Uncertainty-driven Fusion for Conflictive Multiview Data: Beyond View Alignment Assumptions

Puspamalya Sahoo

Uncertainty-driven Fusion for Conflictive Multiview Data: Beyond View Alignment Assumptions

Minor Project submitted in partial fulfillment of the requirements for the degree of

Master of Technology
in
Cryptology and Security

by

Puspamalya Sahoo
[Roll No: CrS2316]

under the guidance of

Malay Bhattacharyya
Associate Professor
Machine Intelligence Unit, Indian Statistical Institute

Anirban Mukhopadhyay
Professor
Department of Computer Science and Engineering, University Of Kalyani

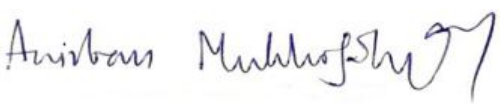
Indian Statistical Institute
Kolkata – 700108, India

July 2025

CERTIFICATE

This is to certify that the Project titled **"Uncertainty-driven Fusion for Conflictive Multiview Data: Beyond View Alignment Assumptions"** submitted by **Puspamalya** to Indian Statistical Institute, Kolkata, in partial fulfillment for the award of the degree of **Master of Technology in Cryptology and Security** is a bonafide record of work carried out by him under my supervision and guidance. The project has fulfilled all the requirements as per the regulations of this Institute. The contents of this project, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Malay Bhattacharyya
Associate Professor, Machine Intelligence Unit
Indian Statistical Institute, Kolkata



Dr. Anirban Mukhopadhyay
Professor, Department of Computer Science and Engineering

Director, Internal Quality Assurance Cell (IQAC)
University of Kalyani

Acknowledgement

I would like to convey my great appreciation to my supervisors, Dr. Malay Bhattacharyya and Prof. Anirban Mukhopadhyay, for their guidance and encouragement throughout the duration of this project. With his guidance, I now have a deep appreciation for the field of machine learning and the research sphere as a whole.

My deepest thanks to the faculties of Indian Statistical Institute, for their support.

Last but not the least, I would like to thank all my family, friends and peers for their continuous help and support. Finally, I would like to thank all those whom I might have missed out on the above list.

Puspamalaya Sahoo
Roll No. CrS2316
Indian Statistical Institute
Kolkata – 700108, India.

Abstract

Multiview learning aims to integrate diverse feature representations to achieve a comprehensive understanding of data. Traditional approaches often assume strict alignment across views, making them ill-suited for real-world scenarios where low-quality conflictive instances, i.e. instances with conflicting information across views are prevalent. Existing methods largely focus on eliminating conflicting instances by discarding them or substituting conflicting views, overlooking the need for practical decision making in such cases. Furthermore, while the recently proposed Reliable Conflictive Multiview Learning (RCML) framework introduces the idea of attaching reliabilities to decision outcomes, it leaves certain theoretical gaps unaddressed, especially prioritization of conflictive views in fusion process in a principled manner.

Contents

List of Tables

Chapter 1

Introduction

In the big data regime, data acquisition techniques have become very much diversified toward different applications. For example, an individual can describe an image from different angles. The same image can be described by different features such as colours, texts, etc. These data for the same entity from different perspectives are called multiview data. Different views represent different features of the data which are fundamentally consistent. Simply taking all the views together and applying traditional machine learning methods lead to the curse of dimensionality [1], also this makes it difficult to expose the latent information present in the data. Hence, multiview learning has become a well-established domain to the machine learning research community who tackles problems involving the availability of multiple sources of data. The use of multiview data helps us to take the advantage of grabbing complementary information, which results in better learning performance. Wei et al. [2] conducted a study that highlighted that multiview learning methods outperform the single view approaches when applied to both uni-view and multiview surface electromyography data streams. At the same time, Tian et al. [3] introduced multiview deep learning and achieved a better performance in the context of detecting epileptic seizures based on encephalograms (EEGs). In another study, Kong et al. [4] used a multiview learning method using a deep neural model for action recognition that achieved a better result compared to single view learning method. Furthermore, with the consideration of deep Gaussian process, Sun et al. [5] represented another multiview approach which had shown better performance in real world mutlview dataset. In another image classification task, Zhang et al. [6] experimented multiview visual classification methods and achieved promising performance.

In general, combining different views during the learning process helps to capture various perspectives of the data which leads to increased preciseness, improved robustness and generalization of the process. In the exploration of multiview learning, people are trying to find an effective method for better utilization of the information collected from different views. As time grows, it is recognized in the research community that the key to multiview learning lies in balancing the consistency and complementarity among different views [7], namely the two principles of multiview learning:

1. **Consensus principle**: The consensus principle refers to the consistency among different views. It means that to some extent multiple views should agree to capture the common features in the data. The fundamental concept of the consensus principle is that there

are some degrees of correlations among different views. For example, in a data containing both image and text data, there must have some similarity between described text objects and the images.

2. **Complementarity principle**: The complementarity principle refers to the complementary nature of different views. In multiview learning, the complementarity principle suggests that each view offers unique but related insights, capturing different aspects or finer details of the data. When these distinct sources of information are combined, they can enrich the overall data representation and boost the learning performance. For instance, in an image recognition task, visual features (like pixel data) can be complemented by textual descriptions or semantic labels to achieve better classification accuracy. Together with the consensus principle—which focuses on the agreement and shared patterns across views—complementarity helps in effectively fusing information. While consensus promotes alignment between views, complementarity values the diversity of each view.

The multiview learning put forward several challenges that have drawn recent attention of researchers and practitioners. These challenges encompass view inconsistency, view complementarity, optimal view fusion, the curse of dimensionality, scalability, limited labels, generalization across domains and others. These are briefly discussed hereunder.

