scale intricate network constituted of a set of layers including an input layer, a couple of hidden layers and an output layer. Each layer has many nerve cells. The inputs of a current player"s nerve cells are the outputs of a former layer"s nerve cells. Given with training data sets, NN learns specific parameters of the whole network with feed-forward or feedback. Due to the complex structure of NN, it is often trapped with long runtime and local minima problem. There are also some derivative NNs, such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), and so on.

In unsupervised learning, there are no labels given with datasets. Algorithms often extract features and patterns by themselves. Usually, according to similarity or distance of data inputs, the models build profound association with the help of internalized heuristics. Clustering methods are representative unsupervised learning algorithms. Compared to supervised learning, or more exactly, classification, dealing with pre-defined labels, clustering does not have any advices or conducts. A data clustering model classifies data in the way that putting objects with similar attributes in the same group (i.e., a cluster). Some typical clustering algorithms have been widely used in various applications. For example, connectivity models, hierarchical clustering, which are constructed based on distance connectivity; centroid models such as k-mean algorithms that use a single vector to describe a class; distribution models such as expectation-maximization algorithms that manipulate data with statistical distributions. Because k-means is a commonly used machine learning algorithm that is applied to many data fusion methods, we discuss it in detail.

- K-means might be the most extensively used clustering method where a structure in data is revealed by minimizing a given objective function. With n data positioned in a d-dimensional space, k points are randomly chosen as clustering centers initially. The distance between every data and the nearest center is calculated. The objective of optimization is to achieve the least distance and local squared-error distortion by recalculating the cluster centers and arranging the distribution plans repeatedly. K-means belongs to variance-based clustering. In fact, clustering is an NP-hard problem, thus there is no general solution. There are some representative efficient models for solving the k-means problem such as those presented in [70] and [71].

In case that the machine learning is based on a given training data set that has incomplete labels, the machine learning is semi-supervised learning. In this case, data inputs with labels will play a leading role in forming a decision boundary. While a large set of data inputs unlabeled will also help in improving the accuracy of decision boundary and the stability of the whole model.

3. Criteria of Machine Learning for Data Fusion

In this section, we list the criteria that a data fusion model or algorithm should satisfy

in order to employ them as evaluation metrics to review the literature in the next section. In what follows, data fusion model, method and algorithm are used interchanged with the same or similar meaning if not specially annotated. Facing the challenges as mentioned in Section 2.3, we propose a list of criteria to comprehensively and thoroughly evaluate the performance of data fusion.

(1) Efficiency (Ef): Efficiency is used to evaluate if a data fusion model makes use of resources economically. In most application scenarios, system resources are limited in terms of computation, bandwidth, storage space and many other aspects. Dealing with as more as possible data in an as shorter as possible time interval with as less as possible system resources should be a universal goal of a fusion model. The efficiency reflected by execution time should be evaluated to demonstrate model advance through comparison with other models.

(2) Quality (Q): Obviously, it is the most important criterion for evaluating a fusion model. What is the direct impact on a fusion algorithm? To which degree does the model improve information accuracy? Quality is the core of data fusion. In a specific application scenario, there should be corresponding assessment metrics. Quality should be inspected by checking if the above questions are answered with sufficient evidence, e.g., experimental results and reasonable explanations.

We divide all the literatures that dealt with fusion quality into two types: the ones with ideal fusion result and the ones without ideal fusion result. For the former, we use Root Mean Squared Error (RMSE) to measure the bias of a calculation result and observation, which directly describes fusion quality:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{M_1}\cdots\sum_{j=1}^{M_k}[R(x_i,\cdots,x_j)-F(x_i,\cdots,x_j)]^2}{M_1 M_2 \cdots M_k}},$$

where $M_1, M_2, \cdots, M_k$ refers to k dimensions of data sets in an application environment. R denotes ideal result and F stands for corresponding calculated fusion result. A smaller RMSE means lower bias between the ideal result and the fusion result, which leads to better fusion quality, apparently.

For the study that does not have referenced data for comparison, different issue deserves a specific analysis. For example, in image fusion, we evaluate fusion quality with the help of Structural Similarity Index (SSIM) of the original images *a, b* and fusion result image *f* [65-66]:

$$Q(a,b,f) = \lambda_a Q_0(a,f) + \lambda_b Q_0(b,f),$$
$$Q_0(a,f) = 4\sigma_{ab}\frac{\bar{a}\bar{b}}{((\bar{a}^2+\bar{b}^2)(\sigma_a^2+\sigma_b^2)}.$$

Where $\bar{a}$ is the mean value of a, σ_a^2 is the variance of a. σ_{ab} is the covariance of a and b. For simple calculation, we use a sliding window to divide the whole problem. We define $\lambda_a(w)$ and $\lambda_b(w)$ as below:

$$\lambda_a(w) = \frac{s(a|w)}{s(a|w)+s(b|w)},$$
$$\lambda_b(w) = 1-\lambda_a(w),$$

where $s(a|w)$ can be any statistical characteristic of image a in window w, such as variance or marginal information. Thus,

$$Q(a,b,f) = |W|^{-1}\sum_{w\epsilon W}(\lambda_a(w)Q_0(a,f|w) + \lambda_b(w)Q_0(b,f|w)).$$

The value of Q is between [-1, 1]. The closer the value of Q to 1, the better fusion quality an algorithm has.

(3) Stability (St): Stability is used to evaluate a fusion model's ability to keep working well in a stable manner in different situations. What we need is not just a

disposable system with expensive costs in installation and debugging. A steady model can persistently achieve high performance. Even with few abnormal situations, expenses are saved in handling exceptions and routine maintenance in reality. In the literature, multiple testing data sets were adopted to examine the stability of a fusion model [25, 26, 27].

(4) Robustness I: Robustness evaluates the strength of a fusion model to resist disturbance. When an underlying environment is changed, fusion quality should be ensured. For example, in a radar system, raw data captured from sensors are not stable all the time. It is highly expected that a fusion algorithm should effectively remove outliners, noises and communication errors as its best. If the fusion model can overcome this problem with a stable fusion result, this model is robust.

(5) Extensibility (Ex): Extensibility means that a data fusion model can be easily further improved and widely used in many situations. For similar application environments with alike targets, the model can be applied in a generic and pervasive way. Extensibility is a valuable feature for wide adoption of the data fusion model in practice.

(6) Privacy (P): In some application scenarios, data used for fusion may be sensitive and private, which induces security requirements on the fusion model. We use privacy to describe such a demand. In the environment where non-public data sets are processed, data should be protected during fusion to avoid any sensitive information leakage in subsequent steps. Which encryption algorithm or privacy protection scheme should be applied and how to manage procedures including but not limited to encryption, fusion, transmission, decryption and storage will be the key objectives of privacy protection.

(7) Tested with real world data sets (Re): In a solid research, experiments are dispensable to testify the performance of a model, prove its effectiveness, and show its advantages. Obviously, the experimental results will be more persuasive if researchers utilize data sets captured from real application scenarios. It is highly preferred if the whole experiments are done in practice rather than in a simulated environment.

4. Machine Learning for Data Fusion

In this section, we review the state of the art of machine learning for data fusion by classifying the current works into three categories: signal level data fusion, feature level data fusion and decision level data fusion. In each category, we review the literature based on the type of machine learning. For each work, we summarize its main contributions and characteristics, and comment on its performance based on the proposed criteria. At the end, we summarize and compare all the reviewed works in Table 1.

4.1 Signal Level Data Fusion

According to the Luo & Kay architecture, the lowest level of data fusion is signal level fusion. With raw data inputs captured from sensors, data outputs with high accuracy, reliability and few noises are captured. Or feature outputs are extracted to directly reflect an aspect in observation. Signal level models are often applied in signal fusion, image fusion (also known as pixel fusion) and other similar scenarios.

Single Level Data Fusion Based on Supervised Learning

As a representative supervised machine learning algorithm, SVM provides a proper fusion function in the signal level. Banerjee et al. [28] proposed a hybrid method for fault detection based on multi-sensor data fusion with SVM, Short Term Fourier Transform (STFT) and a time duration based observer model. The system classifies the state of a system into three kinds: healthy, degraded and failed. The specific scheme of the SVM based fault classifier is described in Figure 4. Raw data from sensors are firstly preprocessed in STFT, which is mainly separated based on the frequency level and amplitude of a signal. Then, an SVM classifier, which has been previously trained with labeled data, will transform the input signals into a high dimensional feature space and separate signals in a linear way into original signals and signals with fault. After that, a sensing system with time duration based observer receives signals from classifiers and judges which state the system is with the help of a threshold. The threshold is a tolerance level of the system. Crossing a safety valve means a signal gives an unwanted response, which will lead to a state change in a finite state model. At last, the output of the system is divided into three states: healthy system (i.e., the state of the system does not change in the observing phase), degraded system (i.e., the change of the system is in a tolerance level), and failed system (there are indeed some signals crossing the safety valve). The proposed model can monitor the working state of a motor in a certain interval of time with a prior alarm if there is any unwanted situation happening. Since the sensors capture data varying nonlinearly, SVM as an excellent nonlinear pattern recognition tool, especially in dynamic procedures ensures accuracy and performance in fault diagnosis at the same time. Classification accuracy and performance of average classification are improved compared with the system without fusion. Experiments on one to ten sensors in the system showed good fusion performance, which implies sound model extensibility. The experiments were performed based on a practical system. However, efficiency, stability, robustness and privacy were not mentioned in this work.

