Applied Soft Computing Journal 171 (2025) 112757

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Three-way multi-label classification: A review, a framework, and new challenges

Yuanjian Zhang [a], Tianna Zhao [b,c,d,*], Duoqian Miao [e,f], Yiyu Yao [g]

[a] School of Computer Engineering and Science, Shanghai University, 200444, China
[b] Institute of Artificial Intelligence on Education, Shanghai Normal University, 200234, China
[c] The Research Base of Online Education for Shanghai Middle and Primary Schools, Shanghai Normal University, 200234, China
[d] Shanghai Engineering Research Center of Intelligent Education and Big data, Shanghai Normal University, 200234, China
[e] Department of Computer Science and Technology, Tongji University, Shanghai, 201804, China
[f] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, 201804, China
[g] Department of Computer Science, University of Regina, Regina, SK S4S 0A2, Canada

ARTICLE INFO

Keywords:
Three-way decision
Decision-theoretic rough set
Sequential three-way decision
Uncertainty
Multi-label classification

ABSTRACT

The multi-label classification task is more challenging than the degenerated case of single-label classification due to diversified uncertainty. Uncertainty in multi-label classification not only involves label dependency but also the inconsistency of label correlation and imbalanced label association. While three-way decision methods excel in characterizing multifaceted uncertainty, developing a well-established three-way decision-based framework for multi-label classification remains challenging. Based on the historical developments of decision-theoretic rough sets and sequential three-way decision, this paper presents a systematic review of representative three-way multi-label classification models. By analyzing the contributions from existing studies, the multifaceted uncertainty of multi-label classification is classified into label uncertainty, correlation uncertainty, and structure uncertainty. To effectively deal with the structure uncertainty, some essential subproblems are identified, and a general three-way-based framework called the multi-label sequential decision-theoretic three-way decision (ML-SD3WD) model is presented. The ML-SD3WD model sequentially handles three-way topic generation, three-way label assignment, and three-way label enhancement by integrating decision-theoretic rough set with sequential three-way decision. Furthermore, some emerging directions in customizing the ML-SD3WD model are delineated. Finally, limitations from both the label side and feature side are discussed and corresponding solutions are offered for uncertainty-driven solutions in practical applications. The results of this review will offer a road map for knowledge discovery in multi-label classification.

Contents

* Corresponding author.
E-mail address: zhaotianna@shnu.edu.cn (T. Zhao).

https://doi.org/10.1016/j.asoc.2025.112757
Received 3 August 2024; Received in revised form 30 November 2024; Accepted 11 January 2025
Available online 18 January 2025

1. Introduction

The ability to leverage underlying knowledge for precision classification on unseen instances has been deemed as a manifestation of human intelligence. To evolve such intelligence on machines, many authors consider concept cognition problems and learn the intension of concepts based on extension information annotated by labels. The generalization of label semantics in scale and relationship has witnessed the extension from single-label classification to multi-label classification [1–3]. Contrary to the single-label case in which concepts are mutually exclusive, the multi-label case (e.g., UCO repository[1]) embraces varying degrees of inconsistency among concepts. The uncertain overlapping nature of concepts is especially ubiquitous in domains including emotion analysis [4,5], video annotation [6,7], and medical diagnosis [8–10].

Label correlation is an essential uncertainty component in developing a multi-label classifier. The reasons are two-fold. Firstly, the annotator may prefer some reliable evidence from the combinations of salient labels, which means the relevancy/irrelevancy of a particular instance-label pair may be referred to when determining another. Secondly, the learner expects to find some data-driven information to resist the overwhelming imbalanced label relevancy and the interpretation of label correlation serves as an auxiliary. During the past decades, the correlated order, spatial representation, and regional boundary constitutes three main directions on label correlation construction, and three trends have correspondingly emerged on label correlation, i.e., from hyperparameter-specified order (including first-order label correlation [11–13], second-order label correlation [14–17], high-order label correlation [18–21]) to a judicious order of label correlation [22–24], from observable feature-label space [25,26] to latent feature-label space [27–29], and from solely global label correlation [30–32] to solely local, both global and local label correlation [33–37]. Despite the increasingly refined label correlations, the robustness of label correlation is still flawed, especially for the case of label distribution learning [38] and label enhancement [39], where numerical labels are predefined or learned from logical labels, respectively. Accordingly, the uncertainty of label correlation remains unsolved and degenerates the classification performance.

Granular computing (GrC) [40–42] is a methodology that approximates the underlying structure of knowledge by granulation. Components in granulation are not limited to a specific mathematical model such as rough set [43], fuzzy rough set [44], intuitionistic fuzzy set [45], evidential theory [46], formal concept analysis [47], and shadowed set [48]. Rather, they are determined by both the characteristics of the data and the identified complexity of uncertainty. These factors, influenced by objective and subjective considerations, lead to a diverse range of granule-based solutions. As the understanding of multi-label semantics deepens, the intensity and scope of multi-faceted uncertainty, including randomness, roughness, and fuzziness, are gradually reduced. For example, Li et al. [49] optimized feature representations via the maximal correlation minimal redundancy principle. Che et al. [50] generated label-specific features by measuring overlapping degrees on label-dependent essential features. Qian et al. [51] extended the neighborhood granularity to solve feature selection in the label distribution learning case. Tan et al. [52] constructed the cost-sensitive features by introducing granular discrimination-based regularization. Lin et al. [53] optimized the common features and label-specific features by introducing mutual information into label correlation. Bian et al. [54] constructed an adaptive membership function to overcome the numeric underflow issue in fuzzy systems for the high-dimensional multi-label classification. Liu et al. [55] introduced a class-imbalance-aware fuzzy information entropy on feature selection for adaptive label enhancement. Zhao et al. [56] developed augmented features by learning quantitative discrimination within the neighborhood.

Three-way decision (3WD) [57] is a field of study related to the theory of granular computing. Originating from semantic explanations of three regions deduced from the rough set, the semantics of threes and decisions in three-way decision have been substantially enriched. A theory of three-way decision has been evolved from the framework of trisecting-and-acting [58], trisecting-acting-outcome (TAO) [59], to the triading-acting-optimizing (TAO) [60] (see Fig. 1). A recent review [61] concludes the development of three-way decision from the philosophy-theory-application triad and who-what-when triad. It is worth mentioning that the TAO structure is problem-dependent in characterizing and reasoning the degrees of uncertainty. This means that the TAO structure adapts to the specific information distribution related to the certainty and uncertainty of the target concepts, allowing for more accurate and context-specific uncertainty management. Consequently, the three-way decision model is conducive to minimizing uncertainty while maintaining interpretability. The superiority of three-way decision has been demonstrated in both theoretical researches and practical explorations, including movement-based TAO model [62–64], change-based TAO model [65–67], uncertainty measure [68,69], concept cognition [70,71], granular ball computing [72–74], conflict analysis [75–77], incremental learning [78,79], outlier

[1] https://www.uco.es/kdis/mllresources/

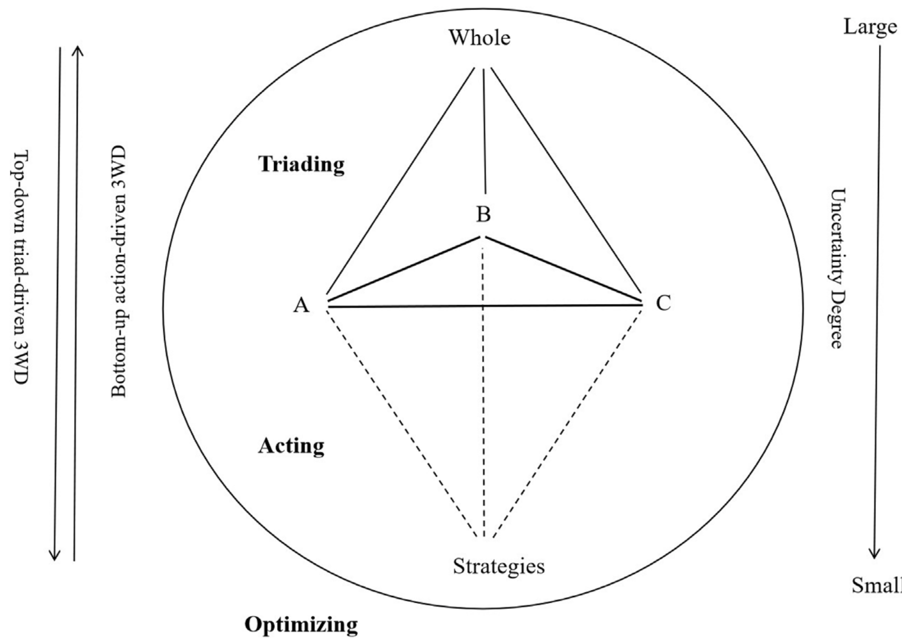

Fig. 1. Illustration of TAO framework.

detection [80,81], clustering analysis [82–84], pearson search [85–87], and privacy-preserving modeling [88–90].

Multi-label classification has become a major topic of three-way decision over decades [61]. Recently, many scholars have employed three-way decision to address the uncertainty in multi-label classification and obtain promising outcomes. For incremental learning, Zhang et al. [91] accelerated the calculation of tripartition for multi-label in the arrival of multiple instances. Liu et al. [92] presented a novel instance similarity distribution-based neighborhood rough set for streaming feature selection. Deng et al. [93] developed an F-fuzzy condition entropy measure to dynamically update the features characterized by an F-double-fuzzy rough set. For active learning, Zhang et al. [94] proposed a granule-based batch learning model and improved the instance selection quality by considering three-way label correlation fitting. For imperfect information processing, Zhu et al. [95] devised an active three-way clustering method to polish the local label correlation for instances with weak labels. Qian et al. [96] developed a three-way label fusion strategy to synthesize the intermediate outputs from tri-training, thus reducing the biased estimation incurred by insufficient consistent instances for partial multi-label classification. Yin et al. [97] introduced multi-scale transformation principle on label space to capture the characteristics fine-grained features for missing label cases.

1.1. Motivation

Numerous extensions from single-label-based three-way decision flourish the solutions of multi-label classification. However, a comprehensive survey that reviews the advancements and accomplishments of three-way multi-label classification is lacking. The detailed motivations on why this survey paper is imperative are as follows.