- **View inconsistency**: In multiview learning, each view usually represents a different side or aspect of the same data. While these views can be combined into a single vector through preprocessing, some important information might get lost in the process. This can lead to inconsistencies between the views. These inconsistencies may be caused by different preprocessing steps, embedding techniques, or feature scales. They can also result from missing data, noisy inputs or labels, or misaligned views. Although some recent methods try to handle issues like misalignment, unmapped data, and noisy correspondences, research in this area is still limited. Moreover, not all views are equally useful for the learning task, especially when different feature extraction methods are used. As a result, managing view inconsistencies remains a key challenge in multiview learning.

- **View complementarity**: While each view contributes distinct information, there also exists potential complementarity across views—where the information from one view can enhance or complete that of another. The key challenge lies in effectively identifying and leveraging this complementary information to improve overall learning performance. This requires a deeper understanding of both the individual characteristics of each view and their interrelationships. However, many existing multiview learning approaches face limitations in this regard. A common issue is the oversimplified assumption that views are entirely independent and inherently complementary. In practice, views often exhibit complex dependencies, and such simplistic assumptions can hinder the effective exploitation of their complementarities. Moreover, several methods primarily focus on aggregating information across views without explicitly modeling or utilizing the complementary aspects, which may lead to suboptimal learning outcomes.

- **Optimal view integration:** the task of integration of different views into a single model is a non-trivial task. The Existing methods can be classified into two types on the basis of the timing of view integration: **early integration** and **late integration**. Early integration integrates the features from different view whereas late integration has

a similarity with ensemble learning. It is yet to be known which fusion method is more effective and demands extensive theoretical research.

Determining how to effectively combine multiple views and evaluate their importance is a key challenge. The fusion process should be able to capture the shared, useful information across views while minimizing noise and redundant features.

Chapter 2

Related Work

2.1 Evidential Deep Learning to Quantify Classification Uncertainty

[8] In this paper, they have provided uncertainty estimation method and approached it from the perspective of Theory of Evidence. They have interpreted softmax, the standard output of a classification deep neural network, as the set of parameters of a categorical distribution. By replacing the parameter set with the parameters of a Dirichlet density function that captures a range of possible softmax outputs, their model does not produce a single prediction; instead, it models softmax outputs by generating a distribution over possible softmax outputs.It can be thought of the Dirichlet density function as a generator of many possible softmax outputs. The deep neural model is being trained on a specific loss function.

The **Dempster–Shafer Theory of Evidence (DST)** is a generalization of the Bayesian theory to subjective probabilities. It assigns belief masses i.e. evidences to subsets of a frame of discernment, which are the set of exclusive possible states, in our case, possible class labels for a sample. A belief mass can be assigned to any subset of the frame, including the whole frame itself, which represents the belief that the truth can be any of the possible states, e.g., any class label is equally likely for a given sample. In other words, by assigning all belief masses to the entire frame, we can express 'I do not know' as an opinion for the truth over possible states. Subjective logic (SL) formalizes the notion of belief assignments in the DST on a framework of discernment as a Dirichlet distribution. Hence, it allows one to use the principles of evidential theory to quantify belief masses and uncertainty through a well-defined theoretical framework. More precisely, Subjective Logic considers a frame of K mutually exclusive singletons (e.g., class labels) by providing a belief mass b_k for each class $k = 1, \ldots, K$ and providing an overall uncertainty mass of u. These $K + 1$ mass values are all non-negative and sum up to one, i.e.,

$$u + \sum_{k=1}^{K} b_k = 1 \tag{2.1}$$

where $u \geq 0$ and $b_k \geq 0$ for $k = 1, \ldots, K$. A belief mass b_k for a singleton k is calculated

using the evidence for the individual classes. Let $e_k \geq 0$ be the evidence derived for the k^{th} singleton, then the belief b_k and the uncertainty u are computed as

$$b_k = e_k/S \text{ and } u = K/S \tag{2.2}$$

where $S = \sum_{i=i}^{K}(e_i+1)$. It can be noted that the uncertainty decreases as the total evidence increases, indicating an inverse relationship between the two. When there is no evidence to any of the classes, the belief for each class is zero and hence the uncertainty is one. Differently from the Bayesian modeling approach, we define *evidence* as a measure of the amount of support collected from data in favor of a sample to be classified into a certain class. An assignment of a belief mass corresponds to a Dirichlet distribution with parameters $\alpha_k = e_k + 1$. That is, opinion for the classes corresponding to a sample can be derived easily from the parameters of the corresponding Dirichlet distribution using this formula $b_k = (\alpha_k - 1)/S, \text{where} S = \sum_{i=1}^{K} \alpha_i$ is said to be the Dirichlet strength. While a standard neural network classifier outputs a probability distribution over possible classes for each sample, a Dirichlet distribution is parametrized by evidence and captures the distribution over these probability assignments themselves. As a result, it models second-order probabilities and uncertainty.