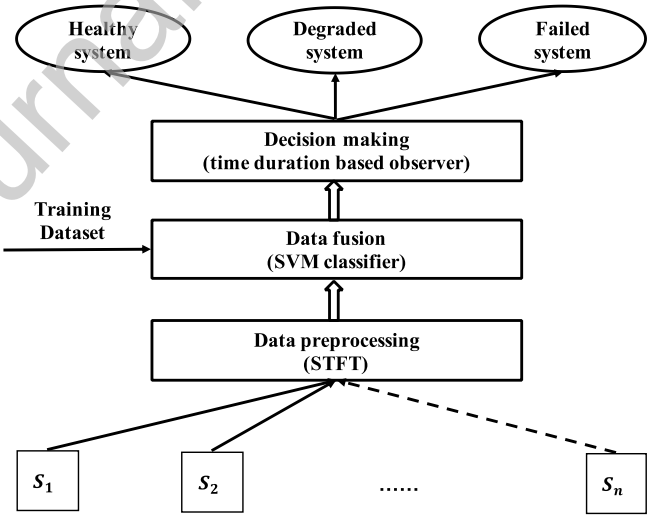

Figure 4. The model structure of [28]

In distributed data fusion systems, a disturbing problem often exists. Raw signals obtained from sensors are usually stored as a large set of samples. While the transmission bandwidth of a data fusion system from sensors to a fusion center is not adequately large, usually with a distinct limit. This causes the problem to transmit these

data sets for next step process within an available time limit before data become expired. Challa et al. [29] optimizes a Bayesian approach to data fusion with SVM, which is used as a technique for compressing information. It minimizes the objective function of SVM to transform input signals to a small set of signals called support vectors, which is described as its approximation function. Other non-support vectors are discarded since related signals do not contain useful information. Correspondingly, a kernel dictionary of the SVM is given for model modification to achieve sound efficiency based on different practical application environments. This model was tested in a density estimator system, which shows excellent performance in data compression. Thus, it performs well in both fusion efficiency and extensibility. On the other hand, it acquires many training samples to certify its strength, thus it does not perform very well in terms of robustness and stability. Through experimental result analysis, we think fusion quality should be further improved.

Fusion based on SVM could overcome fusion challenges regarding imperfect data. Fahmy et al. [30] proposed an improved SVM-based data fusion algorithm. It applies SVM into biometric fusion to fuse iris and fingerprint data to gain high accuracy. However, the performance of traditional Linear SVM is not good enough. The authors emphatically studied a technique called score normalization. Although some literatures previously published assumed that this procedure is not necessary in statistical learning fusion like SVM, this work illustrated the fault of this viewpoint. Essentially, score normalization is an important internal part of SVM procedure that helps transforming raw data of individual factors into a uniform pattern. The normalization method improves efficiency and robustness of the traditional SVM model. What's more, time consumption is also reduced in both training phase and testing phase because SVM can deal with the result directly. A number of score normalization methods were introduced and tested based on Radial Basis SVM with CASIA and FVC2004 databases, which proves the high fusion quality and stability of the enhanced SVM model. However, other criteria were not discussed in this work.

For signal level data fusion in WSN, Back Propagation Neural Network (BP-NN) is a typical solution. However, the BP-NN based fusion model often has long convergence time, which causes low fusion efficiency and a short life cycle of nodes. Shi et al. and Tan et al. improved BP Algorithm-based WSN data fusion from two viewpoints, respectively [40, 41]. Tan et al. applied WSN data fusion in forest fire monitoring [40]. They took advantage of the Levenberg-Marquardt algorithm to ameliorate time and energy consumption in a classical BP-NN. Simulation results indicate improved efficiency compared with ordinary BP-NN algorithms. Correspondingly, Shi et al. optimized the classical BP-NN with the Speed-constrained Multi-objective Particle Swarm Optimization (SMPSO) algorithm. Their method [41] can reach convergence with the least iteration steps compared with a classical BP-NN algorithm and an improved BP-NN algorithm, which shows its efficiency. Simulation results also proved that the proposed algorithm is adaptive in a large-scale network.

Many application environments such as human motion analysis and human-machine interface have a quite crucial need on precise location. Multi-sensors set in different locations capture data of position information of a target from different views. Fusion models are expected to solve the imperfection of these data sets to get complete knowledge of the location of the target. Kolanowski et al. [32] proposed a navigation system based on Elman Artificial Neural Network (ANN), which is good at resolving nonlinear problems especially in prediction. The system first uses Automatic Heading

Reference System (AHRS) to analyze data sets from sensors. The input and output data sets are used to train Elman ANN. Elman ANN model has 9 input neurons and 3 output neurons. There is at least one hidden layer between the input layer and the output layer. There is also a context-sensitive layer only connected to the hidden layer that stores the information of previous hidden layer. The context-sensitive layer can be seen as a representation of feedback. The authors also changed the number of neurons of feedback loop for achieving better performance. Experimental results with Elman ANN show few errors compared with AHRS, which indicates that Elman ANN is an efficient alternative for position detection. The reduction of trigonometric operations and matrix operations makes an improvement on time cost. Thus, this system achieves Efficiency and Quality. However, this work does not discuss other criteria as proposed in Section 3.

Tong et al. [34] proposed an information fusion model for boiler drum water level measurement. There is a crucial need for precise water level measurement in drum because the inbalance between boiler load and feed-water will lead to serious consequences. Differential pressure level measurement is a convenient and efficient method utilized in this problem. However, it does not behave well with regard to robustness facing with disturbances. A Radial Basis Function (RBF) neural network model, which is expected to be able to map highly nonlinear models, was designed to fuse such attributes as operating pressure, operating temperature, water inflow, and so on. Compared with BP-NN, RBF networks solve more problems such as local optimization. With an improved gradient descent algorithm, the RBF neural network can modify error of drum level measurement well. Simulation results show that with a two-step training method, both the number of errors in output and the training time are reduced very quickly, which indicates high efficiency and quality of this fusion model. The authors tested the model with 20 samples, the accuracy of testing results (i.e., the maximum level of the level error is less than 1 millimeter) shows its sound performance regarding Robustness and Stability. But Extensibility and Privacy were not considered in the paper. Neural networks show their strong ability in dealing with a nonlinear problem when it is difficult to be described as a function directly.

Single Level Data Fusion Based on Unsupervised Learning

For both military and nonmilitary usage, multi-radar data fusion is an important technique for target identification and tracking with high accuracy. Shu et al. [31] focused on discriminating and tracking multi-objectives at real-time. In target observation fields, multi-sensor for multi-target tracking is difficult to realize because there is a requirement of discrimination of a goal among lots of targets from the data observed by the same sensor and the combination of data from different sensors with regard to one target. A K-central clustering method was utilized to optimize this target identification and tracking model to find the path of a target. Given with a large set of real-time sensor data without labels, the algorithm is expected to cluster them into valuable categories, which are also the batches of targets. K-central clustering chooses a center of each cluster and trains distribution of points to make the sum of the distances between the centers and other points the minimum. The simulation results indicate that the k-central clustering method solves data association problem efficiently and gains better tracking results compared to original filtering methods, which demonstrate good fusion quality. Efficiency, Stability, Extensibility and Privacy were not considered in this paper. The experiments were carried out in MATLAB, not in a real environment. The model is robust in dealing with data with ambiguity and noise.

In a high-resolution radar system, there is special requirement on the efficiency of data processing on account of the large scale of raw data and the need of real-time fusion in target monitoring or tracking. Li and Wang [39] proposed a fast data fusion algorithm based on clustering. This algorithm divides raw data into clusters based on single dimensional distance. The authors also analyzed the calculation complexity of the proposed algorithm as O(m*n). Experiments showed the outstanding improved fusion efficiency of the model compared to K-means, Hierarchy and some other data fusion algorithms in the same application environment. Particularly, the authors considered serious noises in data collection of the radar system. To enhance the robustness of the algorithm, noise removing is performed at the end of the algorithm.