1. **Complex Uncertainty in Multi-label Classification:** The categories of uncertainty in multi-label classification include more than randomness, fuzziness and roughness. The label ambiguity is compounded by enriched semantics of inconsistency, imbalance, and inaccuracy, which is beyond the scope of the single-label-based uncertainty.
2. **New Challenges in Handling Diversified Uncertainty:** With the diversified uncertainty in observable labels, it is impossible to employ the conventional divide-and-conquer strategy to solve a series of subproblems, which means the label correlation becomes a novel knowledge discovery issue. The classification accuracy will be suboptimal if the structures of concept remains uncertain. To effectively deal with the multi-faceted uncertainty, it is valuable to clarify the representative procedures and analyze the corresponding topology of the TAO model, and a feasible solution for this purpose is conducting a comprehensive review on three-way multi-label classification.
3. **Construction on General Modeling Framework:** As the specialized solutions are unavailable, studies taking model adaptation on three-way multi-label classification are reviewed. The essential functionalities and extension directions are clarified after summarizing the progress in decision-theoretic rough set and sequential three-way decision. Based on these developments, the components of uncertainty in multi-label classification are summarized. With these requirements on uncertainty formulation, a general modeling framework is established. It is crucial for guiding future research and practical applications, ensuring that the models are adaptable and robust across different scenarios.
4. **Supporting Application-Oriented Knowledge Discovery:** To elaborate on how three-way-based model promote the researches in concrete applications, studies from concept-cognitive learning, emotion detection, medical diagnosis, recommender systems, and social network analysis, are critically analyzed. Meanwhile, the limitations and solutions of the proposed framework are discussed, thus inspiring an uncertainty-driven solution.

Despite the challenges and limitations, the proposed framework serves as the backbone of the three-way multi-label classification approach.

1.2. Contribution

By systematically outlining the model developments and highlighting the representative models, this survey can serve as a reference for practitioners in the relevant community to further develop more effective three-way-based solutions for multi-label classifications. The contributions are as follows:

1. A historical review of three-way decision is completed on representative extensions of decision-theoretic rough set and sequential three-way decision from single-label to multi-label. Through

comprehensive investigations on problem formulation, the adaptation of uncertainty-aware modules in three-way-based models are intensively compared.

2. The uncertainty factors in constructing three-way-based multi-label classification are identified as label uncertainty, correlation uncertainty, and structure uncertainty. They denote three kinds of uncertainty distributions, known as the trustworthiness of multiple labels, the topic-based positive and negative correlations among labels, and the granulation strategy, respectively. For comprehension, a hierarchical structure representing the multi-faceted uncertainty as data level, knowledge level, and formulation level, is developed.
3. By leveraging the advantages of the decision-theoretic rough set and sequential three-way decision, a novel three-way-based framework called multi-label sequential decision-theoretic three-way decision (ML-SD3WD) is established to reduce the uncertainty in essential procedures, including topic generation, label assignment, and label enhancement.
4. Challenges in customizing the ML-SD3WD model for multi-label classification are comprehensively discussed, with a focus on the limitations and corresponding solutions of employing three-way decision in real applications.

The remaining parts are organized as follows: the single-label-based prototypes of three-way decision are reviewed in Section 2, which provides the theoretical foundation for three-way-based multi-label classification. Section 3 summarizes the bibliometric analysis on recent studies and three major model adaptation directions, offering valuable insights into the structural uncertainty. Section 4 details the challenges incurred by various kinds of uncertainty and explores a trilevel solution by structural formulation. Section 5 points out some emerging applications, promoting discussions on domain-specific data-driven formulations. Section 6 discusses the limitations and corresponding solutions when employing three-way classification on real application. Finally, Section 7 concludes the paper.

2. 3WD in single-label classification

The decision-theoretic rough set (DTRS) and sequential three-way decision (S3WD) models constitute the foundations of three-way decision and serve as the prototypes of many variations. In this section, the essential components of DTRS and S3WD in binary classification are reviewed. Representative extensions of DTRS and S3WD in multi-class classification are provided, where the concept relations are relaxed from complementary to mutually exclusive.

2.1. DTRS for binary classification

The motivation of decision-theoretic rough set proposed by Yao et al. [98,99] is to offer a reasonable interpretation of how a rough concept X is approximated by three non-overlapping regions from the perspective of the misclassification cost in decision-making. Let $\mathcal{A} = (U, R, Pr)$ denote a triple of available information, where U represents the universe, R represents the equivalence relation that partitions U into a family U/R of nonempty subsets of U. Let $[x] \in U/R$ denote the equivalence class in the partition U/R, which is induced by an object $x \in U$ and $Pr(X\,|[x])$ represents a probability-based measure for $X \subseteq U$ on σ-algebra of U. Suppose the loss function λ_{iP} ($i \in \{P, B, N\}$) denotes potential cost incurred for taking the ith action when an element $x \in [x]$ belongs to X (i.e., $x \in X$), and λ_{iN} ($i \in \{P, B, N\}$) denotes potential cost incurred for taking the ith action when an element $x \in [x]$ belongs to the complementary of X (i.e., $x \in \neg X$), then for the binary classification problem, the six loss functions can be placed as Table 1.

Correspondingly, the expected conditional risk $R(P\,|[x])$, $R(B\,|[x])$, $R(N\,|[x])$ of determining $[x]$ as positive region (denoted as $POS(X)$), boundary region (denoted as $BND(X)$) and negative region (denoted as $NEG(X)$) can be expressed by following Bayesian decision theory [100–102] as:

Table 1
Loss function matrix.

	a_P	a_B	a_N
X	λ_{PP}	λ_{BP}	λ_{NP}
$\neg X$	λ_{PN}	λ_{BN}	λ_{NN}

$$\begin{aligned} R(P\,|[x]) &= \lambda_{PP} Pr(X\,|[x]) + \lambda_{PN} Pr(\neg X\,|[x]), \\ R(B\,|[x]) &= \lambda_{BP} Pr(X\,|[x]) + \lambda_{BN} Pr(\neg X\,|[x]), \\ R(N\,|[x]) &= \lambda_{NP} Pr(X\,|[x]) + \lambda_{NN} Pr(\neg X\,|[x]). \end{aligned} \tag{1}$$

where $Pr(X\,|[x]) + Pr(\neg X\,|[x]) = 1$. In this way, the relative difference in decision-making is cost-sensitive. The optimal tripartition on X and the complementary $\neg X$ is formulated as a problem that simultaneously minimizes expected conditional risks for the three regions as:

$$\begin{aligned} &(P) \quad Decide \quad x \in POS(X) \quad if \quad Pr(X\,|[x]) \geqslant \alpha \quad \wedge \quad Pr(X\,|[x]) \geqslant \gamma, \\ &(B) \quad Decide \quad x \in BND(X) \quad if \quad \beta < Pr(X\,|[x]) < \alpha, \\ &(N) \quad Decide \quad x \in NEG(X) \quad if \quad Pr(X\,|[x]) \leqslant \beta \quad \wedge \quad Pr(X\,|[x]) \leqslant \gamma. \end{aligned} \tag{2}$$

where α, β, and γ are resolved as:

$$\begin{aligned} \alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \\ \beta &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}, \\ \gamma &= \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}. \end{aligned} \tag{3}$$

2.2. DTRS for multi-class classification

The extensions of DTRS in multi-class classification (abbreviated as m-class, where $m > 2$) emphasize the problem transformation on both class relations and loss functions [103–106]. By referring to problem transformation, the new formulation is composed of a number of subproblems in the form of classical DTRS. Let $\Omega = \{C_1, C_2, \ldots, C_m\}$ be a finite set of m classes, for the three actions named as acceptance as the class C_i (denoted as P_i), postpone as the class C_i (denoted as B_i), and reject as the class C_i (denoted as N_i), respectively. Then the DTRS in multi-class can be formulated as follows:

(1) Model of Zhou [104]

In the multi-class classification based on decision-theoretic rough set model, the tripartition takes a one-versus-rest strategy where each class C_i is associated with a set of strategies composed of three actions $\{P_i, B_i, N_i\}$. In other words, there are $3 \times m$ loss functions, where P_i, B_i, and N_i correspond to the decision $[x] \subseteq POS(C_i)$, $[x] \subseteq BND(C_i)$, $[x] \subseteq NEG(C_i)$, respectively. Then the expected conditional risk for acceptance (denoted as $R(P_i\,|[x])$), neutral (denoted as $R(B_i\,|[x])$), and rejection (denoted as $R(N_i\,|[x])$) can be expressed as:

$$\begin{aligned} R(P_i\,|[x]) &= \sum_{j=1}^{j=m} \lambda\left(P_i \middle| C_j\right) Pr(C_j\,|[x]), \\ R(B_i\,|[x]) &= \sum_{j=1}^{j=m} \lambda\left(B_i \middle| C_j\right) Pr(C_j\,|[x]), \\ R(N_i\,|[x]) &= \sum_{j=1}^{j=m} \lambda\left(N_i \middle| C_j\right) Pr(C_j\,|[x]). \end{aligned} \tag{4}$$

where $\lambda\left(P_i \middle| C_j\right)$, $\lambda\left(B_i \middle| C_j\right)$, and $\lambda\left(N_i \middle| C_j\right)$ are the loss incurred by making acceptance, neutral or rejection of class C_i when the actual class is C_j, $Pr(C_j\,|[x])$ is the probability of an instance x being C_j conditioned on the equivalence $[x]$. Thus the tripartition on the concept

X can be deduced by following Bayesian decision theory [100–102] as:

$$
\begin{aligned}
&(P) \quad Decide \quad x \in POS\left(C_i\right) \quad if \quad Pr\left(C_i \mid [x]\right) \geqslant \alpha_i \wedge Pr\left(C_i \mid [x]\right) \geqslant \gamma_i,\\
&(B) \quad Decide \quad x \in BND\left(C_i\right) \quad if \quad Pr\left(C_i \mid [x]\right) < \alpha_i \wedge Pr\left(C_i \mid [x]\right) > \beta_i,\\
&(N) \quad Decide \quad x \in NEG\left(C_i\right) \quad if \quad Pr\left(C_i \mid [x]\right) \leqslant \beta_i \wedge Pr\left(C_i \mid [x]\right) \leqslant \gamma_i.
\end{aligned} \tag{5}
$$

where α_i, β_i, and γ_i are resolved as:

$$
\begin{aligned}
\alpha_i &= \frac{\sum_{j=1, j\neq i}^{j=m} Pr\left(C_j \mid [x]\right)\left(\lambda\left(P_i \middle| C_j\right) - \lambda\left(B_i \middle| C_j\right)\right)}{\lambda\left(B_i \middle| C_j\right) - \lambda\left(P_i \middle| C_j\right)},\\
\beta_i &= \frac{\sum_{j=1, j\neq i}^{j=m} Pr\left(C_j \mid [x]\right)\left(\lambda\left(B_i \middle| C_j\right) - \lambda\left(N_i \middle| C_j\right)\right)}{\lambda\left(N_i \middle| C_j\right) - \lambda\left(B_i \middle| C_j\right)},\\
\gamma_i &= \frac{\sum_{j=1, j\neq i}^{j=m} Pr\left(C_j \mid [x]\right)\left(\lambda\left(P_i \middle| C_j\right) - \lambda\left(N_i \middle| C_j\right)\right)}{\lambda\left(N_i \middle| C_j\right) - \lambda\left(P_i \middle| C_j\right)}.
\end{aligned} \tag{6}
$$