The Dirichlet distribution is a probability density function (pdf) for possible values of the probability mass function (pmf) P. It has K parameters $\alpha = [\alpha_1, \ldots, \alpha_K]$ and is given by

$$D(\boldsymbol{P} \mid \boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} p_i^{\alpha_i - 1} & \text{for } \boldsymbol{P} \in S_K \\ 0 & \text{otherwise} \end{cases}$$

where S_K is the K-dimensional unit simplex,

$$S_k = \left\{ \boldsymbol{P} \mid \sum_{i=1}^{K} p_i = 1 \text{ and } 0 \leq p_1, p_2, \ldots, p_K \leq 1 \right\}$$

and $B(\boldsymbol{\alpha})$ is the K-dimensional multinomial beta function. Consider that there is an opinion $b = \langle 0, \ldots, 0 \rangle$ as the assignment of belief for a 10-class classification problem. Then, the prior distribution for the classification problem becomes a uniform distribution, i.e., $D(\boldsymbol{P} \mid \langle 0, \ldots, 0 \rangle)$ this is nothing but a Dirichlet distribution with all the parameters as 1 . That means there is no evidence for any of the classes, as the belief masses are all zero. This means that the opinion corresponds to the uniform distribution implies total uncertainty, i.e., $u = 1$. Now again consider an another belief masses, $b = \langle 0.8, \ldots, 0 \rangle$ after some training epoch. This implies that the total belief in the opinion is 0.8 and the remaining 0.2 is the uncertainty. Dirichlet strength is calculated as $S = 10/0.2 = 50$, since $K = 10$. Hence, the amount of new evidence for the first class is calculated as $50 \times 0.8 = 40$. Now the opinion would correspond to the Dirichlet distribution $D(\boldsymbol{P} \mid \langle 41, 1 \ldots, 1 \rangle)$.

Foe a given opinion, the expected probability for the $k - th$ class is related to the mean of the corresponding Dirichlet distribution and computed as

$$p_k = \frac{\alpha_k}{S} \tag{2.3}$$

So, what they are trying to establish that the parameters of a Dirichlet distribution for the classification of a sample will be directly proportional to the evidences for each class.

Let us assume that $\boldsymbol{\alpha_i} = \langle \alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK} \rangle$ is the parameters of a Dirichlet distribution for the classification of a sample i, then $(\alpha_{ij} - 1)$ is the total evidence computed by the deep neural network for the classification of the sample i to the jth class. Also for this given parameters, uncertainty of the classification can easily be computed using Equation (2.2).

2.2 Trusted Multi-View Deep Learning with Opinion Aggregation

They developed a multiview deep learning method [9] through simulating opinion aggregation method to achieve better results. The proposed method have used Evidential Deep Learning theory as discussed above to get the opinions from different views and represents the integrated opinion as multiview learning result through opinion aggregation in a specific way as discussed below.

They have theoretically proved that accumulating the evidences from multiple views will decrease the overall uncertainty and increase the prediction accuracy after aggregation of all those opinions. Moreover, they have further extended their method by minimizing the opinion entropy across views to pursue the consistency across multiple views.

Opinion Aggregation with Evidence Accumulation The opinion of a single view has been computed as in Section 2.1 which provide explicit estimation of the uncertainty degree. The method of opinion aggregation with evidence accumulation is described as below.

Opinion aggregation with evidence accumulation. The opinion aggregation with evidence accumulation simply done by evidence parameter addition. Given a sample with V number of views for some classification problem with K classes. a set of evidences $\{\boldsymbol{e}^v\}_{v=1}^V$ collected from V neural networks using the method of evidential theory above and a set of opinions $\{\boldsymbol{\omega}^v\}_{v=1}^V$ in terms of equation (2.2). Then we have the integrated opinion as follows:

$$\boldsymbol{\omega}^{\diamond(V)} = \bigoplus_{v=1}^{V} \boldsymbol{\omega}^v = (\boldsymbol{b}^{\diamond(V)}, u^{\diamond(V)}, \boldsymbol{a}^{\diamond(V)}) \quad \text{For } k = 1, \ldots, K \tag{2.4}$$

we have

$$b_k^{\diamond(V)} = \frac{e_k^{\diamond(V)}}{S^{\diamond(V)}},\ u^{\diamond(V)} = 1 - \sum_{k=1}^{K} b_k^{\diamond(V)},\ a_k^{\diamond(V)} = \frac{1}{K} \tag{2.5}$$

where $S^{\diamond(V)} = \sum_{k=1}^{K}(e_k^{\diamond(V)} + 1)$ is the Dirichlet strength, The combined evidence can be computed in the following way $e_k^{\diamond(V)} = \sum_{k=1}^{K} e_k^v + 1$. Hence we can compute the integrated opinion as follows $\boldsymbol{\omega}^{\diamond(V)} = (\boldsymbol{b}^{\diamond(V)}, u^{\diamond(V)}, \boldsymbol{a}^{\diamond(V)})$. Then the corresponding integrated parameters of the Dirichlet distribution are as follows $\alpha_k^{\diamond(V)} = e_k^{\diamond(V)} + 1$.

2.3 Reliable Conflictive Multi-View Learning

As described in the above paper, in this paper also, they have computed the evidences and opinion using the evidential theory approach [8].

Evidential Multi-view Fusion via Conflictive Opinion Aggregation. They have introduced a new way of view fusion. The conflictive(noise) views of the conflictive multiview data would shows high uncertainty. They were trying to reduce the impact of these uncertainty in the opinion fusion stage. The unaligned views of the conflictive multi-view data which is arrived because of noise or any other error may provide very conflicting opinions with low amount of uncertainty, which indicates that one or more number of views are unreliable. In this case, it is difficult to judge which view is more authentic. As a matter of fact the uncertainty of multi-view learning output should not decrease with the increase of the number of views, but at the same time it should be related to the reasoning behind the fusion method especially when the learning results of two views conflict with each other. To solve this, they have proposed a new conflictive opinion aggregation method.
Definition 1 Conflictive Opinion Aggregation. Let $\omega^A = (\boldsymbol{b}^A, u^A, \boldsymbol{a}^A)$ and $\omega^B = (\boldsymbol{b}^B, u^B, \boldsymbol{a}^B)$ be two be the opinions of view A and B over the same sample, respectively. The conflictive aggregated opinion $\boldsymbol{\omega}^{A\diamond B}$ is calculated in the following manner:

$$\boldsymbol{\omega}^{A\diamond B} = \boldsymbol{\omega}^A \diamond \boldsymbol{\omega}^B = (\boldsymbol{b}^{A\diamond B}, u^{A\diamond B}, \boldsymbol{a}^{A\diamond B}) \tag{2.6}$$

$$\boldsymbol{b}_k^{A\diamond B} = \frac{b_k^A u^B + b_k^B u^A}{u^A + u^B} \tag{2.7}$$

$$u^{A\diamond B} = \frac{2u^A u^B}{u^A + u^B}, \boldsymbol{a}_k^{A\diamond B} = \frac{\boldsymbol{a}_k^A + \boldsymbol{a}_k^B}{2} \tag{2.8}$$