Similarly, Wang et al. [36] proposed a hierarchical clustering algorithm based on the K-means method for multi-target tracking. In this paper, target tracking problems with targets detected by multiple radars were described in detail. For example, target route is irregular, radar tracks are not uniform in time or have no common interval. To solve these problems, a hierarchical clustering model was built. After data preprocessing, Hausdorff distance that describes the similar level between tracking data sets was defined and calculated. Data sets with Hausdorff distance become a class and constitute a cluster search tree. According to the clustering algorithm, similar classes are merged into a new class to build the hierarchical clustering tree. At last, an improved K-means algorithm was designed to deal with final clustering, which is also the most important fusion process. Tests with real radar data showed the effectiveness, stability and also the high tracking accuracy of the algorithm.

As one of hot topics in WSN, anomaly detection is attracting more and more attention. There are a number of distinctions between WSN and ordinary networks, which might lead to many serious problems if we simply transplant traditional outlier detection techniques into the WSN environment. Firstly, WSN has severe resource constraints especially in battery life, computational capacity and also communication overload, which makes it hard to afford expensive or complicated computation. It also has high demand for online and real time detection without prior knowledge because of the characteristics of data in W– -- distributed streaming data. Guo et al. [26] proposed an anomaly detection model to solve the above issues. A lightweight data fusion algorithm named Piecewise Aggregate Approximation (PAA) was proposed to compress raw data collected by sensors, which greatly reduces transmission overload. Then, K-Means, an unsupervised detection algorithm improved with Artificial Immune System (AIS) completes classification of normal data and abnormal data, namely outliner detection. Compared with other WSN detection algorithms, this model not only consumes less energy and time, but also offers a higher detection rate and a lower false alarm rate. Thorough experimental result comparison and analysis show comprehensiveness of this work. Besides, experiments based on virtual and real data demonstrate the stability and effectiveness of the model.

To gain good fusion efficiency, routing protocol design becomes an important issue in WSN. An appropriate routing protocol considers many factors such as the topology of the whole network, the capability of fusion nodes, the time limits of valid signals captured by sensors. Xiao and Liu [33] provided a routing protocol based on Un-even clustering and a simulated annealing algorithm. Compared to the classical protocol LEACH (Low Energy Adaptive Clustering Hierarchy), two obvious differences are un-even initial clustering and dynamic time interval for cluster head reselection. At the start of the protocol, the base station clusters all nodes based on their position

information and energy information with the simulated annealing algorithm. Sensor members transmit their data sets to their corresponding cluster head in the next phase. Cluster heads execute data fusion and send the fused information to its next hop. There is a threshold about the residual energy of the cluster head to examine if it is suitable to continue in real-time. If it is not, a new cluster head will be chosen by the base station immediately and a new round will begin. The proposed protocol prolongs the alive time of the whole network and reduces total energy consumption. It improves the performance of a WSN with a distributed data fusion function from the view of resource consumption. Thus, fusion efficiency is improved a lot. Experiments were performed based on a sensor simulation tool without real environment tests. Other criteria were not mentioned in this work.

The weighted algorithm based on fuzzy logic is a classical data fusion algorithm. Due to its excellent performance in calculating weighted factors and dealing with imprecise data, the fusion algorithms based on fuzzy logic have been paid much attention. However, raw data in WSN do not adapt to the traditional weighted fuzzy logic algorithm ideally because invalid data appear frequently during data collection in a real-world environment, which might lead to serious measure deviation. Wang et al. proposed an improved fusion method with k-mean clustering [38] aiming to solve this problem. The K-means clustering method is applied to preprocess raw data before calculating the weighted factors. They divided raw data into different clusters and the error data with high variance are arranged into specific clusters. Thus, fusion quality can be improved by reducing the weights of data in these clusters that contain error or useless data. Experiments with simulated datasets showed its better fusion accuracy compared to traditional weighted fuzzy logic algorithms and other two fusion models. Theoretically, the method achieves better fusion efficiency and quality, and is also robust facing with noises. It is a pity that this method was not evaluated with real-world data sets.

Along with the great development of the Internet and e-commerce, online shopping becomes more and more popular. Consumers need to acquire as much information as possible about the products they are interested in, including opinions of other consumers. Yan et al. [35] proposed an algorithm for reputation generation and recommendation provision based on opinion mining and fusion. Opinions are firstly filtered to eliminate unrelated or spam opinions. Then, similar opinions are fused and clustered into a specific opinion set. A number of opinion clusters are then generated. In addition, the voting or cited opinions of original opinions are also properly fused into main opinion clusters. The scale of raw data set is greatly reduced for generating a reputation value with high efficiency. Experimental results based on real-world data from both Chinese and English Amazon websites show the accuracy (quality) and stability of the algorithm. The authors also discussed the generality of the algorithm by indicating that it can be applied to generate reputation of many different entities through opinion fusion. It can also support such ways that people express their attitudes as votes and comments in nature languages. Thus, this fusion model performs well in terms of extensibility.

Alyannezhadi et al. [37] proposed a data fusion algorithm based on clustering for uncertainty systems. The systems with a number of unidentified characteristics or mathematical models are usually called unknown systems. In this case, we do not know explicit patterns of the system, which would make researchers fall into trouble in processing data. In [37], a data fusion algorithm was proposed, which contains three parts including clustering, prediction and updating. In the clustering part, subsets of raw

data are generated and then a multi-layer perceptron (MLP) is trained with data to optimize its prediction ability. It is worth noting that the data in training sets are timely. At last, fusion results are updated in the whole system. In unknown systems, a prominent problem is data inconsistency and uncertainty, which is also the main problem solved by this model. Experiment results with real data sets of temperature from five Internet companies show the elimination of data inconsistency and also the robustness of the algorithm. This algorithm is also possible to be applied into other known or unknown multi-sensor data fusion scenarios, which shows its potential of extensibility. Efficiency, Stability and Privacy were not mentioned in the paper.

4.2 Feature Level Data Fusion

In feature level data fusion, data inputs can be either data or features extracted already. As an output, we can obtain refined characteristics or features in the form of other patterns that can be applied to other targets, or data in a higher level, i.e., decisions. Information derived from this process is more polished and comprehensive to show various characteristics of data compared with the signal level data fusion. In what follows, we review the recent advances about feature level data fusion.

Feature Level Data Fusion Based on Supervised Learning

SVM performs well in feature fusion [44]. Pouteau et al. proposed an SVM-based selective fusion algorithm for solving a land cover classification problem [44]. The authors compared a variety of previous fusion models in this field and stated that SVM acquires the best performance because of its ability for processing the data from both mono-source and multi-source. Most simplex multi-source fusion models applied in remote sensing may face deteriorative accuracy in some scenarios with classes utilized by a non-relevant source. On the contrary, selective SVM can deal with it with the integration of mono-source classification and multi-source fusion. Experiments with real data sets showed the effectiveness and stability of the algorithm [44]. What's more, it is not limited to be applied in tropical rainforest classification, as tested in this paper. It is applicable in solving other remote sensing problems with multi-sensory and Geographic Information System (GIS) data, which implies good extensibility of the algorithm. However, other criteria were not discussed in this work.

Starzacher and Rinner proposed an embedded real time multisensory data fusion scheme based on ANN, SVM and NBC (Naïve Bayes Classifiers) [43]. In an embedded real-time environment, there are not affluent resources in each data processing node. However, there is a strict requirement on processing time of an applied fusion algorithm because of the high speed and instantaneity of data scream. The embedded multi-sensor fusion system proposed in [43] includes several sensor nodes distributed in three layers, a single center node and an assisted sensor node to help a single node make decisions. Three fusion methods were tested in an embedded test platform with four real-world datasets. Classification execution time and classification rates were used to measure the performance of models. Experiments result showed that SVM has the least classification time and the three algorithms all perform better than the classical methods. On the other hand, classification rates are influenced by many reasons. As a whole, these three fusion methods perform well in the embedded system with reasonable performance.

Ranking SVM, which transforms a learn-to-rank problem into a formalized binary classification solved by SVM, has become a hot topic nowadays. Cao et al. [42]

employed the ranking SVM into a meta-search engine based on fusion. The meta-search engine in this paper is a cross-media engine, which approves both text-based retrieval and content-based retrieval. The meta-search engine is expected to have the ability of distributing the requests from users to several member search engines and then merging results into a whole list. The key point in this engine we pay attention to is "result fusion", which integrates the results from all member search engines and figure out a comprehensive rank list. Common literatures in this field often give different engines a common weight by ignoring the specific condition and performance of each single member search system. This paper solved this problem with the help of supervised learning to obtain appropriate fusion weights. The ranking SVM model transforms the ranking problem into a binary classification problem by modifying a function form. For a document from the result sets, the algorithm firstly selects features and builds training sets based on users" orders. Then the constraint relationships and a linear final merged function help in training weights of the features. The final score of a new testing set is also computed by the function above. In simulations, the authors used many parameters to evaluate the precision of the fusion model. Results showed that the ranking SVM obtains higher accuracy and better performance than other methods in terms of all assessment measurements. The efficiency of the model was not mentioned in this paper, which requests further study. The experiments were conducted based on a large amount of data from WikipediaMM2008 database.