(2) Model of Jia et al. [105]

In the multi-class three-way decision-theoretic rough set (MCDTRS) model, the tripartition takes a one-versus-rest strategy where each class C_i is associated with a set of strategies composed of three actions $\mathcal{A} = \{a_{P_i}, a_{B_i}, a_{N_i}\}$. Different from Zhou's model [104] whose loss functions include $3\times m$ combinations, MCDTRS is composed of $6\times m$ combinations for m-class. As shown in (7), the expected conditional risk for positive, boundary, and negative class on ith class is calculated by Bayesian decision theory [100–102].

$$
\begin{aligned}
R_{P_i} &= \lambda_{PP}^i Pr\left(C_i \mid [x]\right) + \lambda_{PN}^i \left(1 - Pr\left(C_i \mid [x]\right)\right),\\
R_{B_i} &= \lambda_{BP}^i Pr\left(C_i \mid [x]\right) + \lambda_{BN}^i \left(1 - Pr\left(C_i \mid [x]\right)\right),\\
R_{N_i} &= \lambda_{NP}^i Pr\left(C_i \mid [x]\right) + \lambda_{NN}^i \left(1 - Pr\left(C_i \mid [x]\right)\right).
\end{aligned} \tag{7}
$$

where $Pr\left(C_i \mid [x]\right)$ denotes the probability of x being the class of C_i. The six loss functions λ_{PP}^i, λ_{PN}^i, λ_{BP}^i, λ_{BN}^i, λ_{NP}^i, λ_{NN}^i, represent the loss incurred by accepting (P), delaying (B), and rejecting (N) x as ith class when the x actually belongs to the ith class (P) or not (N). Similarly, the decision on x follows the principle of risk minimization:

$$
\begin{aligned}
&(P) \quad Decide \quad x \in POS\left(C_i\right) \quad if \quad Pr\left(C_i \mid [x]\right) \geqslant \alpha_i,\\
&(B) \quad Decide \quad x \in BND\left(C_i\right) \quad if \quad \beta_i < Pr\left(C_i \mid [x]\right) < \alpha_i,\\
&(N) \quad Decide \quad x \in NEG\left(C_i\right) \quad if \quad Pr\left(C_i \mid [x]\right) \leqslant \beta_i.
\end{aligned} \tag{8}
$$

where the α_i and β_i are determined as:

$$
\begin{aligned}
\alpha_i &= \frac{\lambda_{PN}^i - \lambda_{BN}^i}{\left(\lambda_{PN}^i - \lambda_{BN}^i\right) + \left(\lambda_{BP}^i - \lambda_{PP}^i\right)},\\
\beta_i &= \frac{\lambda_{BN}^i - \lambda_{NN}^i}{\left(\lambda_{BN}^i - \lambda_{NN}^i\right) + \left(\lambda_{NP}^i - \lambda_{BP}^i\right)}.
\end{aligned} \tag{9}
$$

2.3. S3WD for binary classification

The motivation of S3WD presented by Yao and Deng [107,108] is to formulate a top-down information processing solution that can be examined at a finer granulation level with more detailed information when there is a need or benefit for doing so [109]. This means the roughness of a given concept X can be gradually resolved and the three-way decision is degenerated as a two-way decision eventually (see Fig. 2).

Let $\{0, 1, \ldots, n\}$ denotes $n+1$ ($n \geqslant 1$) levels of granularity, with 0 and n representing the finest and coarsest granularity, respectively. The decision-making procedure performs sequentially at levels $n, n-1, \ldots, 1$ and eventually reaches the two-way decision at level 0. For the ith level, the descriptor of the instance x ($Des_i(x)$) follows the assumption below:

$$
Des_0(x) \preceq Des_1(x) \preceq \ldots \preceq Des_i(x) \preceq \ldots \preceq Des_n(x) \tag{10}
$$

where the relation $\preceq$ denotes a "finer than" relationship for two descriptors at different granularity. It is worth mentioning that the tripartition at a particular level can be any concrete three-way model like DTRS. Let U_{i+1} be the set of instances in boundary region at level $i+1$ and stipulate $U_{n+1} = U$, then the general form of S3WD at a particular level can be expressed as:

$$
\begin{aligned}
POS_{(\alpha_i,\beta_i)}\left(v_i\right) &= \left\{x \in U_{i+1} \middle| v_i\left(Des_i(x)\right) \succeq_i \alpha_i\right\},\\
BND_{(\alpha_i,\beta_i)}\left(v_i\right) &= \left\{x \in U_{i+1} \middle| \beta_i \prec_i v_i\left(Des_i(x)\right) \prec_i \alpha_i\right\},\\
NEG_{(\alpha_i,\beta_i)}\left(v_i\right) &= \left\{x \in U_{i+1} \middle| v_i\left(Des_i(x)\right) \preceq_i \beta_i\right\}.
\end{aligned} \tag{11}
$$

where $POS_{(\alpha_i,\beta_i)}\left(v_i\right)$, $BND_{(\alpha_i,\beta_i)}\left(v_i\right)$, and $NEG_{(\alpha_i,\beta_i)}\left(v_i\right)$ denote the positive, boundary and negative regions at ith step. $v_i(\cdot)$ is an evaluation function that measures the relationship between its value and the corresponding three-way-based thresholds (α_i and β_i, which satisfies $\beta_i \prec_i \alpha_i$). For level 0, the evaluation function $v_i(\cdot)$ compares its value with the two-way-based threshold γ_0 on U_1.

2.4. S3WD for multi-class classification

The extensions of S3WD on multi-class classification [110–116] explore different combinations between the instance descriptor ($Des_i(x)$) and the structure of thresholds $\left(\alpha_i, \beta_i\right)$ at every level of granularity. Nevertheless, they all share the three-way-based decision-making structures with the refinement of granularity[2]. In what follows, the structures of three representatives are enumerated to show how the extensions of S3WD are completed.

(1) Model of Savchenko [110]

In the sequential database enumeration with the sequential three-way decision model, the S3WD is employed for statistical recognition of the objects. With the refinement of the grid, the method constructed a group of segmentations with identical sizes and generated the instance descriptor by calculating gradient orientations based on Kullback–Leibler divergence (a.k.a. KL-divergence). To minimize ambiguity within classes, the tripartition at the ith granularity is realized in a one-versus-one strategy by comparing the KL-divergence (denoted as $\rho_{KL}^{(i)}\left(X_1, X_2\right)$) from different classes against the thresholds $\left(\rho_0^{(i)}, \rho_1^{(i)}\right)$.

$$
\begin{aligned}
POS_{\rho_1^{(i)}} &= \left\{\rho_{KL}^{(i)}\left(X_1, X_2\right) \middle| X_1, X_2 \in \mathbf{X}, \rho_{KL}^{(i)}\left(X_1, X_2\right) > \rho_1^{(i)}\right\},\\
BND_{\rho_0^{(i)};\rho_1^{(i)}} &= \left\{\rho_{KL}^{(i)}\left(X_1, X_2\right) \middle| X_1, X_2 \in \mathbf{X}, \rho_0^{(i)} \leqslant \rho_{KL}^{(i)}\left(X_1, X_2\right) \leqslant \rho_1^{(i)}\right\},\\
NEG_{\rho_0^{(i)}} &= \left\{\rho_{KL}^{(i)}\left(X_1, X_2\right) \middle| X_1, X_2 \in \mathbf{X}, \rho_{KL}^{(i)}\left(X_1, X_2\right) < \rho_0^{(i)}\right\}.
\end{aligned} \tag{12}
$$

The model determines the classification at level 0 by selecting the least unreliable level l^* from the boundary region among all n levels.

$$
l^* = \underset{i\in\{1,2,\ldots,n\}}{\arg\max}\ \hat{Pr}^{(i)}\left(W_{v^{(i)}} \mid X\right) \tag{13}
$$

where

$$
\hat{Pr}^{(i)}\left(W_{v^{(i)}} \mid X\right) = \frac{\exp\left(-\rho_{KL}^{(i)}\left(X^{(i)}, X_{v^{(i)}}^{(i)}\right)\right)}{\sum_{r=1}^{R} \exp\left(-\rho_{KL}^{(i)}\left(X^{(i)}, X_r^{(i)}\right)\right)}
$$

denotes the estimation of the posterior probability.

(2) Model of Qian et al. [111]

In the variable multigranulation sequential three-way decision based on multiple thresholds (VMS3WDMT) model, the S3WD is introduced to extend a variable multigranulation rough set. For the hth level of granularity, the tripartition takes the one-versus-one strategy with the fixed instance descriptor (denoted as $Pr\left(D_q^h \middle| [x]_{E_i}\right)$) and granule-dependent thresholds. Concretely, it determines the positive, boundary

[2] The two-way decision module is optional for S3WD if the non-commitment case is allowed. Therefore, the S3WD generalizes as a multi-stage version of conventional three-way-based solutions (e.g., DTRS).

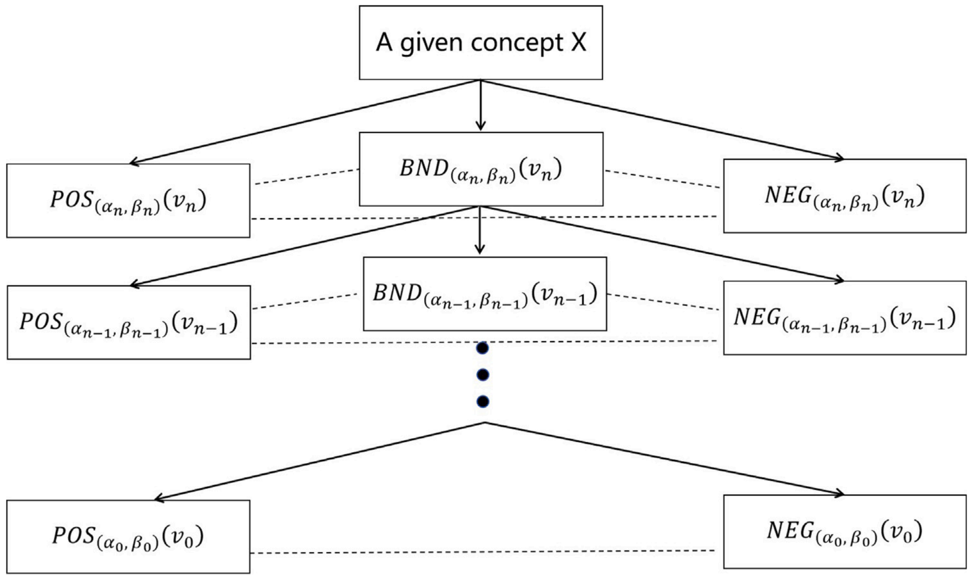

Fig. 2. Illustration of sequential three-way decision.