The opinion $\boldsymbol{\omega}^{A\diamond B}$ represents the combination of the opinions of view A and view B. This combination is achieved by mapping the belief opinions to evidence opinions using a bijective mapping between multinomial opinions and the Dirichlet distribution. The above combination rule establishes that the quality of the combined new opinion is proportional to the opinions have been combined. In other words, when a highly uncertain opinion is combined with a more certain original opinion, the uncertainty of the new opinion is larger than the original opinion. Following Definition 1,For more than two views the joint opinion $\boldsymbol{\omega}$ can be computed with the following rule:

$$\boldsymbol{\omega} = \boldsymbol{\omega}^1 \diamond \boldsymbol{\omega}^2 \diamond \ldots \diamond \boldsymbol{\omega}^V$$

According to the above fusion rules, the final multi-view joint opinion can be obtained and thus get the final probability of each class and the overall uncertainty.

Chapter 3

Methodology

3.1 Proposed Method

3.1.1 Problem Definition

Suppose we are given with a dataset with V views, $\bar{N}$ normal instances and $\tilde{N}$ conflictive(noisy) instances. We use $\boldsymbol{x}_n^v \in \mathbb{R}^{D_v}$ $(v = 1, 2, \ldots, V)$ to denote the characteristic vector for the v-th view of the n-th instance $(n = 1, \ldots, N)$, where D_v is the dimensionality of the v^{th} view. The one-hot vector $y_n \in \{0, 1\}^K$ denotes the true label of the n^{th} instance, where K is the number of overall categories. The training examples $\left\{ \{\boldsymbol{x}_n^v\}_{v=1}^V, \boldsymbol{y}_n \right\}_{n=1}^{N_{\text{train}}}$ contain $\bar{N}_{\text{train}}$ normal instances. The other $\bar{N} - \bar{N}_{\text{train}}$ normal instances and $\tilde{N}$ conflictive(noisy) instances form the test set. We would like our model to predict accurately y_n for the test instances along with a measurement of uncertainty $u_n \in [0, 1]$ to measure how our model is confident about its output , let us name it as decision reliability and computed as $(1 - u_n)$.

Definition 3.1 (Conflictive Opinion Aggregation through Decreasing Uncertainty) *Let $\omega^A = (\boldsymbol{b}^A, u^A, \boldsymbol{a}^A)$ and $\omega^B = (\boldsymbol{b}^B, u^B, \boldsymbol{a}^B)$ be the opinions of views A and B, respectively, over the same instance. The conflictive aggregated opinion [10] $\boldsymbol{\omega}^{A \diamond B}$ is calculated in the following manner:*

$$\begin{aligned} \boldsymbol{\omega}^{A \diamond B} = \boldsymbol{\omega}^A \diamond \boldsymbol{\omega}^B &= (\boldsymbol{b}^{A \diamond B}, u^{A \diamond B}, \boldsymbol{a}^{A \diamond B}) \\ \boldsymbol{b}_k^{A \diamond B} &= \frac{b_k^A u^B + b_k^B u^A}{u^A + u^B} \\ u^{A \diamond B} = \frac{2u^A u^B}{u^A + u^B}, \boldsymbol{a}_k^{A \diamond B} &= \frac{\boldsymbol{a}_k^A + \boldsymbol{a}_k^B}{2} \end{aligned}$$

For aggregating more than two views, we assume that the uncertainties $(u^1, u^2, \ldots u^V)$ of the respective opinions $(\boldsymbol{\omega}^1, \boldsymbol{\omega}^2, \ldots \boldsymbol{\omega}^V)$ are sorted in decreasing order, i.e., $u^1 \geq u^2 \geq \ldots \geq u^V$

The aggregated joint opinion $\boldsymbol{\omega}$ is computed sequentially in the following order:

$$\boldsymbol{\omega} = ((\ldots(((\boldsymbol{\omega}^1 \diamond \boldsymbol{\omega}^2) \diamond \boldsymbol{\omega}^3) \diamond \boldsymbol{\omega}^4) \diamond \ldots \diamond \boldsymbol{\omega}^{V-1}) \diamond \boldsymbol{\omega}^V) \tag{3.1}$$

This ordering ensures that **views with lower uncertainty (i.e., higher confidence) is given greater weight** during the fusion process, thus minimizing the overall uncertainty of the aggregated opinion. It is being proved in the lemma.