A typical Artificial Neural Network-based sensor fusion method was developed in an online tool wear estimation environment [27]. In a manufacturing process, a monitoring tool wear plays an important role to avoid degradation of product quality caused by serious tool wear. Great demand on online tool wear estimation leads to the research of data fusion. This paper provides a classical neural network-based fusion model including data preprocessing, feature extraction and feature fusion. Training data sets and testing data sets with tool wear condition obtained from optical microscope were used to train the neural networks offline. Thus, the system can provide tool wear estimation as soon as the features of the tool are given online. Different feature groups were extracted and tested in order to assess and acquire the best estimation result. Tests based on training data sets generated from both laboratory and an industrial environment with different noise levels showed the practicability and effectiveness of the system. Therefore, this method is robust and effective.

Feature Level Data Fusion Based on Unsupervised Learning

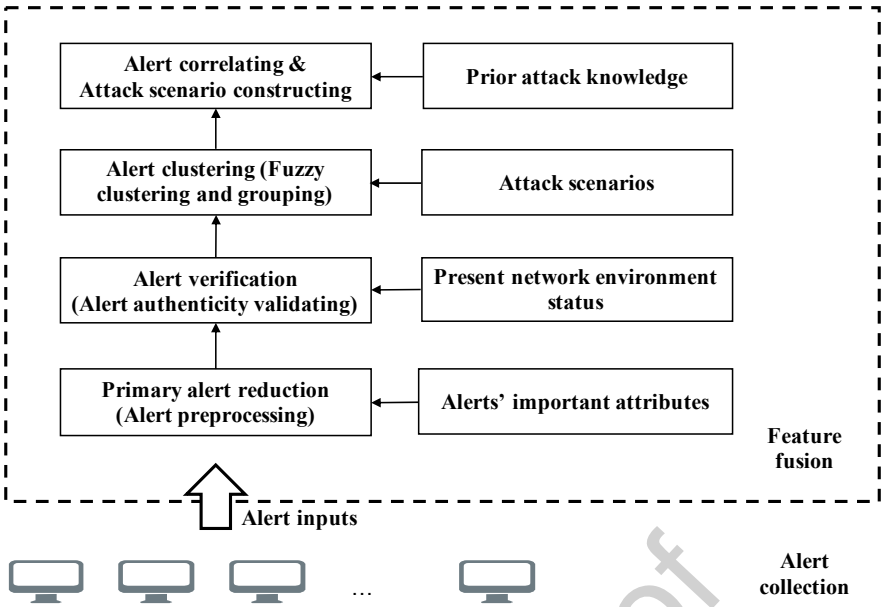

Figure 5. Alert fusion model of [25]

Intrusion Detection Systems (IDS) discriminate attacks and maintain system stability. However, many alerts detected by IDSs have many kinds of problems. They are in large scales or inferior quality, which consumes many system resources and takes long time to deal with. In some conditions, up to 99% of alerts detected by IDSs are false or repetitive. To resolve these problems, there are many models provided. Xiao et al. [25] proposed a hierarchical fusion system with four fusion layers to process alerts. Figure 5 shows the architecture of the alert fusion model. After alert pretreatment, data sets first come into primary alert reduction. This module compares some important attributes, such as protocol type, source IP, target IP, and so on, of different alerts arrived during a temporal window. When all attributes are same in the two alerts, which means the alerts are repetitive, these two alerts should be firstly combined. Then, alert verification module is responsible for validating authenticity of alerts and eliminating false alerts. Alert verification compares alerts regarding both the information of alert itself and its target machine in order to achieve high fusion quality. This module periodically scans the protected network environment for gaining high efficiency. Many false alerts and irrespective ones are eliminated with this way and the burden of services can also be reduced. Next, fuzzy clustering methods are used to classify alerts, which mainly groups the alerts based on attack scenarios. The model groups the alerts into clusters with their target IP. Alerts with the same target IP are clustered into one group. Then, the fuzzy similarity matrix of each group is generated. At last, the alerts are divided by the fuzzy clustering model with the help of an appropriate threshold. Based on attack knowledge, alerts in the same class are correlated and attack scenarios are constructed then. Experiments based on two test data sets showed the performance in redundant alerts reduction, so the quality of the system is good. Efficiency, Robustness and Privacy were not mentioned in the paper. Tests over two real world datasets showed the stability of the system. What's more, the system can work with any effective alert detection methods, so it also has good extensibility.

4.3 Decision Level Data Fusion

In order to further fuse some information that has already been generated to reveal some decisions of a task, we come to the highest level - decision level data fusion. We need not only the decision derived from single perspective, but also the one with a global view. Thus, decision level fusion often appears right before final decisions are made. Compared to low-level fusion, decision fusion methods often generate a preliminary classification and can fuse different types of data and obtain accurate fusion results.

Decision Level Data Fusion Based on Supervised Learning

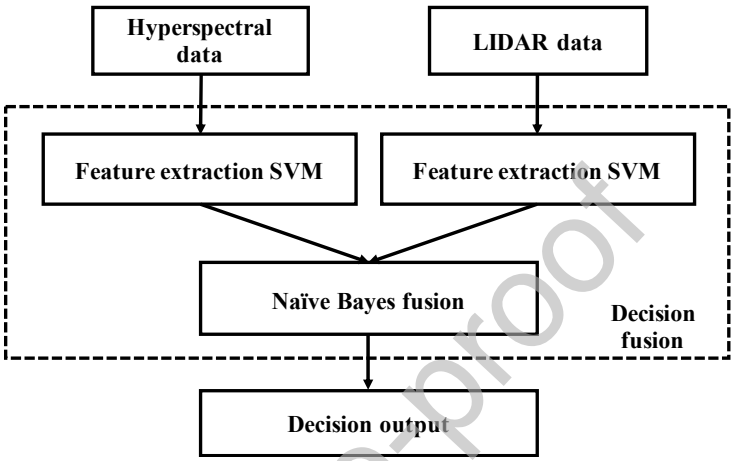

Figure 6. Flowchart of the proposed fusion method in [45]

Bigdeli et al. [45] proposed a typical decision fusion model based on multiple SVM and Naïve Bayes. Fusion of light detection and ranging (LIDAR) and hyperspectral data was discussed in the field of remote sensing data from multiple sensors. Figure 6 shows the classifier fusion system proposed in this paper. Firstly, a set of features, which contain valuable information to distinguish objectives in the next steps, are extracted from LIDAR data and hyperspectral data, respectively. After that, a one-against-one multi-class SVM method based on radial basis function (RBF) kernel is utilized to classify the features captured in the previous phase. SVM classifiers are used in each feature space. At last, a classical fusion method, Naïve Bayes model fuses data sets from single classifiers. The authors used overall accuracy and kappa coefficient as the evaluation metrics of model performance. The proposed model shows better results than the usage of original LIDAR, hyperspectral data or any other simple integrated models of these two kinds of data. Experimental results based on the data sets captured around the University of Houston campus from an official mapping organization showed that this method effectively maximizes the advantage attributes of LIDAR and hyperspectral through feature extraction, feature classification and decision fusion. A detailed fusion performance comparison and evaluation analysis were given in this paper, which is a marked advantage. However, Efficiently, Robustness, Extensibility, Privacy and Stability were not mentioned in this paper.

As a preliminary version of [47], Giorgio et al. provided a similar system of anomaly-based intrusion detection in [48]. It has multiple classifiers. Features in each traffic connection and data packet are subdivided into three groups -- intrinsic features, traffic features and content features based on the characteristics of the feature. Each

feature set maps with a corresponding classifier. Feature sets can mostly describe normal and abnormal network patterns so that the classifiers can distinguish attack pattern by training with a large group of given data sets. The authors implemented a three-layer neural network as classifier and applied five different fusion rules to verify system effectiveness. Results showed that the Multiple Classifier System provides a trade-off between detection rate, false alarm rate and generalization abilities compared to the approach with individual classifier that deals with all extracted features. In addition, A-Posteriori DCS fusion technique can provide the best overall performance in terms of false alarms rate, error rate and average cost.