and negative class on the qth class (denoted as D_q^h) if the corresponding conditional probabilities of more than δ-percentage of equivalence classes of the qth class are greater than the threshold α_i^h, smaller than the threshold β_i^h or otherwise. Particularly, VMS3WDMT degenerates as optimistic multigranulation S3WD (OMS3WDMT) if $\delta = \frac{1}{m}$, and pessimistic multigranulation S3WD (PMS3WDMT) if $\delta = 1$.

$$\begin{aligned}
POS_{\mathbb{E}}^{(\alpha^h,\beta^h)}\left(D_q^h\right) &= \bigcup_{\mathbb{T}\subseteq\mathbb{E}\wedge|\mathbb{T}|/m\geqslant\delta} \left\{ x\in U^h \,\middle|\, \bigwedge_{E_i\in\mathbb{T}} Pr\left(D_q^h \,\middle|\, [x]_{E_i}\right) \geqslant \alpha_i^h \right\},\\
BND_{\mathbb{E}}^{(\alpha^h,\beta^h)}\left(D_q^h\right) &= \bigcup_{\mathbb{T}\subseteq\mathbb{E}\wedge|\mathbb{T}|/m\geqslant\delta} \left\{ x\in U^h \,\middle|\, \bigwedge_{E_i\in\mathbb{T}} \beta_i^h < Pr\left(D_q^h \,\middle|\, [x]_{E_i}\right) < \alpha_i^h \right\},\\
NEG_{\mathbb{E}}^{(\alpha^h,\beta^h)}\left(D_q^h\right) &= \bigcup_{\mathbb{T}\subseteq\mathbb{E}\wedge|\mathbb{T}|/m\geqslant\delta} \left\{ x\in U^h \,\middle|\, \bigwedge_{E_i\in\mathbb{T}} Pr\left(D_q^h \,\middle|\, [x]_{E_i}\right) \leqslant \beta_i^h \right\}.
\end{aligned} \tag{14}$$

where $U^h = POS_{\mathbb{E}}^{\left(\alpha^{h-1},\beta^{h-1}\right)}\left(D_q^{h-1}\right) - BND_{\mathbb{E}}^{\left(\alpha^{h-1},\beta^{h-1}\right)}\left(D_q^{h-1}\right)$ represents that given a family of equivalence relation (i.e., $\mathbb{E}$), the hth processing from the uncertain component of universe U at step $h-1$. The model does not require bipartition at level 0 as it focuses on the variation of tripartition rather than classification.

(3) Model of Xu et al. [113]

Three DTRS-based S3WD models are formulated by taking one-versus-one (OVO), one-versus-rest (OVR), and one-versus-multiple (OVM) strategies. The formulations of one-versus-rest and one-versus-multiple at ith level of granularity are similar in both instance descriptor $\left(Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right)\right)$ and structures of thresholds $\left(\alpha_i^j,\beta_i^j\right)$, however the loss functions $(\lambda_{\cdot\cdot})$ are class-dependent. The three-way structure of OVR and OVM are as follows:

$$\begin{aligned}
POS\left(D_i^j\right) &= \left\{ x\in U_i \,\middle|\, Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right) \geqslant \alpha_i^j \right\},\\
BND\left(D_i^j\right) &= \left\{ x\in U_i \,\middle|\, \beta_i^j < Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right) < \alpha_i^j \right\},\\
NEG\left(D_i^j\right) &= \left\{ x\in U_i \,\middle|\, Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right) \leqslant \beta_i^j \right\}.
\end{aligned} \tag{15}$$

where U_i is the ith processing from the uncertain component of universe U at step $i-1$. For the OVR model that takes the one-versus-rest strategy, the thresholds $\left(\alpha_i^j,\beta_i^j\right)$ are determined as:

$$\begin{aligned}
\alpha_i^j &= \frac{\lambda^j_{P_i\neg D_i^j} - \lambda^j_{B_i\neg D_i^j}}{\left(\lambda^j_{P_i\neg D_i^j} - \lambda^j_{B_i\neg D_i^j}\right) + \left(\lambda^j_{B_iD_i^j} - \lambda^j_{P_iD_i^j}\right)},\\
\beta_i^j &= \frac{\lambda^j_{B_i\neg D_i^j} - \lambda^j_{N_i\neg D_i^j}}{\left(\lambda^j_{B_i\neg D_i^j} - \lambda^j_{N_i\neg D_i^j}\right) + \left(\lambda^j_{N_iD_i^j} - \lambda^j_{B_iD_i^j}\right)}.
\end{aligned} \tag{16}$$

For the OVM model that takes the one-versus-multiple strategy, the thresholds $\left(\alpha_i^j,\beta_i^j\right)$ are determined via Bayesian decision theory [100–102].

$$\begin{aligned}
\alpha_i^j &= \frac{\sum_{k=1,k\neq j}^{k=m} Pr\left(D_i^k \,\middle|\, [x]_{C_i}\right)\left(\lambda^j_{P_iD_i^k} - \lambda^j_{B_iD_i^k}\right)}{\lambda^j_{B_iD_i^j} - \lambda^j_{P_iD_i^j}},\\
\beta_i^j &= \frac{\sum_{k=1,k\neq j}^{k=m} Pr\left(D_i^k \,\middle|\, [x]_{C_i}\right)\left(\lambda^j_{B_iD_i^k} - \lambda^j_{N_iD_i^k}\right)}{\lambda^j_{N_iD_i^j} - \lambda^j_{B_iD_i^j}}.
\end{aligned} \tag{17}$$

For the one-versus-one strategy, the structures of threshold become $\left(\alpha_i^{jk},\beta_i^{jk}\right)$ while the instance descriptor ($Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right)$) remains unchanged.

$$\begin{aligned}
POS\left(D_i^{jk}\right) &= \left\{ x\in U_i^{jk} \,\middle|\, Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right) \geqslant \alpha_i^{jk} \right\},\\
BND\left(D_i^{jk}\right) &= \left\{ x\in U_i^{jk} \,\middle|\, \beta_i^{jk} < Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right) < \alpha_i^{jk} \right\},\\
NEG\left(D_i^{jk}\right) &= \left\{ x\in U_i^{jk} \,\middle|\, Pr\left(D_i^j \,\middle|\, [x]_{C_i}\right) \leqslant \beta_i^{jk} \right\}.
\end{aligned} \tag{18}$$

where U_i^{jk} represents the instances with the jth and kth class at the ith processing from the uncertain component of universe U at step $i-1$, the thresholds $\left(\alpha_i^{jk},\beta_i^{jk}\right)$ are determined by Bayesian decision

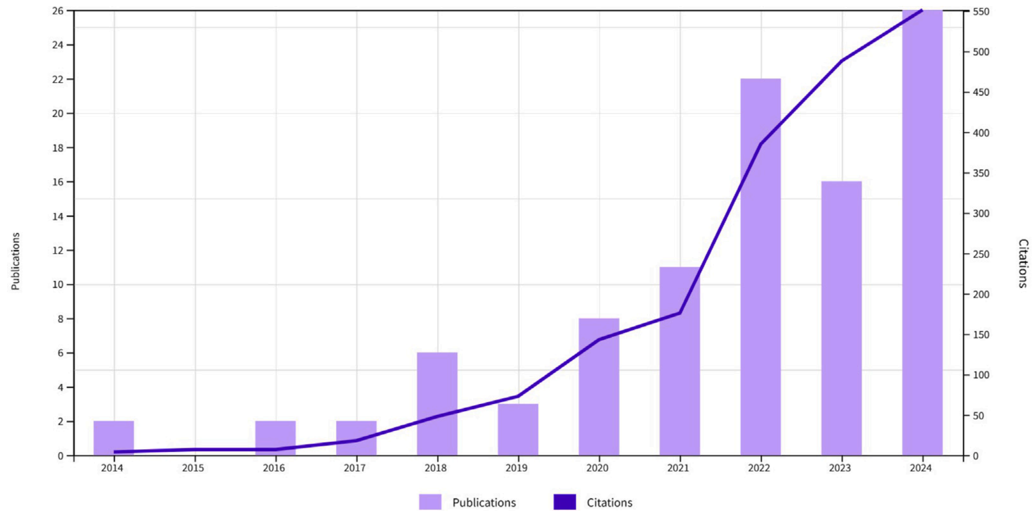

Fig. 3. Publications and citations of three-way multi-label classification.

theory [100–102] as:

$$\alpha_i^{jk} = \frac{\lambda^j_{P_i D_i^k} - \lambda^j_{B_i D_i^k}}{\left(\lambda^j_{P_i D_i^k} - \lambda^j_{B_i D_i^k}\right) + \left(\lambda^j_{B_i D_i^j} - \lambda^j_{P_i D_i^j}\right)},$$
$$\beta_i^{jk} = \frac{\lambda^j_{B_i D_i^k} - \lambda^j_{N_i D_i^k}}{\left(\lambda^j_{B_i D_i^k} - \lambda^j_{N_i D_i^k}\right) + \left(\lambda^j_{N_i D_i^j} - \lambda^j_{B_i D_i^j}\right)}. \tag{19}$$

3. Representative 3WD-based models for multi-label classification

The concept relations in multi-label is further relaxed from mutually exclusive to uncertain degrees of co-existence. This generalization frustrates the pipeline of single-label-based three-way decision with the characteristics of concatenating solutions of subproblems based on well-defined divide-and-conquer strategies. Fortunately, scholars have explored some promising directions in approximating the unseen label correlation by extending the representation and reasoning of uncertainty. Consequently, the three-way-based solutions are much more diversified. In this section, a summary of bibliometric analysis (see Section 3.1) is firstly conducted to present the landscape of the three-way-based multi-label classification. Detailed comparisons on the representative studies are trisected (i.e., *Extensions on DTRS* (see Section 3.2), *Extensions on S3WD* (see Section 3.3), and *Other Extensions* (see Section 3.4)) based on the prototype models.

3.1. Bibliometric analysis

To evaluate the trends of activity degrees for three-way multi-label classification, publications and citations on Web of Science are collected with the keywords "three-way decision*" or "TAO" or "trisecting-acting-outcome" or "trisecting-acting-optimizing" or "*rough set*" or "formal concept analysis" or "intuitionistic fuzzy set" or "evidential theory" or "shadowed set" and "multi-label" or "multilabel" or "label enhancement" or "label distribution learning" as topics from 2014 to 2024 (see Fig. 3). It can be observed that, most of the 98 papers are published in the past three years (i.e., 2022–2024), while the citations surge rapidly since 2021 and reach 550 in 2024. Both indicators show that the three-way multi-label classification is an emerging domain.