Definition 3.2 (Conflictive Degree) *Given opinions $\boldsymbol{\omega}^A$ and $\boldsymbol{\omega}^B$ for the views A and B over an instance, the conflictive degree [10] between $\boldsymbol{\omega}^A$ and $\boldsymbol{\omega}^B$ is defined as:*

$$c(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B) = c_p(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B) \cdot c_c(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B) \tag{3.2}$$

where $c_p(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B)$ is the projected distance between $\boldsymbol{\omega}^A$ and $\boldsymbol{\omega}^B$, $c_c(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B)$ is the conjunctive certainty between $\boldsymbol{\omega}^A$ and $\boldsymbol{\omega}^B$, which can be formulated as follows:

$$c_p(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B) = \frac{\sum_{k=1}^{K} \left| p_k^A - p_k^B \right|}{2}, \tag{3.3}$$

$$c_c(\boldsymbol{\omega}^A, \boldsymbol{\omega}^B) = (1 - u^A)(1 - u^B). \tag{3.4}$$

Intuitive Interpretation of the Conflictive Degree

The conflictive degree c captures two key scenarios:

1. **Minimal Conflict** ($c \approx 0$) **:**
   This case arises in either of two the following cases:
   - The projected probability distributions of two opinions are identical. i.e. $c_p = 0$
   - One or both views are vacuous, i.e., have maximum uncertainty mass ($u = 1$), resulting in $c_c = 0$.
2. **Maximal Conflict** ($c \approx 1$) **:**
   This scenario occurs when the opinions have non-identical projected probabilities and are both credible, meaning:
   - The uncertainty masses are minimal ($u = 0$) hence $c_c = 1$.
   - Evidence is accumulated into a single but different class by the two views.

   The uncertainty masses are minimal ($u = 0$)

Accordingly, two views A and B are said to be:

- **Highly conflictive** if both the **conjunctive certainty** c_c and **projected distance** c_p are closed to 1, indicating strong disagreement between the views A and B.
- **Consistent** if both the **conjunctive certainty** c_c and **projected distance** c_p are closed to 0, indicating agreement or a lack of confident information.

Intuitively, this metric ensures two things: (1) The case where $c = 0$ arrive when the same probability distributions are being observed for two different views, which indicates no conflict between these two opinion or $c_c = 0$ i.e one or both of the views have highest uncertainty mass i.e. 0. (2) $c = 1$ is attained when opinions are present but with completely different projected probabilities. On the other hand, when $c_c = 1$, it says that the opinions are considered reliable, that means they have zero uncertainty mass. Two views A and B are said to be highly conflictive if c_c and c_p is closed to 1 and they are said to be consistent both if c_c or c_p are closed to zero.

3.2 Loss Function

In this section, the training DNN to get the multiview opinion is introduced. Traditional Deep Neural Network can be converted into evidential Deep Neural Network with little modification as described in (Sensoy, Kaplan, and Kandemir 2018). In this modification the output softmax layer is replaced with an activation layer (e.g., ReLU) to get the non-negative output of this layer as evidence. By doing so, the parameters of the Dirichlet distribution can be computed. Let $\{\boldsymbol{x}_n^v\}_{v=1}^V, \boldsymbol{e}_n^v = f^v(\boldsymbol{x}_n^v)$ represent the evidence vector computed by the network for the classification. Then $\boldsymbol{\alpha}_n^v = \boldsymbol{e}_n^v + 1$ is the parameters of the corresponding Dirichlet distribution. In the case of conventional neural network-based classifiers, the cross-entropy loss is employed generally.In our case we need to adapt the cross-entropy loss to take into account the evidence-based approach:

$$L_{ace}(\boldsymbol{\alpha}_n) = \int [\sum_{j=1}^{K} -y_{nj} \log p_{nj}] \frac{\prod_{j=1}^{K} p_{nj}^{\alpha_{nj}-1}}{B(\boldsymbol{\alpha}_n)} d\boldsymbol{p}_n = \sum_{j-1}^{K} y_{nj}(\psi(S_n) - \psi(\alpha_{nj}) \tag{3.5}$$

where $\psi(\cdot)$ is the digamma function.
The above loss function does not guarantee that the evidence generated by the incorrect labels is lower. To address this issue, we can introduce an additional term in the loss function, namely the Kullback-Leibler (KL) divergence as follows:

$$\begin{aligned} L_{KL}(\boldsymbol{\alpha}_n) &= KL[D(\boldsymbol{P}_n \mid \tilde{\boldsymbol{\alpha}}_n) \mid\mid D(\boldsymbol{P}_n \mid \mathbf{1})] \\ &= \log(\frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_{nk})}{\Gamma(K) \prod_{k=1}^{K} \Gamma(\tilde{\alpha}_{nk})}) + \sum_{k=1}^{K} (\tilde{\alpha}_{nk} - 1)[\psi(\tilde{\alpha}_{nk}) - \psi(\sum_{j=1}^{K} \tilde{\alpha}_{nj})] \end{aligned} \tag{3.6}$$

where $D(\boldsymbol{P}_n \mid \mathbf{1})$ is the uniform Dirichlet distribution, $\tilde{\boldsymbol{\alpha}}_n = \boldsymbol{y}_n + (1 - \boldsymbol{y}_n) \bigodot \boldsymbol{\alpha}_n$ is the Dirichlet parameters after removal of the non-misleading evidence from predicted parameters $\boldsymbol{\alpha}_n$ for the n-th instance, and $\Gamma(\cdot)$ is the gamma function.
Therefore, given the Dirichlet distribution with parameter $\boldsymbol{\alpha}_n$ for the n-th instance, the loss is:

$$L_{acc}(\boldsymbol{\alpha}_n) = L_{ace}(\boldsymbol{\alpha}_n) + \lambda_t L_{KL}(\boldsymbol{\alpha}_n) \tag{3.7}$$

where $\lambda_t = \min(1.0, t/T) \in [0, 1]$ is the annealing coefficient, t is the index of the current training epoch, and T is the annealing step. By gradually increasing the influence of **KL** divergence in loss, premature convergence of misclassified instances to uniform distribution can

be avoided.