Decision Level Data Fusion Based on Unsupervised Learning

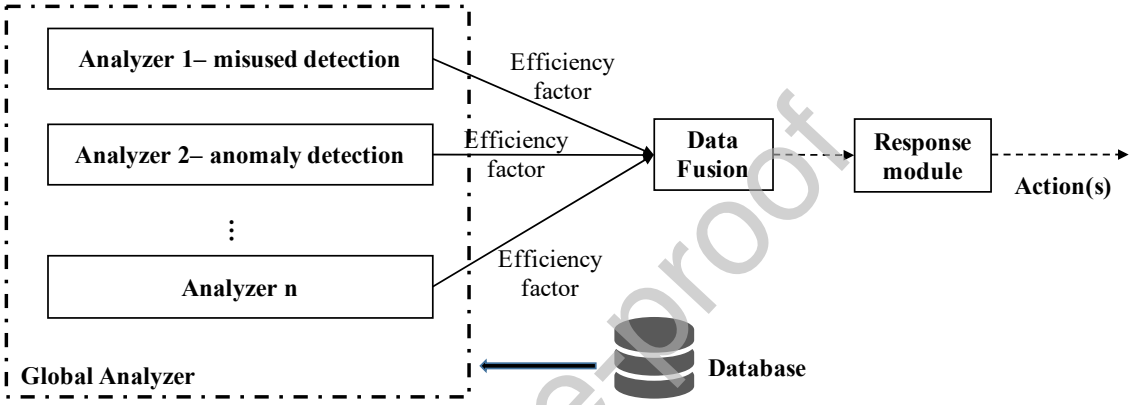

Figure 7. Fusion architecture discussed in [46]

Fessi et al. [46] proposed a data fusion model based on clustering for intrusion detection to resolve the weakness of some existing literatures on clustering, such as the lack of ability in detecting composite attacks and constructive attacks, the ignoring of efficiency and overmuch of human intervention. The architecture of the intrusion detection system is described in Figure 7. It is a centralized system that contains sensors as observers to detect data sets, a global analyzer containing a data fusion component, a response module for activating actions and database. A number of analyzers inside the global analyzer are set to detect different events about attacks with different methods based on misuse detection or anomaly detection. An efficiency factor of each analyzer is used to evaluate its accuracy, performance and robustness. Some partial decisions are made by a number of analyzers and then are sent to the fusion component for gaining a global security view of the whole system. Both the events sniffed by the analyzers and the efficiency factor of each analyzer are taken into account by the clustering operation. A data fusion clustering model partitions events from the analyzers into new clusters based on attack behaviors. As a whole, the adaptivity of this model, which is mainly realized by the settings of analyzers, in different attack scenes and composite attacks is remarkable. The author illustrated the function of proposed algorithm with an example, but they did not set any simulations or experiments to prove its performance. What's more, other properties, such as Robustness, Stability and tests based on real world data sets were not mentioned.

Intrusion detection systems fall into two main categories, anomaly-based IDS and signature-based IDS. The anomaly-based IDSs model the normal network and malicious

behaviors that can be detected with different features compared to the normal model. An important advantage of anomaly-based IDS is its ability of detecting unknown intrusions, but its high false alarm rate cannot be ignored. Giacinto et al. [47] proposed a multiple modular system with a one-class classifier to implement anomaly-based detection. The authors divided network connections into groups on account of the service of each connection. In other words, each group describes a set of similar packets in view of "service". Three one-classifier algorithms were applied to realize the classification. In each group, extracted features are classified and compared with the normal model, and then decisions from classifiers will be generated and fused into an overall conclusion. Another peculiarity of this system is that it subdivides false alarm rate into distributed modules. Thus, people can adjust the threshold of the similarity in the detection, which affects the detection rate further. Experimental results showed that the multiple modular system can provide higher detection rate than a single classifier that deals with whole features. The experiments were based on dataset DARPA 1998, which is a popular real data set. However, Privacy was not discussed in this work although there is a strong need to protect security and privacy of the data used in intrusion detection.

Clustering is also used for decision fusion in the last step of fusion process [49]. Chen et al. proposed a deep learning-based nuclear power crack detection algorithm [49]. Nuclear power crack inspection is an important component of nuclear applications in case of incidents. Some vision-based crack detection algorithms were proposed, but there are still open issues in tiny cracks and noisy patterns detection. This paper solved this problem with a Naïve Bayes and clustering-based fusion model. With the former modules" crack detection results aggregated in tubelets, Naïve Bayes discards false positive tubelets and the clustering model groups the tubelets for a whole crack with Euclidean distance in order to make final decision. This algorithm was tested with real crack datasets. Experiments showed its improved effectiveness compared with the past methods. Thus, it has sound Quality and Stability. With the outstanding advantage in detecting robust and noisy patterns, this algorithm performs quite well regarding Robustness. Due to the computational consumption of some parts in the algorithm, Efficiency is not ideal, which becomes a part of future work. Other criteria were not concluded in this paper.

4.4 Comparison and Discussion

In Section 4, we comprehensively review the existing works about machine learning for data fusion. To conclude, we compare all the models/methods/algorithms involved in this section in Table 1 with regard to their fusion types, application scenarios, applied machine learning methods, main challenges to overcome, and satisfactory with the proposed criteria. The notations used to evaluate the performance of data fusion are introduced below.

- **Efficiency (Ef)**

-Yes (Y): The algorithm provides highly efficient data fusion or there are discussions on fusion efficiency in experiments and evaluation.

-No (N): The algorithm does not promote efficiency or efficiency was not discussed.

- **Quality (Q)**

-High (H): The algorithm improves quality as the main concern and provides detailed evaluation to prove its effectiveness or enough experiment results to show good data

fusion quality.

-Low (L): The algorithm intends to deal with low fusion quality. Nevertheless, performance analysis is too rough or experiment results are not adequate. Alternatively, there is no significant performance gain.

-No (N): Fusion quality was not discussed or not obviously promoted.

- **Stability (St)**

-Yes (Y): The algorithm performs well in a stable way, which is supported with experimental results.

-No (N): The algorithm is not stable or this property was not concerned in the paper.

- **Robustness (R)**

-Yes (Y): The algorithm performs well in a fluctuant environment with the support of experimental results, or robustness was only theoretically discussed.

-No (N): The algorithm is not robust or this property was not concerned in the paper.

- **Extensibility (Ex)**

-Yes (Y): The algorithm can be applied into other application scenarios theoretically or illustrated with experiments.

-No (N): This property was not concerned.

- **Privacy (P)**

-Yes (Y): The algorithm can ensure data security in data fusion, data privacy was taken into consideration, or this problem was concerned theoretically.

-No (N): This problem was not considered in study.

- **Tested with real world data sets (Re)**

-Yes (Y): The proposed model was tested with the data sets captured from real world environments or in practice.

-No (N): The data sets used in experiments were all simulated or authors did not talk about the sources of data sets or there are no any data-based experiments provided at all.

Table 1. Summary and Comparison of Machine Learning Methods for Data Fusion

References	Fusion types	Application scenarios	Machine learning methods	Challenges to overcome	Ef	Q	St	R	Ex	P	Re
[28]	signal	Motor fault detection	SVM	Dynamic fusion	N	H	N	N	Y	N	Y
[29]	signal	Distributed data fusion	SVM	Fusion location	Y	L	N	N	Y	N	N
[30]	signal	Biometric Fusion	SVM	Data imperfection	Y	H	Y	Y	N	N	Y
[40]	signal	WSN	BP neural network	Data type	Y	N	Y	N	N	N	N
[41]	signal	WSN	SMPSO-BP neural network	Data imperfection	Y	N	Y	N	Y	N	N
[32]	signal	Navigation system	Elman neural network	Data imperfection	Y	H	N	N	N	N	Y
[34]	signal	Drum level measurement	RBF neural network	Data imperfection	Y	H	Y	Y	N	N	N
[31]	signal	Multi-objectives real-time tracking	k-central clustering	Data association	N	H	N	Y	N	N	N
[39]	signal	High resolution radar system	Clustering	Dynamic fusion	Y	N	Y	Y	N	N	Y
[51]	signal	Radar data fusion	Cell clustering	Data imperfection	N	L	N	Y	N	N	N
[36]	signal	Multi-target tracking	K-means	Data imperfection	N	H	Y	N	N	N	Y
[26]	signal	WSN anomaly detection	K-Means	Fusion Location	Y	H	Y	Y	N	N	Y
[33]	signal	WSN	Un-even clustering/Simulated annealing algorithm	Data imperfection	Y	N	N	Y	Y	N	N
[38]	signal	WSN	K-means	Data imperfection	Y	H	Y	N	N	N	N
[35]	signal	Reputation generation	Clustering	Data imperfection	Y	H	Y	N	Y	N	Y