The distribution of publication titles in the 98 studies shows the journal preference. In Fig. 4, the numbers above the journal name record how many publications are published in the journal, and the areas on the chart is in proportional to the publication count. Among the 25 SCI-indexed journals, the top 10 journals that favor this topic are Information Sciences, Applied Soft Computing, Knowledge-based Systems, International Journal of Approximate Reasoning, International Journal of Machine Learning and Cybernetics, Applied Intelligence, IEEE Transactions on Fuzzy Systems, IEEE Access, Pattern Recognition, and Fuzzy Sets and Systems.

The corresponding topic distribution is shown in Fig. 5 . The presented 31 topics come from the co-occurrence of keywords in a publication, and each has co-occurred at least 3 times. The nodes represent the frequently discussed topics for three-way multi-label classification, while thickness of the topic-to-topic edges shows the degree of relevancy. Based on the color count, it can be observed that these topics are partitioned into four different categories. The keywords with the top 10 largest total link strength (i.e., the size of the dots in Fig. 5) are feature selection, rough sets, multi-label learning, fuzzy rough sets, uncertainty, feature extraction, redundancy, label enhancement, multi-label classification, and label correlation. A more in-depth observation on the co-authorship network is available in Fig. 6. All the selected 35 authors have published at least 3 papers and achieved at least 10 citations in this area. The nodes represent the active authors for three-way multi-label classification, while thickness of the author-to-author edges shows the strength of co-authorship. The collaboration is clustered into 8 groups, and the authors with the top 10 largest total link strength (i.e., the size of the dots in Fig. 6) are Tengyu Yin, Tianrui Li, Hongmei Chen, Keyu Liu, Lin Sun, Jiucheng Xu, Weiping Ding, Wenbin Qian, Duoqian Miao, and Zhong Yuan.

3.2. Extensions of DTRS

The labels in multi-label classification are sparsely distributed, thus the evaluation metrics prioritize the accuracy of the relevant labels. The motivation of extending DTRS for multi-label classification is to deal with the related cost-sensitive components including unreliable label association [117], doubtful label correlation [118], and ambiguous feature redundancy [119] in problem formulation. How the DTRS modules are extended will be discussed later.

(1) Model of Zhao et al. [117]

In the multi-granular threshold with a three-way-based label enhancement (MGT-LEML) model, the DTRS is employed to determine the instances with unreliable label association on an arbitrary label l_i for label enhancement. This is completed by estimating the label-based

Fig. 4. Distribution of publication titles on three-way multi-label classification.

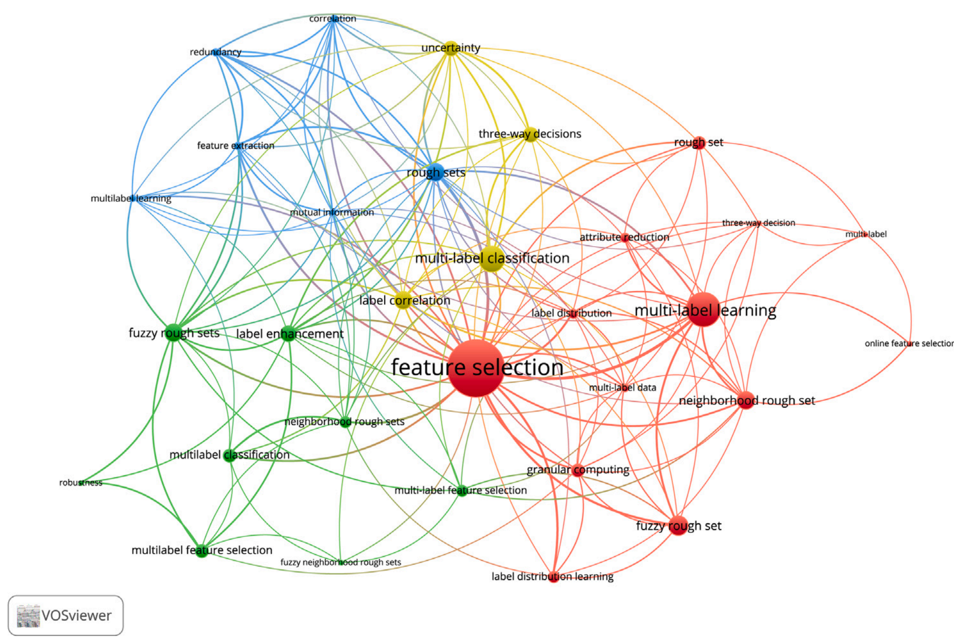

Fig. 5. Major topics in three-way multi-label classification.

misclassification cost.

$$
\begin{aligned}
Cost\left(P_{l_i}|[x]\right) &= \lambda_{PP}Pr\left(l_i|[x]\right) + \lambda_{PN}Pr\left(\neg l_i|[x]\right),\\
Cost\left(B_{l_i}|[x]\right) &= \lambda_{BP}Pr\left(l_i|[x]\right) + \lambda_{BN}Pr\left(\neg l_i|[x]\right),\\
Cost\left(N_{l_i}|[x]\right) &= \lambda_{NP}Pr\left(l_i|[x]\right) + \lambda_{NN}Pr\left(\neg l_i|[x]\right).
\end{aligned}
\tag{20}
$$

where $Pr\left(l_i|[x]\right)$ denotes the probability of an instance x that is with label l_i by scaling the output of a linear model into $[0,1]$. By following the Bayesian decision theory [100–102], the tripartition on the reliability of $\mathbf{x}$ on l_i is given as:

$$
\begin{aligned}
POS\left(l_i\right) &= \left\{x \in U \,\middle|\, Pr\left(l_i|[x]\right) \geqslant \alpha_i\right\},\\
BND\left(l_i\right) &= \left\{x \in U \,\middle|\, \beta_i < Pr\left(l_i|[x]\right) < \alpha_i\right\},\\
NEG\left(l_i\right) &= \left\{x \in U \,\middle|\, Pr\left(l_i|[x]\right) \leqslant \beta_i\right\}.
\end{aligned}
\tag{21}
$$

where the thresholds α_i and β_i are determined as:

$$
\begin{aligned}
\alpha_i &= \frac{\lambda_{PN} - \lambda_{BN}}{\left(\lambda_{PN} - \lambda_{BN}\right) + \left(\lambda_{BP} - \lambda_{PP}\right)},\\
\beta_i &= \frac{\lambda_{BN} - \lambda_{NN}}{\left(\lambda_{BN} - \lambda_{NN}\right) + \left(\lambda_{NP} - \lambda_{BP}\right)}.
\end{aligned}
\tag{22}
$$

(2) Model of Che et al. [118]

In the multi-label with three-way decision (ML-3WD) model, the DTRS is introduced to characterize the roughness of local label correlation. The corresponding misclassification cost is regarded as an indicator representing the fitness of local label relevance in problem formulation:

$$
\begin{aligned}
&\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{W}\mathbf{X} - \mathbf{Y}\|_F^2 + \frac{\sigma}{2}\|\mathbf{W}\mathbf{W}^\top - \hat{\mathcal{R}}|_{\mathbf{D}}\|_F^2 + \phi\|\mathbf{W}\|_1\\
&\quad s.t.\quad \|\mathbf{w}_j\|_2 = 1,\quad j = 1,\dots,s
\end{aligned}
\tag{23}
$$

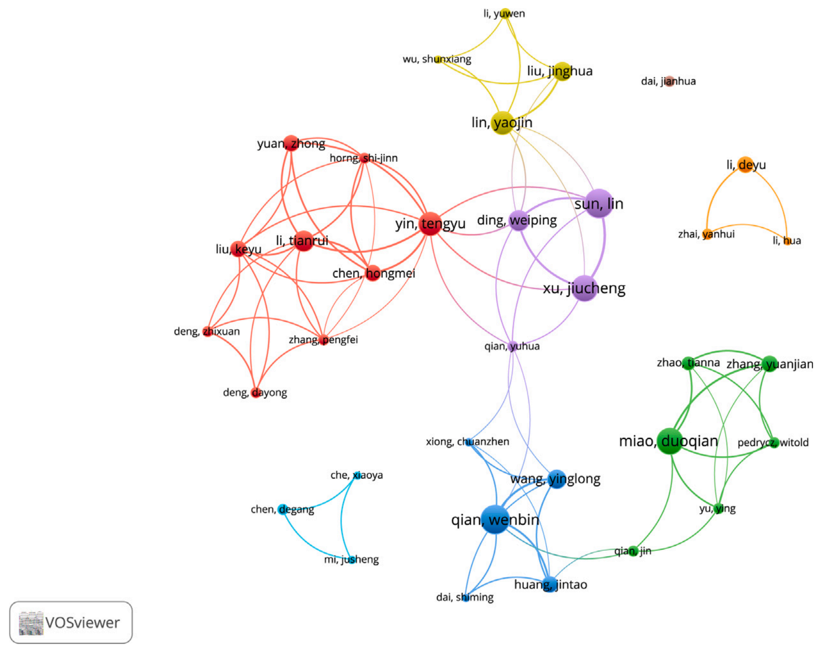

Fig. 6. Cooperation network with more than 3 papers and 10 citations in three-way multi-label classification.

where $\mathbf{W}$ is the coefficient matrix of the multi-label classification model, $\mathbf{X}$ is a training set with the corresponding label matrix as $\mathbf{Y}$. $\hat{\mathcal{R}}|_{\mathbf{D}}$ is a matrix with the dimensionality of $|L| \times |L|$ whose element is an aggregation of pairwise label correlation on dataset $\mathbf{D} \subseteq D$ (denoted as $r_{ij}(x_k)$). σ and ϕ are trade-off parameters, while s is the dimensionality of features. Note that the tripartition in this case is relevant to the positive, negative, and undetermined pairwise label correlation ($r_{ij}(x_k)$) instead of the affiliation of x_k to a specific label l_i.

$$\begin{aligned} &(LP) \quad Decide \quad r_{ij}(x_k) = +1 \quad if \quad Pr\left(D_{ij} \middle| [x_k]\right) \geqslant \alpha_{ij}(x_k), \\ &(LB) \quad Decide \quad r_{ij}(x_k) = 0 \quad if \quad \beta_{ij}(x_k) < Pr\left(D_{ij} \middle| [x_k]\right) < \alpha_{ij}(x_k), \\ &(LN) \quad Decide \quad r_{ij}(x_k) = -1 \quad if \quad Pr\left(D_{ij} \middle| [x_k]\right) \leqslant \beta_{ij}(x_k). \end{aligned} \tag{24}$$

where the thresholds $\alpha_{ij}(x_k)$, $\beta_{ij}(x_k)$ are the product of $\alpha_i(x_k)\,\alpha_j(x_k)$ and $\beta_i(x_k)\,\beta_j(x_k)$, respectively, which are determined as:

$$\begin{aligned} \alpha_j(x_k) &= \frac{\sum_{t=1}^{m}(1-\xi_{tj})\,\omega_{tj} d\left(x_{kt}^{+}, x_t^{+}\right)}{\sum_{t=1}^{m}(1-\xi_{tj})\,\omega_{tj} d\left(x_{kt}^{+}, x_t^{+}\right) + \sum_{t=1}^{m}\xi_{tj}\omega_{tj} d\left(x_{kt}^{-}, x_t^{-}\right)}, \\ \beta_j(x_k) &= \frac{\sum_{t=1}^{m}\xi_{tj}\omega_{tj} d\left(x_{kt}^{+}, x_t^{+}\right)}{\sum_{t=1}^{m}\xi_{tj}\omega_{tj} d\left(x_{kt}^{+}, x_t^{+}\right) + \sum_{t=1}^{m}(1-\xi_{tj})\,\omega_{tj} d\left(x_{kt}^{-}, x_t^{-}\right)}. \end{aligned} \tag{25}$$

where $0 \leqslant \xi_{tj} \leqslant 0.5$ is the risk avoidance parameter, ω_{tj} is the weight of tth feature on jth label, $d\left(x_{kt}^{+}, x_t^{+}\right)$ and $d\left(x_{kt}^{-}, x_t^{-}\right)$ represent the distance of tth feature of $\mathbf{x}$ to the maximal (x_t^{+}) and minimum value (x_t^{-}) of tth feature.