In order to ensure the consistency of results between different opinions during training, minimizing the degree of conflict between opinions was adopted. The consistency loss for the instance $\{\boldsymbol{x}_n^v\}_{v=1}^V$ is calculated as:

$$L_{con} = \frac{1}{V-1}\sum_{p=1}^{V}\left(\sum_{q\neq p}^{V} c(\boldsymbol{\omega}_n^p, \boldsymbol{\omega}_n^q)\right) \tag{3.8}$$

Hence to conclude, the overall loss function for a specific instance $\{\boldsymbol{x}_n^v\}_{v=1}^V$ is calculated as:

$$L = L_{acc}(\boldsymbol{\alpha}_n) + \beta\sum_{v=1}^{V} L_{acc}(\boldsymbol{\alpha}_n^v) + \gamma L_{con} \tag{3.9}$$

3.3 Proof of lemmas

Lemma 3.1 *The aggragation of opinion $\boldsymbol{\omega}^A$ and opinion $\boldsymbol{\omega}^B$, i.e. $\boldsymbol{\omega}^{A\diamond B} = \boldsymbol{\omega}^A \diamond \boldsymbol{\omega}^B$ corresponds to averaging the view-specific evidences $\boldsymbol{e}^{A\diamond B} = \frac{1}{2}(\boldsymbol{e}^A + \boldsymbol{e}^B)$.*

Proof 3.1 *Let $\boldsymbol{\omega}^i = (\boldsymbol{b}^i, u^i, \boldsymbol{a}^i)$ and $\boldsymbol{\omega}^j = (\boldsymbol{b}^j, u^j, \boldsymbol{a}^j)$ be the multinomial opinion corresponding to the i-th and j-th view of the same sample respectively. After averaging belief fusion, $\boldsymbol{\omega} = (\boldsymbol{b}, u, \boldsymbol{a})$ is the integrated multinomial opinion from the multinomial opinion ω^i and ω^j. Correspondingly, e_k^i, e_k^j and e_k are the k-th category of evidence for the i-th, j-th and integrated view of the same sample.*

$$b_k = \frac{e_k}{S}, u = \frac{K}{S}, S = \sum_{k=1}^{K}(e_k + 1), \tag{3.10}$$

$$b_k = \frac{b_k^i u^j + b_k^j u^i}{u^i + u^j}, \tag{3.11}$$

$$u = \frac{2u^i u^j}{u^i + u^j}, a_k = \frac{a_k^i + a_k^j}{2}. \tag{3.12}$$

From the above three equations, e_k can be updated as follows:

$$\begin{aligned} e_k = b_k S &= \frac{b_k K}{u} \\ &= \frac{b_k^i u^j + b_k^j u^i}{u^i + u^j} \cdot \frac{K(u^i + u^j)}{2u^i u^j} \\ &= \frac{K}{2} \cdot \frac{b_k^i u^j + b_k^j u^i}{u^i u^j} \\ &= \frac{K}{2} \cdot \frac{\left(\frac{e_k^i}{S^i} \cdot \frac{K}{S^j} + \frac{e_k^j}{S^j} \cdot \frac{K}{S^i}\right)}{\frac{K}{S^i} \cdot \frac{K}{S^j}} \\ &= \frac{e_k^i + e_k^j}{2} \end{aligned}$$

Hence, we have $\boldsymbol{e}^{A\diamond B} = \frac{1}{2}(\boldsymbol{e}^A + \boldsymbol{e}^B)$.

Lemma 3.2 *For the opinion aggregation method, after aggregating a new opinion* $\boldsymbol{\omega}^a$ *into the original opinion* $\boldsymbol{\omega}^o$*, if the uncertainty mass of the new opinion* u^a *is lesser than the original uncertainty mass* u^o*, the uncertainty mass of the combined opinion* u *will be lesser than the original one; and if the original uncertainty mass is larger than that of new opinion then the combined opinion will be greater than the original one.*

Proof 3.2 *We know*

$$u = \frac{2u^a u^o}{u^a + u^o} = \frac{1}{\frac{1}{2}(1 + \frac{u^o}{u^a})} \cdot u^o.$$

The uncertainty mass of the aggregated opinion and the original opinion is:

$$\begin{cases} u < u^o, & if u^a < u^o \\ u = u^o, & if u^a = u^o \\ u > u^o, & if u^a > u^o \end{cases}$$

Lemma 3.3 *Let* $\boldsymbol{e}^1, \boldsymbol{e}^2, \ldots, \boldsymbol{e}^V$ *be evidence vectors from* V *different views ordered such that their uncertainty masses satisfy:*

$$u^1 \geq u^2 \geq \ldots \geq u^V$$

Define the sequential fusion as follows:

$$\begin{aligned} \boldsymbol{E}^2 &= e^1 \diamond e^2 \\ \boldsymbol{E}^v &= \boldsymbol{E}^{v-1} \diamond \boldsymbol{e}^v \quad for \quad v = 3, 4, \ldots, V \end{aligned}$$

Then, this fusion order minimizes the overall uncertainty of the final combined evidence $\boldsymbol{E}^V$ *among all possible fusion sequences.*

Proof 3.3 $\boldsymbol{e}^i = (e_1^i, e_2^i, \ldots, e_K^i)$ *and* $\boldsymbol{e}^j = (e_1^j, e_2^j, \ldots, e_K^j)$ *be two evidences with corresponding uncertainty masses* u^i, u^j *such that* $u^i \geq u^j$*. Let* $r, s \in \mathbb{N}$ *such that* $r < s$*.*

From Lemma 3.1 we have

$$\boldsymbol{E}^V = (\frac{1}{2^V}) \cdot \boldsymbol{e}^1 + (\frac{1}{2^{V-1}}) \cdot \boldsymbol{e}^2 + \ldots + (\frac{1}{2^2}) \cdot \boldsymbol{e}^{V-1} + (\frac{1}{2}) \cdot \boldsymbol{e}^V$$

By definition, the uncertainty mass u *of an evidence vector* $\boldsymbol{e} = (e_1, e_2, \ldots, e_K)$ *is inversely proportional to total evidence mass, i.e.,*

$$u \propto \frac{1}{\sum_{k=1}^{K} e_k}$$

Therefore, from $u^i \geq u^j$*, it follows that:*

$$
\begin{aligned}
u^i &\geq u^j \\
\Rightarrow \sum_{k=1}^{K} e_k^i &\leq \sum_{k=1}^{K} e_k^j \\
\Rightarrow \frac{1}{2^s}(\sum_{k=1}^{K} e_k^i) &< \frac{1}{2^r}(\sum_{k=1}^{K} e_k^j)
\end{aligned}
$$

This implies that placing a more uncertain view earlier in the fusion sequence results in a lower contribution to the final aggregated evidence, thereby letting the more certain view contribute more to the aggregated evidence.