[37]	signal	Unknown system	Clustering/MLP	Data inconsistency	N	H	N	Y	Y	N	Y
[44]	feature	Land cover classification	SVM	Data imperfection	N	H	Y	N	Y	N	Y
[43]	feature	Embedded real-time fusion	ANN	Data imperfection	Y	H	Y	N	Y	N	Y
[43]	feature	Embedded real-time fusion	SVM	Data imperfection	Y	H	Y	N	Y	N	Y
[43]	feature	Embedded real-time fusion	NBC	Data imperfection	Y	H	Y	N	Y	N	Y
[42]	feature	Meta-search engine	Ranking SVM	Data imperfection	N	H	Y	N	N	N	Y
[27]	feature	Tool wear estimation	Artificial Neural Network	Dynamic fusion	Y	N	N	Y	N	N	Y
[50]	feature	Gesture Recognition	SVM	Data imperfection-Classifier	Y	H	Y	N	N	N	N
[52]	feature	Moving target indication	Self-organizing clustering	Data imperfection	N	L	N	N	Y	N	N
[25]	feature	Intrusion detection	Fuzzy clustering	Data imperfection	N	H	Y	N	Y	N	Y
[45]	decision	Remote sensing data fusion	SVM	Data imperfection	N	H	N	N	N	N	Y
[48]	decision	Intrusion detection	Neural Network	Data imperfection	Y	H	Y	N	N	N	Y
[46]	decision	Intrusion detection	Clustering	Data imperfection	N	L	N	N	Y	N	N
[47]	decision	Intrusion detection	K-Means & v-SVC	Data imperfection	N	H	Y	N	N	N	Y
[49]	decision	Nuclear power crack detection	Clustering	Data imperfection	N	H	Y	Y	N	N	Y

Ef: Efficiency, Q: Quality, R: Robustness, St: Stability, Ex: Extensibility, P: Privacy, Re: Tested with real world data sets.
Y: Concluded or did well or discussed theoretically; N: Not mentioned.
H: Did well especially in Q; L: Concluded but analysis was not adequate in terms of Q.

Based on Table Ⅰ, we summarize our review as below.

Among all the studies reviewed in this section, the methods of signal level data fusion are distinctly overwhelming with nearly half of all the reviewed papers [26, 28-41, 51]. Some works fused features extracted from raw data to acquire better fusion quality [25, 27, 42-44, 50, 52]. In [45-49], researchers extracted information and fused decisions in a high level.

During survey, we observe that the application environment of data fusion with machine learning are in variety. Representative fusion scenarios include but not limited to WSN systems [29, 33, 38, 40, 41], radar tracking and remote systems [31, 36, 45, 51, 52], intrusion detection [25, 46-48], reputation generation [35], mechanical engineering scenarios [27-28, 34], and so on. More and more machine learning-based fusion is needed in all kinds of fields. Most of the reviewed works solved the “data imperfection” problem in data fusion. Beyond that, some works applied in distributed systems and WSN figure out location fusion problem with SVM and K-Means [29, 38]. We hold such an opinion that the machine learning methods cannot solve all challenges of data fusion, such as data confliction due to the limitation caused by its nature.

Data fusion models are based on many typical machine learning methods. Supervised learning methods such as SVM [28, 29, 30, 42-44, 45, 47, 50] and NN [27, 32, 34, 40, 48] were widely applied. Correspondingly, clustering models [17, 19, 23, 46, 48, 52, 66, 68] and K-Means [26, 36, 38, 47] were also adopted to improve fusion effectiveness and performance. SVM is good at dealing with data with high dimensions, while NN is more adept at learning from imperfect and uncertain data or when a system is difficult to be described with a linear formula. There is no direct relationship between fusion types and machine learning methods. Usually machine learning methods are good at handling classifying problem during fusion process.

Many of the data fusion models treat fusion quality as the most important requirement without any discussion on fusion efficiency [25, 28, 31, 36-37, 42, 44-47,

49, 51-52]. In the models that mainly concern about fusion quality, most literatures provided expatiation about performance evaluation to exhibit their significant improvement on quality. Experiments were usually performed to show the advantages of the proposed models by comparing them with the results of other previous models. However, fusion efficiency was paid little attention. In some existing signal level fusion models, efficiency was discussed in distributed fusion applications [29, 33].

We also find that most fusion models perform well in terms of stability, which shows their strong ability of steady operation in actual applications. However, few existing works concerned about robustness and extensibility, and few experiments testified the performance of data fusion on these two aspects. In addition, few existing literatures considered the security of training sets, even in the field of intrusion detection. Security and privacy issues request urgent investigation in some specific data fusion fields. We also note that many existing works only focus on achieving a single research objective without comprehensively fulfilling all performance requirements and criteria.

Besides, more than half of the reviewed works evaluated the performance of their proposed models with data sets captured from real application environments. However, some experiments were conducted in simulated environments due to multiple reasons and difficulties of testing in practice. Only few works researched their models in reality and most of them are related to computer science. Some works even did not expound the source of data they used for experiments.

5. Open Issues and Future Research Directions

Based on the detailed survey reported in Section 4, we further indicate a number of open issues and suggest some future research directions.

5.1 Open Issues

First, the machine learning methods used for data fusion are simplex. As we discussed in Section 4.4, most of machine learning models mentioned for data fusion are based on SVM, clustering and neural networks, which are classical methods and simple neural networks. SVM and clustering methods often aim at classifying with high accuracy. NN is suitable for describing uncertain complex systems. Nevertheless, the power of machine learning methods should be far more than this. Taking one example, deep learning is considered as a significant research field in artificial intelligence in next 10 years. Deep learning describes the techniques that simulate complex neural systems of humans. Compared with simple neural networks, more hidden layers inserted into the network would give the system better accuracy and learning quality. The lack of deep learning methods for data fusion motivate us to explore new thoughts.

Second, researchers pay little attention to fusion efficiency. Refer to Table 1, past work focuses more on fusion quality than fusion efficiency. Some works even did not discuss or evaluate this important property at all. The most obvious disadvantage of machine learning methods is its computational complexity and huge consumption of computing and system resources. Machine learning often needs large sets of data for training, which also brings difficulty into actual applications. Since there will be a good deal for specific needs of miniature devices in the future, which are not affordable for complicated computation due to limited resources, the study for optimizing the efficiency of data fusion models becomes necessary.

Third, comprehensive concern of data fusion is missed. Based on Table 1, few literatures discussed Robustness and Extensibility. Some literatures did not testify if

their models are stable in an unsteady environment with experimental results. These requirements should be fundamental for a fusion model. Some works consider little about the models‟ effectiveness in practical use. Taking Robustness as an example, data with serious imprecision, inconsistency and noises often occurred, a model that cannot handle this circumstance well will be practically limited. A similar argument is put on Extensibility. Simply improving data fusion accuracy and quality, but ignoring other properties will lead to an imperfect model, while a comprehensive model that satisfies all expected criteria should be urgently studied.

Finally, few existing literatures take account of data privacy and security. Machine learning methods have a great need to deal with a large scale of data sets to ensure learning quality and fusion accuracy. However, using original data in machine learning could cause sensitive information leakage. This problem can be particularly acute in the Internet related applications such as intrusion detection, attack analysis, and location tracking. Private information about identities and positions of data providers could be disclosed if the proposed model cannot manage it well.

5.2 Future Research Directions

Based on the above indicated open issues, we move up to propose some potential future research directions.

First direction is to explore more application scenarios for machine learning based data fusion. After the great development of machine learning for data fusion in decades, it is gratifying to see a wide range of models applied into different scenarios, such as intrusion detection, target identification and tracking for military and nonmilitary utilization, human-computer interaction, navigation and geographic utilization, and so on. What‟s more, there are many other application scenarios that are expected to use machine learning based data fusion methods. The strong ability of machine learning in nonlinear mapping provides additional opportunities for data fusion. Supervised learning models represented by SVM and Random Forest do well with high dimensional data and their flexibility makes them suitable for solving more problems. ANN models are especially good at modeling multifarious nonlinear networks that are difficult to describe with functions directly. With a growing demand of IoT and smart devices, there are more industrial fields with numerous data sets that can be promoted by applying machine learning based data fusion methods.