(3) Model of Yu et al. [119]

In the variable-degree multi-granulation decision-theoretic rough set multi-label feature selection (VMFS) model, the DTRS is incorporated to explore the distribution of redundant features. The bth feature is relatively redundant to the multi-label information system if the importance of feature subset ($IMP_{B-\{b\}}$) remains unchanged as compared with the origin (IMP_B) by observing the proportion between coarse and refine decision function at least v-percentage of the corresponding equivalence classes.

$$\begin{aligned} R(c) &= IMP_B - IMP_{B-\{b\}} \\ &= \frac{\sum_U card_B\left(RDF^{v}_{(\alpha,\beta),B}(x)\right) + \varphi(B)}{\sum_U card_B\left(RDC^{v}_{(\alpha,\beta),B}(x)\right) + \Phi(B)} \\ &\quad - \frac{\sum_U card_{B-\{b\}}\left(RDF^{v}_{(\alpha,\beta),B-\{b\}}(x)\right) + \varphi(B-\{b\})}{\sum_U card_{B-\{b\}}\left(RDC^{v}_{(\alpha,\beta),B-\{b\}}(x)\right) + \Phi(B-\{b\})} \end{aligned} \tag{26}$$

where $card(\cdot)$ is the function representing set cardinality, while $\varphi(\cdot)$ and $\Phi(\cdot)$ are two smoothing factors to make compensation of λ_1 and λ_2 if the numerator or denominator reaches 0 $(\lambda_1 \leqslant \lambda_2)$. $RDF^{v}_{(\alpha,\beta),B}(x)$ and $RDC^{v}_{(\alpha,\beta),B}(x)$ are the coarse and fine decision function on feature set B with the thresholds (α, β), $0 \leqslant v \leqslant 1$ denotes the variable degree. Finally, the tripartition on label approximation is given as:

$$\begin{aligned} POS^{v}_{(\alpha,\beta),B} &= \bigcap_{x \in X} RDF^{v}_{(\alpha,\beta),B}(x), \\ BND^{v}_{(\alpha,\beta),B} &= \bigcup_{x \in X} RDC^{v}_{(\alpha,\beta),B}(x) - \bigcap_{x \in X} RDF^{v}_{(\alpha,\beta),B}(x), \\ NEG^{v}_{(\alpha,\beta),B} &= U - BND^{v}_{(\alpha,\beta),B}. \end{aligned} \tag{27}$$

3.3. Extensions of S3WD

The S3WD-based multi-label classification solutions [120,121] emphasize the minimization of essential components in boundary regions. The boundary region is generated via information fusion on fine-grained features or labels.

(1) Model of Zhao et al. [120]

In the 3WDLE model, the S3WD is introduced to conduct label enhancement on unreliable classifications from a logical label-based model on unseen instances, and keep the classification unchanged otherwise. The unreliable classifications come from the boundary region whose instances are close to the label-specific hyperplane and surrounded by heterogeneous in neighborhood.

$$\mathbf{X}_2^{(\mathbf{d},\rho)} = \bigcup_{c=1}^{q}\left\{ x_j \middle| x_j \in \mathbf{X}_2^{c+} \cup \mathbf{X}_2^{c-} \bigwedge \sum_c sgn\left(\sum_c u_{jc} \leqslant \sum_c u_{ic}\right) \leqslant n_2^{le} \right\}, \forall i \neq j \tag{28}$$

where $\mathbf{X}_2^{c+}$ and $\mathbf{X}_2^{c-}$ denote the positive and negative instances close to the hyperplane of the cth label. u_{jc} is the fundamental instance descriptor which denotes the local uncertainty degree of x_j on cth label. $sgn(\cdot)$ is a sign function that returns 1 if the condition holds and 0 otherwise. n_2^{le} is a threshold which is related to the distribution of u_{jc} as:

$$n_2^{le} = \left\lfloor \frac{\sum_{g=1}^{m} \sum_{j=1}^{n_2} sgn\left(u_{jg}\right)}{\sum_{g=1}^{m} sgn\left(\sum_{j=1}^{n_2} u_{jg}\right)} \right\rfloor \tag{29}$$

where u_{jc} is the fraction of the weighted proportion of heterogeneous samples within the neighborhood of x_j (denoted as ρ_{jc}) to the distance of the sample to the hyperplane on cth label (denoted as d_{jc}) .

The three-way decision on label association is constructed in a label-dependent fashion, which is trisected by the relationship between classification and threshold τ_l for positive and negative region and the unreliability of classification for boundary region. The classification result is either determined by the logical-label-based model (i.e., $f_l(\cdot)$) or by the numerical-label-based model (i.e., $f_n(\cdot)$).

$$\begin{cases} POS_{l_i}(x_j) = & \left\{ sgn\left(f_l\left(x_j\right) \geqslant \tau_l\right) \middle| x \notin \mathbf{X}_2^{(\mathbf{d},\rho)} \right\} \\ BND_{l_i}(x_j) = & \left\{ sgn\left(f_n\left(x_j\right)\right) \middle| x \in \mathbf{X}_2^{(\mathbf{d},\rho)} \right\} \\ NEG_{l_i}(x_j) = & \left\{ sgn\left(f_l\left(x_j\right) < \tau_l\right) \middle| x \notin \mathbf{X}_2^{(\mathbf{d},\rho)} \right\} \end{cases} \tag{30}$$

(2) Model of Yu et al. [121]

In the three-way graph convolutional neural network (TW-GCN) model, S3WD is introduced to optimize the uncertain neighborhood-based relation within the layers of a graph convolutional network (GCN). Concretely, the node embedding in GCN is refined sequentially via the three-way neighborhood-based relation for the middle l layers. Specifically, the three-way-based graph convolution layers are iteratively defined as:

$$\mathcal{H} = Pr\left(\mathcal{H}^{l+1} \middle| \mathcal{H}^{l+1} = \sigma\left(N, \mathcal{H}^l, W^l\right), N \in \hat{\mathcal{A}}, \hat{\mathcal{D}}, \hat{\mathcal{U}}\right) \tag{31}$$

where $\sigma(\cdot)$ denotes the activation function like Relu and $W^l \in \mathbb{R}^{|U|\times|d|}$ denotes the weight matrix of the lth layer. $\hat{\mathcal{A}}$, $\hat{\mathcal{D}}$, and $\hat{\mathcal{U}}$ constitute the instance descriptor denoting the normalized term of the three attribute-level neighborhood-based three-way relation matrices.

The underlying three-way relation stems from advantages neighborhood, disadvantage neighborhood, and uncertain neighborhood of u_i under the attribute p_k. They are defined as follows:

$$\begin{aligned} POS_{p_k}\left(u_i\right) &= \left\{ u_j \in U \middle| \left(\left| f\left(u_i, p_k\right) - f\left(u_j, p_k\right) \right| > \gamma_{p_k} \right) \right. \\ &\qquad \left. \wedge \left(f\left(u_i, p_k\right) - f\left(u_j, p_k\right) > 0 \right) \right\}, \\ BND_{p_k}\left(u_i\right) &= \left\{ u_j \in U \middle| \left| f\left(u_i, p_k\right) - f\left(u_j, p_k\right) \right| < \gamma_{p_k} \right\}, \\ NEG_{p_k}\left(u_i\right) &= \left\{ u_j \in U \middle| \left(\left| f\left(u_i, p_k\right) - f\left(u_j, p_k\right) \right| > \gamma_{p_k} \right) \right. \\ &\qquad \left. \wedge \left(f\left(u_i, p_k\right) - f\left(u_j, p_k\right) < 0 \right) \right\}. \end{aligned} \tag{32}$$

where $f\left(u_i, p_k\right)$ denotes the value of u_i on attribute p_k, $|\cdot|$ represents the symbol of absolute value, the threshold γ_{p_k} denotes the average difference among two arbitrary instances u_i and u_j on attribute p_k. Finally, the two-way decision combines both the Euclidean and non-Euclidean features and is made at the $l+1$th layer as:

$$\begin{aligned} &\mathcal{H}^{l+1} = \sigma\left(\hat{E}_d \mathcal{H}^l W^l\right) \\ &where \quad \hat{E}_d = \varepsilon_{E_d}^{-1} E_d\left(u_i, u_j\right), \\ &\qquad E_d\left(u_i, u_j\right) = \begin{cases} 1 & R_d\left(u_i, u_j\right) \leqslant k, \\ 0 & otherwise. \end{cases} \end{aligned} \tag{33}$$

where $\mathcal{H}^l$ is the characteristic matrix at the lth layer, $\sigma(\cdot)$ is an activate function, W^l is the convolutional kernel at the lth layer represented by a weight matrix, ε is the degree matrix, E_d is an indicator function, which preserves the instances within the k neighborhood. $R_d(\cdot,\cdot)$ is a distance ranking function.

3.4. Other extensions

In addition to extensions of decision-theoretic rough sets and sequential three-way decision, three-way multi-label classification has seen a variety of model adaptation solutions. Table 2 to Table 5 summarize the prototypes, focused uncertainty, motivations, and contributions of the selected 70 papers. The functionality of the four dimensions are as follows:

1. Prototypes: This explains the foundation of the proposed model. By taking appropriate adaptations on these prototypes, the model's ability to handle uncertainty becomes stronger.
2. Focused Uncertainties: This explains the specific types of uncertainty addressed in the study. The classification accuracy will be significantly improved if a well-defined granular structure is developed to describe the crucial concepts.
3. Motivations: This explains the rationality on adaptation strategy. It highlights the limitations of the original prototype in handling multi-label classification and justifies the need for improvements.
4. Contributions: This outlines the significance of the proposed three-way multi-label solutions. It emphasizes how these solutions enhance the understanding and management of multi-label data.