Hence, combining views in the increasing order of certainty ***(i.e., the decreasing order of uncertainty)*** *minimizes the uncertainty in the final opinion* $\boldsymbol{E}^V$.

Chapter 4

Experimental Results

In this chapter we take a look at the results obtained by applying the existing and proposed methods

4.1 Design of Experiments

In this section, we evaluate the model on 6 real-world multiview datasets.

4.1.1 Experimental Setup

Datasets

- **HandWritten** contains 2000 images of handwritten numerals from $'0'$ to $'9'$, with 200 patterns per class. It represents using six feature sets taking outputs from CNN.
- **Scene15** contains 4485 images from 15 indoor and outdoor scene categories. Three types of feature GIST, PHOG , and LBP have been extracted.
- **PIE** contains 680 examples from 68 classes. Intensity, LBP, and Gabor as 3 views have been extracted.
- **Caltech101** comprises 8677 images from 101 classes. The first 10 categories have been selected. It is represented using six feature sets.
- **CUB** consists of 11788 instances which are associated with text descriptions of 200 different categories of birds. In this study,the first 10 categories have been focused and image features are extraced using GoogleNet and corresponding text features are extracted using doc2vec.
- **Colored-MNIST** includes 18835 instances of digits with RGB Coloured backgrounds which consist of three colors (red, green, blue) for each number.A color bias is intentionally introduced, meaning that certain digits are more likely to be associated with specific

Dataset	Size	Classes	Dimensionality
HandWritten	2000	10	240/76/216/47/64/6
Scene15	4485	15	20/59/40
PIE	680	68	484/256/279
Caltech101	2386	10	48/40/254/1984/512/928
CUB	600	10	1024/300
Colored-MNIST	1200	10	768/576

Table 4.1: Dataset Summery

Normal_data		Conflict_data		Gaussian_noise	
random_order	specific_order	random_order	specific_order	random_order	specific_order
HandWritten, 2000					
96.95 ± 00.78	$\mathbf{97.35 \pm 00.86}$	$\mathbf{92.75 \pm 01.25}$	91.40 ± 01.61	94.92 ± 01.51	$\mathbf{97.37 \pm 00.84}$
Scene15, 4485					
70.12 ± 01.45	$\mathbf{71.38 \pm 01.36}$	59.77 ± 01.27	$\mathbf{61.18 \pm 01.63}$	67.02 ± 01.25	$\mathbf{69.04 \pm 01.38}$
PIE, 600					
93.01 ± 01.89	$\mathbf{93.26 \pm 01.70}$	81.98 ± 02.80	$\mathbf{83.16 \pm 03.22}$	92.57 ± 02.31	$\mathbf{98.83 \pm 02.61}$
Caltech101, 2386					
$\mathbf{93.26 \pm 01.24}$	92.47 ± 01.26	$\mathbf{88.37 \pm 01.38}$	86.99 ± 01.43	92.86 ± 01.42	$\mathbf{92.88 \pm 01.48}$
CUB , 600					
91.33 ± 02.56	$\mathbf{91.58 \pm 01.91}$	70.67 ± 03.51	$\mathbf{72.74 \pm 03.39}$	91.25 ± 01.98	$\mathbf{91.66 \pm 01.66}$
CNIST, 1200					
40.46 ± 02.01	$\mathbf{43.17 \pm 02.24}$	$\mathbf{32.83 \pm 03.04}$	32.29 ± 02.11	41.25 ± 02.50	$\mathbf{42.20 \pm 03.22}$

Table 4.2: Performance comparison between normal data and data with gaussian noise

colors. 1200 instances are chosen and RGB and HOG features have been extracted as two views.

To create a test set with conflictive instances, we apply the following transformation: For the noisy views, three types of noises —— Normal, Laplace, and Uniform —— are introduced with certain level of standard deviations (σ) applied to 20% of the test instances. For each method (RCML and the proposed approach), we perform 10 independent runs both on the clean (normal) and noisy datasets, and report the mean and standard deviation of the classification accuracy to ensure a robust and reliable evaluation.

4.1.2 Results

Table 2, 3 and 4 show the classification performance on normal and 3 types of noisy test sets.

From the preceding three performance tables, it is evident that the proposed method outperforms the baseline(RCML) across most datasets under both normal and noisy conditions, with the exception of Caltech101. A plausible explanation for this deviation is the significant class imbalance present in the Caltech101 dataset.