Another future research direction is the use of more complex and large-scale learning techniques into data fusion. As talked above, we place expectations on deep learning, which combines supervised learning and unsupervised learning to construct learning hierarchy, namely the network. Especially in some scenarios that relate to a large amount of data, Deep learning can gain much more improved performance and prediction precision than past learning algorithms. According to [4], there have been some efficient models appeared to deal with fusion problems with deep learning. In [53], a deep belief network based data fusion scheme was proposed for ball screw fault detection. Nevertheless, there might be some following challenges introduced at the same time. The effectiveness of deep learning can only be ensured with mass data and high resource consumption. How to ensure the applicability of deep learning based fusion models in small devices and how to make trade-off between fusion efficiency and quality are additional issues that should be solved. Except for the issues mentioned above, we are also looking forward to researches on deep composite intelligent applications.

There is also a serious security need on fusion models. Information privacy is in urgent need to be protected in both fusion process and machine learning process. Experiments involved in the above reviewed works are mostly performed with testing data sets. It will be extremely dangerous if transplanting the model into actual utilization directly because of the exposure of all data sets. Without any security protection, sensitive information can be recovered and acquired from fusion results. Besides, a central device that preforms fusion might become vulnerable facing to attacks. We need fallback or other solutions for model's better practicability. Trustworthy data fusion with security and privacy protection is highly required to be ensured.

At last, data fusion model performance evaluation should be more well-founded. As mentioned in Section 4.4, there are some works that did not testify their model with real data sets. Laere [54] studied the state and difficulties of information fusion performance evaluation in reality. The author took an overview of 52 data fusion publications, only 6% works evaluate the model in real scenarios. Laere also explained the difficulties, which impede data fusion research, and gave suggestions. Further research should improve the quality of model performance evaluation. A more holistic model evaluation should be conducted to prove the effectiveness of data fusion based on machine learning.

6. Conclusions

This paper has made a comprehensive review on the literature about machine learning for data fusion. We first provided basic background knowledge about data fusion and machine learning. We further proposed a number of criteria to evaluate the works reviewed in this paper for the purpose of commenting their pros and cons remarkably. We carefully reviewed the recent literature based on the level of fusion taken apart in and the type of machine learning, and then used a table to summarize our main review results. On the basis of our survey, we went ahead to specify a number of open issues and proposed some future research directions that deserve further investigation. This study provides a concise and comprehensive reference for researchers and practitioners in the field of machine learning for data fusion.

Acknowledgments

This work is sponsored by the NSFC (grants 61672410, 61802293 and U1536202), Academy of Finland (grants 308087 and 314203), National Postdoctoral Program for Innovative Talents (grant BX20180238), the Project funded by China Postdoctoral Science Foundation (grant 2018M633461), the Fundamental Research Funds for the Central Universities (grant JB191504), and the 111 project (grants B16037).

Conflict of Interest

We claim that we have no conflict of interest with other researchers with regard to the paper:

"A Survey on Machine Learning for Data Fusion"

submitted to Information Fusion.

Paper authors:

Tong Meng, Xuyang Jing, Zheng Yan, Witold Pedrycz

References

[1] D. L. Hall and J. Llinas, An introduction to multisensor data fusion, Proceedings of the IEEE, 85 (1) (1997) 6-23.
[2] C. Federico, A review of data fusion techniques, The Scientific World Journal, (2013) 1-19.
[3] A. L. Samuel, Some studies in machine learning using the game of checkers. I, Computer Games I, (1988) 335–365.
[4] F. Alam, R. Mehmood, I. Katib, N. N. Albogami and A. Albeshri, Data fusion and IoT for smart ubiquitous environments: a survey, IEEE Access, 5 (2018) 9533-9554.
[5] S. Gite and H. Agrawal, On context awareness for multisensor data fusion in IoT, Springer India, 381 (2016) 85-93.
[6] I. M. Pires, N. M. Garcia, N. Pombo, and F. Flórez-Revuelta, From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices, Sensors, 16 (2) (2016) 184.
[7] G. Navarro-Arribas and V. Torra, Information fusion in data privacy: a survey, Information Fusion, 13 (4) (2012) 235-244.
[8] N. Faouzi, H. Leung and A. Kurian, Data fusion in intelligent transportation systems: progress and challenges – A survey, Information Fusion, 12 (1) (2012) 4-10.
[9] I. Corona, G. Giacinto, C. Mazzariello, F. Roli and C. Sansone, Information fusion for computer security: state of the art and open issues, Information Fusion, 10 (4) (2009) 274-284.
[10] J. Yao, V. Raghavan and Z. Wu, Web information fusion: a review of the state of the art, Information Fusion, 9 (4) (2008) 446-449.
[11] S. Liao, et al. Data mining techniques and applications – a decade review from 2000 to 2011, Expert Systems with Applications, 39 (12) (2012).
[12] C. Rudin and K. L. Wagstaff, Machine learning for science and society, Machine Learning, 95 (1) (2014) 1-9.
[13] J. Qiu et al., A survey of machine learning for big data processing, EURASIP Journal on Advances in Signal Processing, 2016 (1) (2016).
[14] Q. Zhang et al., A survey on deep learning for big data, Information Fusion, 42 (2018) 146-157.
[15] F.E. White, Data Fusion Lexicon, (1991).
[16] Z. Yan, J. Liu, L. T. Yang, W. Pedrycz, Data Fusion in Heterogeneous Networks, Information Fusion, 53 (2020) 1-3.
[17] C.L. Bowman and M.S. Murphy, Description of the VERAC NSource tracker/correlator, Naval Res Lab. Report R-01O-80, (1980).
[18] C.L. Bowman and C.L. Morefield, Multisensor fusion of target attributes and kinematics, in Proceedings of Decision and Control including the Symposium on Adaptive Processes, 1980 19th IEEE Conference on IEEE, 1981.
[19] R. Luo and M. Kay, Multisensor integration and fusion: issues and approaches, SPIE Sensor Fusion, 931 (1988) 42-49.

[20] B. Dasarathy, Sensor fusion potential exploitation-innovative architectures and illustrative applications, Proceedings of IEEE, 85 (1) (1997) 24-38.
[21] Steinberg, Alan N, C. L. Bowman, and F. E. White, Revisions to the JDL data fusion model, Proceedings of SPIE - The International Society for Optical Engineering, 3719 (1999) 430-441.
[22] S. Ayed, H. Trichili, and A. M. Alimi, Data fusion architectures: a survey and comparison, in Proceedings of International Conference on Intelligent Systems Design and Applications, 2016, pp. 277-282.
[23] X. Jing, Z. Yan, and P. Witold, security data collection and data analytics in the Internet: a survey. IEEE Communications Surveys & Tutorials, 21 (1) (2018) 586-618.
[24] B. Khaleghi, et al., Multisensor data fusion: a review of the state-of-the-art, Information Fusion, 14 (1) (2013) 28-44.
[25] S. Xiao, Y. Zhang, X. Liu, and J. Gao, Alert fusion based on cluster and correlation analysis, in Proceedings of International Conference on Convergence and Hybrid Information Technology, 2008, pp. 163-168.
[26] X. Guo, D. Wang and F. Chen, An anomaly detection based on data fusion algorithm in wireless sensor networks, International Journal of Distributed Sensor Networks, 2015 (2015) 1-10.
[27] N. Ghosh et al., Estimation of tool wear during CNC milling using neural network-based sensor fusion, Mechanical Systems & Signal Processing, 21 (1) (2017) 466-479.
[28] T.P. Banerjee and S. Das, Multi-sensor data fusion using support vector machine for motor fault detection, Information Sciences, 217 (24) (2012) 96-107.
[29] S. Challa, M. Palaniswami and A. Shilton, Distributed data fusion using support vector machines, International Conference on Information Fusion, 2 (6) (2013) 881-885.
[30] M. S. Fahmy et al., Biometric fusion using enhanced SVM classification, in Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing IEEE, 2008, pp. 1043-1048.
[31] H. Shu, Y. Wang and J. Jiang, Multi-radar data fusion algorithm based on K-central clustering, in Proceedings of International Conference on Fuzzy Systems and Knowledge Discovery, 2007, pp. 617-621.
[32] K. Kolanowski, A. Swietlika, R. Kapela, J. Pochmara and A. Rybarczyk, Multisensor data fusion using Elman neural networks, Applied Mathematics & Computation, 319 (2017) 236-244.
[33] L. Xiao and Q. Liu, A data fusion using un-even clustering for WSN, in Proceedings of International Conference on Advanced Intelligence and Awareness Internet, 2012, pp. 216-219.
[34] W. Tong, B. Li, X. Jin, Y. Yang and Q. Zhang, A study on model of multisensor information fusion and its application, in Proceedings of International Conference on Machine Learning and Cybernetics, 2006, pp. 3073-3077.
[35] Z. Yan, X. Jing, and W. Pedrycz, Fusing and mining opinions for reputation generation, Information Fusion, 36 (2017) 172-184.
[36] H. Wang et al., An algorithm based on hierarchical clustering for multi-target tracking of multi-sensor data fusion, In Proceedings of Control Conference. IEEE, 2016, pp. 5106-5111.
[37] M. M. Alyannezhadi, A. A. Pouyan, and V. Abolghasemi, An efficient algorithm for multisensory data fusion under uncertainty condition, Journal of Electrical Systems & Information Technology, (2016).
[38] F. Wang et al., An improved fusion method of fuzzy logic based on K-means clustering in WSN, Journal of North University of China, 35 (6) (2014) 699-703.
[39] Z. Li and X. Wang, High resolution radar data fusion based on clustering algorithm, in Proceedings of IEEE International Workshop on Database Technology and Applications, 2010, pp. 1-4.
[40] J. Tan and H. Gan, WSN data fusion scheme based on improved BP neural network, Journal of Residuals Science & Technology, 13 (7) (2016).
[41] S. Li et al., WSN data fusion approach based on improved BP algorithm and clustering protocol, in Proceedings of 2015 27th Chinese Control and Decision Conference (CCDC) IEEE, 2015.
[42] Y. Cao, T. J. Huang and Y. H. Tian, A ranking SVM based fusion model for cross-media meta-search engine, Frontiers of Information Technology & Electronic Engineering, 11 (11) (2011) 903-910.
[43] A. Starzacher and B. Rinner, Embedded realtime feature fusion based on ANN, SVM and NBC, in Proceedings of IEEE International Conference on Information Fusion, 2009, pp. 482-489.
[44] R. Pouteau and S. Benoît, SVM selective fusion (self) for multi-source classification of structurally complex tropical rainforest, IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 5 (4) (2012) 1203-1212.
[45] B. Bigdeli, F. Samadzadegan and P. Reinartz, A decision fusion method based on multiple support vector machine system for fusion of hyperspectral and LIDAR data, International Journal of Image & Data Fusion, 5 (3) (2014) 196-209.