Four groups are generated according to the differences of extension directions:

1. Extensions from rough set: This group (see Table 2) includes 20 papers that develop three-way via extension on rough set-based models, while the labels are complete.
2. Extensions from fuzzy rough set: This group (see Table 3) includes 20 papers that develop three-way via extension on fuzzy rough set-based models, while the labels are complete.
3. Extensions beyond rough set: This group (see Table 4) includes 14 papers that develop three-way via extension on prototypes other than rough set, while the labels are complete.
4. Extensions for flawed label: This group (see Table 5) includes 16 papers that develop three-way solutions for missing or noisy labels.

Meanwhile, to comprehensively analyze the progress on the label description-based three-way formulation, the studies that concerning the label distribution learning and label enhancement are marked by solid dot (i.e., •) and hollow circular dot (i.e., ∘) from Table 2 to Table 5. Studies without any kinds of dots represent the three-way-based formulations with logical-based multi-label classification.

From Table 2 to Table 5, the number of papers in each group (i.e., 20/20/14/16) implies that most three-way multi-label classification solutions are originated from rough set based theory, and the studies for flawed labels (i.e., missing labels or noisy labels) requires more investigations. There are 15 papers and 9 papers formulating three-way multi-label solutions for label distribution and label enhancement, with the group-based publications (9/3/0/3) for label distribution and (1/3/1/4) for label enhancement, respectively. Meanwhile, the relevant studies on the label distribution and label enhancement emerge from 2020 and 2022. As label importance information provides stronger supervision in dealing with multi-label classification, the uncertainty analysis on the two directions deserve more efforts.

Some interesting findings from the selected four dimensions are as follows:

1. It is easy to observe that most of these prototypes are soft computing techniques, which includes neighborhood rough sets,

Table 2
Representative three-way multi-label classification solutions (Group 1: Extensions from rough set).

Authors	Prototypes	Focused uncertainties	Motivations	Contributions
Yu et al. [122]	Neighborhood rough set	Bias between visual similarity and semantic similarity.	Projection from visual feature space to semantic concepts are uncertain.	Adopted an algorithm adaptation fashion on neighborhood rough set, and generated multi-output based on neighborhood relationship.
Li et al. [123]	Pawlak rough set	Feature selection	Specialized reduct for multi-label is preliminary.	Defined a complementary decision reduct on Pawlak rough set and constructed the classifier based on deduced features.
Zhang et al. [124]	Probabilistic rough set	Label-specific features Label correlation	Characterizing the boundary of ambiguious labels is difficult.	Constructed a label-specific three-way decision representation, where three regions are determined by attribute reductions from probabilistic rough set.
Liang et al. [125]	Multi-granulation rough set	Optimal granularity selection	Existing multi-granulation-based reduct cannot fully describe the uncertainty.	Suggested an optimal attribute reduction method by extending the optimistic and pessimistic operators in the sense of granules to refine positive reducts .
Fan et al. [126]	Neighborhood rough set	Feature selection Label ambiguity	The difference between single label and multi-label requires an extension for attribute reduction algorithms.	Deduced label positive region preserving, neighborhood label positive region preserving, label dependency and the neighborhood label dependency reduction by extending positive region-based attribute reductions.
Wen et al. [127]	Neighborhood rough set	Label ambiguity Heterogeneous sample pairs	ML-*k*NN degenerates significantly if neighborhood contains instances in boundary region.	Constructed a label-specific neighborhood relation to assist weight adjustments of features in *k*NN model based on the discernibility of labels.
Wu et al. [128]	Neighborhood rough set	Label-specific features Label correlation	Advancement in label correlation is promising but lacks analysis on uncertainty.	Developed a novel neighborhood relation with the embedded label correlation information measured by label similarity for feature selection.
Fan et al. [129]	Neighborhood rough set	Feature selection	Sample-level calculation for attribute reduction is time-consuming.	Constructed an inter-class based distance to generate neighborhood-based monotonic attribute reduction and increase the distance robustness via k pairs from instances with heterogeneous classes.
Wang et al. [130]	Pawlak rough set Membership function	Feature selection Granularity-aware reducts	Insufficient discrimination can be extracted from single-level label representation.	Explored the label representations on macroscopic-level, mesoscopic-level, and microscopic-level to reduce the computational complexity of feature reduction.
Liu et al. [131]	Pawlak rough set	Feature selection Knowledge consistency Knowledge independence	Uncertainty degrees towards the consistency and independence of knowledge granules are lacking, thus limiting the knowledge extracted from label space.	Defined the structures of knowledge consistency and knowledge independence from the perspective of granularity and features to generate the knowledge consistency-independence index for feature selection.
• Deng et al. [132]	Neighborhood rough set	Feature selection Label ambiguity	Studies on selecting features from high-dimensional data or data with redundant features with label distribution are lacking.	Developed a novel neighborhood fuzzy entropy as feature evaluation metric to extend the capability of feature selection with neighborhood rough set.
• Kou et al. [133]	Neighborhood rough set	Feature selection Label correlation	Mixed order of label correlations in label distribution learning has not been studied.	Presented a neighborhood description by feature value representation, and leverage label correlations by association mining with label distribution.
• Liu et al. [134]	Neighborhood rough set	Feature selection	Qualitative labels cannot discriminate the significance of two feature subsets in some cases and the local relationship between features and labels is neglected.	Extended the neighborhood relation for multi-label by collaborating with variable neighborhood structure,and improved the label-specific features by embedding label significance.
• Qian et al. [135]	Neighborhood rough set	Feature selection	Traditional mutual information cannot be directly employed for label distribution, which limits the effectiveness of label distribution learning.	Extended the mutual information to multi-labelfeature selection by incorporating the feature complementarity without discretization.
• Qian et al. [136]	Neighborhood rough set	Feature selection Feature dependency	High-dimensional label distribution learning suffer from inefficiency with numerical label representation.	Defined cost-sensitive approximations by neighborhood-based label distribution while maintaining the monotonicity variation of reducts.
• Qian et al. [137]	Neighborhood rough set	Feature selection Label ambiguity	Large concept priority has only been studied in single-label and the interpretability in multi-label case is lacking.	Transform the instance-level similarity to granular-ball based on adaptive similarity for enhanced label distribution and tripartition the usefulness of features.
∘ Qian et al. [138]	Linear discriminative analysis Granular-ball	Feature selection Label correlation	Linear discrimination-based feature selection does not support the label distribution, resulting in flawed knowledge extraction.	Proposed a two-stage framework to learn the label enhancement and corresponding reducts iteratively in terms of granular ball representation.
• Liu et al. [139]	Neighborhood rough set	Feature selection Neighborhood granulation	Neighborhood rough set deals with mixed type of data without breaking the data structures, and is promising for streaming features.	Extended the neighborhood relation for the feature selection of online streaming feature and maintain the classification accuracy.
• Chen et al. [140]	Neighborhood rough set	Feature selection Label correlation	Incremental instances require fast knowledge updating. Neighborhood rough set is an appropriate choice since it preserves the data structure without requiring prior knowledge.	Introduced mutual information into neighborhood construction for label correlation learning and extended to the dynamic updating scenario by sequentially handling analyses on feature importance, significance, and redundancy.
• Lu et al. [141]	Neighborhood rough set	Feature selection Instance hierarchy	Feature selection in label distribution should take the advantages of hierarchical relationships among data to improve feature discriminability, and granulation process can fill this gap to improve features' identification.	Presented a variable precision neighborhood rough set by fusing local and global metrics within a multi-granularity hierarchical framework to identify more discriminative and relevant features for label distribution learning.

Table 3
Representative three-way multi-label classification solutions (Group 2: Extensions from fuzzy rough set).