Normal_data		Conflict_data		Laplace_noise	
random_order	specific_order	random_order	specific_order	random_order	specific_order
HandWritten, 2000					
96.95 ± 00.78	$\mathbf{97.35 \pm 00.86}$	$\mathbf{90.77 \pm 01.02}$	90.52 ± 01.15	94.57 ± 01.24	$\mathbf{96.74 \pm 00.72}$
Scene15, 4485					
70.12 ± 01.45	$\mathbf{71.38 \pm 01.36}$	58.50 ± 01.18	$\mathbf{59.64 \pm 00.91}$	66.29 ± 01.76	$\mathbf{67.95 \pm 01.31}$
PIE, 600					
93.01 ± 01.89	$\mathbf{93.26 \pm 01.70}$	69.19 ± 03.14	$\mathbf{81.98 \pm 01.71}$	83.60 ± 03.15	$\mathbf{91.61 \pm 02.18}$
Caltech101, 2386					
$\mathbf{93.26 \pm 01.24}$	92.47 ± 01.26	$\mathbf{87.94 \pm 01.42}$	87.13 ± 01.19	$\mathbf{92.61 \pm 01.56}$	92.19 ± 00.96
CUB , 600					
91.33 ± 02.56	$\mathbf{91.58 \pm 01.91}$	71.08 ± 03.94	$\mathbf{71.08 \pm 02.52}$	87.83 ± 01.79	$\mathbf{89.66 \pm 02.39}$
CNIST, 1200					
40.46 ± 02.01	$\mathbf{43.17 \pm 02.24}$	12.95 ± 02.46	$\mathbf{33.12 \pm 03.18}$	14.38 ± 01.47	$\mathbf{41.08 \pm 02.50}$

Table 4.3: Performance comparison between normal data and data with Laplacian noise

Normal_data		Conflict_data		Uniform_noise	
random_order	specific_order	random_order	specific_order	random_order	specific_order
HandWritten, 2000					
96.95 ± 00.78	$\mathbf{97.35 \pm 00.86}$	$\mathbf{92.72 \pm 01.13}$	91.4 ± 01.46	96.42 ± 00.77	$\mathbf{97.52 \pm 00.71}$
Scene15, 4485					
70.12 ± 01.45	$\mathbf{71.38 \pm 1.36}$	60.22 ± 1.39	$\mathbf{61.90 \pm 1.68}$	67.49 ± 1.66	$\mathbf{69.28 \pm 1.52}$
PIE, 680					
93.01 ± 01.89	$\mathbf{93.26 \pm 01.70}$	74.11 ± 04.03	$\mathbf{85.07 \pm 02.99}$	86.25 ± 01.89	$\mathbf{93.67 \pm 01.80}$
Caltech101, 2386					
$\mathbf{93.26 \pm 01.24}$	92.47 ± 01.26	$\mathbf{89.85 \pm 01.05}$	87.44 ± 01.25	$\mathbf{93.49 \pm 00.91}$	92.15 ± 01.20
CUB , 600					
91.33 ± 02.56	$\mathbf{91.58 \pm 01.91}$	$\mathbf{72.50 \pm 01.66}$	70.25 ± 03.83	89.08 ± 01.98	$\mathbf{90.67 \pm 01.89}$
CNIST, 1200					
40.46 ± 02.01	$\mathbf{43.17 \pm 02.24}$	11.50 ± 01.83	$\mathbf{32.00 \pm 02.57}$	13.78 ± 01.55	$\mathbf{41.50 \pm 03.46}$

Table 4.4: Performance comparison between normal data and data with Uniform noise

Additionally, a comparative analysis between the proposed method and the baseline was conducted on the conflictive dataset. The results indicate that the performance in this setting is not consistent across datasets. This inconsistency can be attributed to the methodology employed in generating the conflictive dataset. Specifically, the process by which conflict was introduced by the author may not accurately reflect realistic or semantically meaningful interview disagreements, thereby limiting the reliability of the evaluation in this scenario.

Chapter 5

Conclusion and Future Work

In this work, we have studied the key features and challenges of multiview learning. We discussed the uncertainty computation using evidential theory. We further discussed two methods for view-specific opinion aggregation. A main contribution of our work is the theoretical proof that aggregating opinions in a specific order-starting with the most uncertain views-results in the lowest overall uncertainty. This leads to more reliable decisions and helps reduce conflicts, especially when the data contains noise. Our experiments across several datasets show that this specific fusion order improves classification accuracy in most cases. We argue that the methodology used to construct the conflictive data set in the RCML paper lacks a strong logical foundation. Developing a more meaningful approach to generating multiview conflictive data is an important direction for future work. In future, this can be extended to various application areas.

Bibliography

[1] E. Keogh and A. Mueen, "Curse of dimensionality," in *Encyclopedia of machine learning*, pp. 257–258, Springer, 2011.

[2] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surfaceelectromyography-based gesture recognition by multi-view deep learning," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 10, pp. 2964–2973, 2019.

[3] X. Tian, Z. Deng, W. Ying, S. Choi, K, D. Wu, B. Qin, J. Wang, H. Shen, and S. Wang, "Deep multi-view feature learning for eeg-based epileptic seizure detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 1962–1972, 2019.

[4] Y. Kong, Z. Ding, J. Li, and Y. Fu, "Deeply learned view-invariant features for cross-view action recognition.," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 3028–3037, 2017.

[5] S. Sun, W. Dong, and Q. Liu, "Multi-view representation learning with deep gaussian processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4453–4468, 2021.

[6] T. Q. Zhang C, Cheng J, "Multi-view image classification with visual, semantic and view consistency," *IEEE Transactions on Image Processing*, vol. 29, pp. 617–627, 2020.

[7] X. C. Xu C, Tao D, "A survey on multi-view learning," *arXiv preprint arXiv*, vol. 1304.5634.

[8] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," *In Advances in Neural Information Processing Systems*, vol. 31, 2018.

[9] W. Liu, Y. Chen, and T. Denoeux, "Trusted multi-view deep learning with opinion aggregation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, pp. 7585–7593, 2022.

[10] X. Cai, S. Jiajun, G. Ziyu, Z. Wei, W. Yue, and G. Xiyue, "Reliable conflictive multi-view learning," *Association for the Advancement of Artificial Intelligence*, 2024.