[46] B.A. Fessi, S. BenAbdallah, Y. Djemaiel and N. Boudriga, A clustering data fusion method for intrusion detection system, in Proceedings of 11th IEEE International Conference on Computer and Information Technology, 2011, pp. 539-545.
[47] G. Giacinto, R. Perdisc, Del Rio M and F. Roli, Intrusion detection in computer networks by a modular ensemble of one-class classifiers, Information Fusion, 9 (1) (2008) 69-82.
[48] G. Giorgio, F. Roli and L. Didaci, Fusion of multiple classifiers for intrusion detection in computer networks, Pattern Recognition Letters, 24 (12) (2003) 1795-1803.
[49] C. Fu and R. Mohammad, NB-CNN: deep learning-based crack detection using convolutional neural network and naïve bayes data fusion, IEEE Transactions on Industrial Electronics, 65 (5) (2018) 4392-4400.
[50] Z. He, Accelerometer based gesture recognition using fusion features and SVM, Journal of Software, 6 (6) (2011) 1042-1049.
[51] H. Shu, The application of cell-based clustering algorithm dealing with radar data fusion, in Proceedings of 2008 Congress on Image and Signal Processing, 2008.
[52] D. Qiu et al., The study of self-organizing clustering neural networks and applications in data fusion, in Proceedings of IEEE World Congress on Intelligent Control & Automation, 2008.
[53] L. Zhang and H. Gao, A deep learning-based multi-sensor data fusion method for degradation monitoring of ball screws, in Proceedings of Prognostics Syst. Health Manage. Conf. (PHM-Chengdu), 2016, pp. 1–6.
[54] J. Laere, Challenges for IF performance evaluation in practice, in Proceedings of 12th International Conference on Information Fusion, 2009.
[55] D. L. Hall and J. Llinas, An introduction to multisensor data fusion, Proceedings of IEEE, 85 (1) (2002) 6–23.
[56] E. Soltanmohammadi and M. Naraghi-Pour, Context-based unsupervised data fusion for decision making, in Proceedings of International Conference on International Conference on Machine Learning, 2015, pp. 2076-2084.
[57] K. Lin, T. Liu, and H. Ge, A clustering hierarchy based on data fusion in wireless sensor networks, in Proceedings of International Conference on Computational Intelligence & Software Engineering, 2009, pp. 1-4.
[58] L. Snidaro, J. Garcia, and J. Llinas, Context-based information fusion: a survey and discussion, Information Fusion, 25 (2015) 16-31.
[59] R. Nowak, R. Biedrzyck, and J. Misiurewicz, Machine learning methods in data fusion systems, in Proceedings of 19th International Radar Symposium, 2012, pp.400-405.
[60] K. Julisch, Clustering intrusion detection alarms to support root cause analysis, ACM Transactions on Information & System Security, 6 (4) (2013) 443-471.
[61] C. Völker and P. Shokouhi, Data aggregation for improved honeycomb detection in concrete using machine learning–based algorithms, in Proceedings of International Symposium Non-Destructive Testing in Civil Engineering, 2015, pp. 30-47.
[62] SB. Ayed, H. Trichili and AM. Alimi, Data fusion architectures: a survey and comparison, in Proceedings of International Conference on Intelligent Systems Design & Applications, 2016, pp. 277-282.
[63] M. Gheisari and G. Wang, A survey on deep learning in big data, in Proceedings of IEEE International Conference on Computational Science and Engineering, 2017, pp. 173-180.
[64] C. L. Bowman and C. L. Morefield, Multisensor fusion of target attributes and kinematics, in Proceedings of Decision and Control including the Symposium on Adaptive Processes, 1980 19th IEEE Conference on IEEE, 1981.
[65] Gemma Piella, New quality measures for image fusion, in Proceedings of the 7th International Conference on Information Fusion, 2004, pp. 542-546.
[66] Z. Wang, A. C. Bovik, A universal image quality index, IEEE Signal Processing Letters, 9 (3) (2002) 81-84.
[67] L. F. Pau, Sensor data fusion, Journal of Intelligent and Robotic systems, (1988) 103-116.
[68] E. Wilfried, A review on system architectures for sensor fusion applications, Springer, (2007).
[69] Ren C. Luo, C. C. Yih, K. L. Su, Multisensor fusion and integration: approaches, applications, and future research directions, IEEE Sensors Journal, 2 (2) (2002) 107-119.
[70] T. Kanungo, D. M. Mount, et al., An efficient k-means clustering algorithm: analysis and implementation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (7) (2002) 0-892.
[71] J. Matousek, On approximate geometric k-clustering, Discrete & Computational Geometry, 24 (1) (2000) 61-84.

[72] S. S. Liu, L. F. Zhang, Z. Yan, Predict pairwise trust based on machine learning in online social networks: a survey, IEEE Access, 6 (1) (2018) 51297-51318.
[73] L. F. Wei, W. Q. Luo, J. Weng, Y. J. Zhong, X. Q. Zhang, and Z. Yan, Machine learning-based malicious application detection of android, IEEE Access, 5 (1) (2017) 25591-25601.
[74] H. Q. Lin, G. Liu, Z. Yan, Detection of application-layer tunnels with rules and machine learning, The 12th International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (SpaCCS2019), 2019, pp. 441-455.
[75] J. Z. Wang, Z. Yan, L. T. Yang, B. X. Huang, An approach to rank reviews by fusing and mining opinions based on review pertinence, Information Fusion, 23 (2015) 3–15.
[76] W. X. Ding, X. Y. Jing, Z. Yan, L. T. Yang, A survey on data fusion in Internet of Things: towards secure and privacy-preserving fusion, Information Fusion, 51 (2019) 129-144.
[77] X. Y. Jing, Z. Yan, X. Q. Liang, W. Pedrycz, Network traffic fusion and analysis against DDoS flooding attacks with a novel reversible sketch, Information Fusion, 51 (2019) 100-113.
[78] G. Q. Li, Z. Yan, Y. L. Fu, H. L. Chen, Data fusion for network intrusion detection: a review, Security and Communication Networks, (2018) 2018.
[79] Z. Yan, J. Liu, L. T. Yang, N. Chawla, Big data fusion in Internet of Things, Information Fusion, 40 (2018) 32-33.
[80] Z. Yan, J. Liu, A. V. Vasilakos, L. T. Yang, Trustworthy data fusion and mining in Internet of Things, Future Generation Computer Systems, 49 (2015) 45–46.
[81] J. Liu, Z. Yan, L. T. Yang, Fusion - an aide to data mining in Internet of Things, Information Fusion, 23 (2015) 1-2.
[82] X. Y. Jing, J. J. Zhao, Q. H. Zheng, Z. Yan, W. Pedrycz, A reversible sketch-based method for detecting and mitigating amplification attacks, Journal of Network and Computer Applications, 142 (2019) 15-24.