Authors	Prototypes	Focused uncertainties	Motivations	Contributions
Vluymans et al. [142]	Fuzzy rough set	Label ambiguity Neighborhood consensus	Approximation operators in fuzzy rough set extends the uncertainty reasoning of labelset-based aggregation.	Developed a consensus prediction with ordered weighted average aggregations based on neighborhood relation of labelsets via fuzzy rough operators-based quality measure.
Li et al. [143]	Fuzzy rough set	Feature selection Label roughness	Fuzzy rough set deals with feature space and label space separately, ignoring the connection between features pace and label space.	Extended the capability of learning internal correlation between feature space and label space in fuzzy rough set by merging multiple kernelized information from corresponding spaces and constructed a robust kernelized fuzzy rough set.
Tan et al. [144]	Fuzzy rough set	Feature selection Topic-based label correlation	Discernibility matrix has not been studied for multi-label case to explore sample and label correlations.	Introduced fuzzy rough discrimination matrix to evaluate and weight the sample discrimination ability of features for feature selection.
Qian et al. [145]	Fuzzy rough set	Feature selection Label correlation	Traditional fuzzy discernibility matrix cannot handle complicated discrimination relation in multi-label case.	Extended a fuzzy label discernibility relation and a fuzzy relative discernibility relation from fuzzy discernibility relation for feature selection.
Che et al. [146]	Fuzzy rough set	Topic-based label correlation	Consistency of crucial features imply the similar distribution on subset-based label correlation.	Explored sample-level positive and negative label correlation and simplified local attribute reductions via discernibility matrix on fuzzy positive region preserving.
Xu et al. [147]	Fuzzy rough set Neighborhood rough set	Feature selection Instance similarity	Feature importance are exploited from single perspective and ignore uncertainty in boundary region.	Constructed a mixed measure combining fuzzy neighborhood conditional entropy and approximate accuracy of fuzzy neighborhood for feature selection.
Yao et al. [148]	Fuzzy rough set	Feature selection Instance similarity	Existing rough set-based methods cannot effectively characterize the ability of features to distinguish multilabel sample pairs and are time-consuming in processing large-scale multi-label data.	Improved the discerning ability of feature sets from the perspective of discernibility relations by defining relative discernibility pair matrix in the fuzzy rough set and accelerated the computation with the combination of sampling and ensemble strategies.
Che et al. [149]	Fuzzy rough set	Topic-based label correlation	The local label correlation is easily trapped by over-fitting issue, while the decisive influence of features generated by local label attribution should be an alternative solution for local label correlation learning.	Provided a feasible solution to estimate the distribution of local label correlation at instance level based on kernelized fuzzy rough set and improved the applicability on unseen instances by redefining label relevance matrix as the integral average of the distribution.
Zhong et al. [150]	Fuzzy rough set	Feature selection Feature redundancy	Redundancy on the impact of selected feature set is limited, resulting in the feature redundancy and incomplete collections of discriminative features.	Partitioned the feature redundancy into label-independent redundancy and label-dependent redundancy and employed fuzzy conditional mutual information based on the results of redundancy analysis between candidate and selected features.
Yin et al. [151]	Fuzzy rough set	Feature selection Feature interactivity	Importance of feature interaction is neglected, which incurs the contradict estimation on class relevancy degree when they are with and without other features.	Explored the multi-neighborhood granularity representation of multi-label and the interactivity between features by measuring fuzzy dependence from both algebraic view and information view.
Jiang et al. [152]	Fuzzy rough set Neighborhood rough set	Feature selection Label correlation	Divergence-based fuzzy rough set is noisy-sensitive and cannot leverage the relevance among all labels, especially for the uncertainty induced by upper approximation.	Enriched the self-information representation of upper approximation and lower approximation to describe feature distinguishing ability by exploring divergence-based fuzzy neighborhood relation on three-level uncertainty measure.
Dai et al. [153]	Fuzzy rough set	Feature selection Label correlation	The informativeness of features w.r.t. labels may be different, and the ignorance degenerates the effectiveness of selected features.	Introduced strongly relevant label gain and label mutual aid information into feature significance by combining fuzzy mutual information and fuzzy conditional mutual information, which improves the estimation of positive correlation between features.
Dai et al. [154]	Fuzzy rough set	Feature selection Label correlation	Although fuzzy mutual information is effective in characterizing uncertainty, it has high computational cost for multi-label with high-dimensionality.	Established an optimization framework based on fuzzy mutual information to deduce the feature weights, which supervised the learning of embedded features from both correlation perspective and regression perspective.
Liu et al. [155]	Fuzzy rough set Synthetic minority oversampling	Class imbalance	Complexity of generating synthetic data for minority classes in multi-label arise significantly.	Exploring the distribution of labelset-aware synthetic data by combining fuzzy rough set and optimized intra-cluster-based instances.
• Qian et al. [156]	Fuzzy rough set Neighborhood rough set	Feature selection Label ambiguity	Reducing the feature dimensionality of label distribution learning is challenging due to the label ambiguity.	Accelerate the feature reduction on label distribution learning by integrating granular ball with the fuzzy rough set.
∘ Cai et al. [157]	Fuzzy rough set	Feature selection Instance similarity	Single distance metric overlooks the variability of label descriptions for different datasets.	Suggested a multi-metric function to learn fuzzy similarity relationship via metric learning and deduce feature selection based on the extended approximation operators.
∘ Deng et al. [158]	Fuzzy rough set	Feature selection Label ambiguity	The effectiveness of selected features are limited by the flawed selecting criteria, which does not consider the structural information including relevancy between labels and redundancy among features.	Proposed a criterion for feature selection via label enhancement called weighted label relevancy and label-fuzzy redundancy, which acquires distinguished knowledge in positive regions from a general label-fuzzy rough set.
∘ Sun et al. [159]	Fuzzy rough set Neighborhood rough set	Feature selection Label dependency	Existing approaches does not fully leverage the correlations between features and labels, while the variants of ant colony optimization is promising in speedup and feature stability.	Optimized the fusion strategy of roughness degree by integrating algebraic and information viewpoints from a new adaptive fuzzy neighborhood rough set, which further improved the initial representation in ant colony optimization.
• Li et al. [160]	Fuzzy rough set	Label-specific features Label correlation	Abundant information in irrelevant labels are omitted in limited studies concerning label distribution, resulting in the biased estimation of label-specific features.	Constructed multi-label-specific features by simultaneously leveraging label importance from instance similarity and label correlation from fuzzy rough set.
• Dai et al. [161]	Fuzzy rough set Discernibility pair theory	Feature selection Instance similarity	Existing discernibility pair theory is not applicable to label distribution case, and even for the feature selection with label enhancement, the relevance and importance between labels are considered rarely.	Customized the discernibility pair theory for speedup of feature selection and explored the approximated representation of soft relevance between objects and labels via fuzzy rough set.

Table 4
Representative three-way multi-label classification solutions (Group 3: Extensions beyond rough set).

Authors	Prototypes	Focused uncertainties	Motivations	Contributions
Denoeux et al. [162]	Dempster-Shafer theory	Label ambiguity Label lattice	The partial label association in multi-label is set-valued, thus compatible with DS theory.	Formulated the multi-label classification problem as a synthesis of the mass function represented by tuples under the context of the nearest neighborhood.
Masson et al. [163]	Dempster-Shafer theory	Label ambiguity Label ranking	Quantitative solutions fail to support analysis on partial order cases like label ranking.	Suggested a partial order structure tailored for belief function to deal with multi-label classification.
Mehravaran et al. [164]	Dempster-Shafer theory Rough set theory	Feature selection Label correlation	Existing feature selection strategy fails to evaluate the importance of label correlation.	Presented a feature selection method for correlation label based on overlapping nature of label-specific features by regarding belief function as heuristic information of attribute reduction.
Chen et al. [165]	Network embedding	Label ambiguity	Existing network embedding only learns the local structures of nodes under a single granularity, while the beneficial global structure is neglected.	Developed a multi-granular network embedding to improve the informativeness of embedded network by repeatedly employing quotient space from coarsest to finest, where the granulation criteria are flexible and the global structure within the same candidate node is preserved.
Nápoles et al. [166]	Recurrent neural network	Label ambiguity Feature sparsity	Sparse features degenerates the classification performance, and imputation incur bias for distribution.	Developed a three-block structure to sequentially generate high-level features, learn relationship, and conduct aggregation by incorporating with the cognition processing.
Cui et al. [167]	Deep hashing	Semantic correlation	Deep hashing cannot exploit the multi-level similarities between instances, especially for the semantic correlation at fine granularity.	Developed a multi-central ranking loss to preserve complex semantic correlations of multi-label images with low quantization error, where the hash centers count is learnable and semantic similarity is measured in a metric space.
Wu et al. [168]	Formal concept analysis	Label correlation	Existing concept-cognitive learning fails to establish the robust association between structured features and multi-label.	Constructed a correlation concept-cognitive learning method for the optimization of feature concepts and formulation of positive and negative label correlation concepts.
Li and Xu [169]	Variational estimation	Label correlation	Knowledge distillation is degenerated by the inherent uncertainty in the knowledge transfer process.	Constructed a multi-model combination strategy for category enhancement with three loss functions to distill knowledge from intra-class, interclass, and prediction probability deviation.
Yu et al. [170]	Graph convolutional neural networks	Label ambiguity Instance similarity	The inference capability of graph convolutional neural networks has not been exploited in analyzing instance relations, while the equally importance of attributes incurs suboptimal solution.	Optimizing the instance similarity by incorporating attention mechanism with multiple difference matrices, where the adaptive adjustment is realized by introducing attention mechanism on feature subset with varying granularity.
Peng et al. [171]	Bidirectional encoder Representation from Transformation	Label ambiguity Long-tail labels	A single vector for the whole text cannot capture intact discriminative information, while label-specific vectors tends to over-emphasizing the importance of tail labels.	Constructed a label-adaptive representation framework to enhance the improvement of features from both layer-aspect and fragment-aspect, which is composed of three components named as base feature generation, representation pool building, and adaptive label matching, respectively.
Zou et al. [172]	Gradient-based methods Information theory	Label-specific feature Label correlation	The importance of different features on label correlation should be different to mitigate the gap between feature correlation and label correlation.	Developed a second-order regularization term to consider three-way feature interaction between actual decrease and expected reduction on both feature and label side.
Wei et al. [173]	Recurrent neural network Graph neural network	Semantic correlation Instance similarity	The imbalanced semantic similarity within labels can degenerate the discrimination, while the generalization capability of tail-label is limited.	Incorporated multi-level constraint augmentation with label association attention to alleviate the impact of imbalanced instances and optimize the association between text and levels.
Liu et al. [174]	*k*-nearest neighborhood	Label ambiguity Instance similarity Label correlation	Oversampling imbalanced data can be skewed if a single neighborhood is selected, and degenerate the label correlation.	Proposed an adaptive strategy for natural neighbor construction to approximate the underlying local label distribution and the structure of label correlation.
∘ Qian et al. [175]	Linear regression	Feature selection Feature ambiguity Label ambiguity	Conventional multi-label learning cannot effectively handle the vagueness incurred by the difference among all associated semantics in the label space.	Extending the uncertainty capability of label enhancement by developing fuzzy neighborhood discrimination index and fuzzy label similarity.

Table 5
Representative three-way multi-label classification solutions (Group 4: Extensions for flawed label).

Authors	Prototypes	Focused Uncertainties	Motivations	Contributions
Sun et al. [176]	Fuzzy rough set Neighborhood rough set	Feature selection Missing label Label correlation	Shortage of uncertainty measures for multi-label limits the quality of selected features, especially for the missing labels case.	Proposed fuzzy neighborhood entropy-based measures to approximate the semantics and relationships among ambiguous labels with the criterion maximum relevance minimum redundancy.
Liang et al. [177]	Neighborhood rough set	Feature selection Missing label Instance similarity	Hyperparameter setting in existing stream feature selection requires domain knowledge, and these configurations cannot be easily adapted with missing label case.	Proposed a filtering strategy based on neighborhood rough set for streaming feature selection without parameter configurations.
Shu et al. [178]	Neighborhood rough set	Feature selection Missing label Label correlation Feature redundancy	Hyperparameter neighborhood radius cannot reflect data characteristics, while granular ball determines radius in a data-driven way.	Developed a granular-ball-based mutual information measure as feature significance on recovered labels, which is learned by simultaneously optimizing the feature matrix and label matrix.
Gull and Aguilar [179]	Multi-assignment clustering Fuzzy theory	Label ambiguity Missing label Instance similarity	Shortage of multiple cluster assignment cannot support pattern recognition for one-versus-many label associations.	Defined a semi-supervised algorithm to automatically generate overlapping clusters by optimizing three membership thresholds.
Liu et al. [180]	k-modes clustering Formal context	Label dependency Missing label	Single-level-based dependency exploration overlook the multi-level characteristic of label dependencies.	Develop a meso-granularity description on complete labels by granularity-aware mutual information and complete the missing label based on stochastic granules.
Qian et al. [181]	Neighborhood rough set Bayes' rule	Feature selection Noisy label Label confidence	Most partial label learning methods cannot learn the underlying factor that incurs the noisy label, thus the model robustness on unseen instances cannot be guaranteed.	Learning the key distinguishing features among labels based on a two-stage framework, which selects the most beneficial features by extending granular ball firstly, and adaptively adjust the feature weights based on joint effects from global to local.
Li et al. [182]	Multilabel twin support vector machine	Label-specific features Noisy label Topic-based label correlation	The hyperplane of twin support vector machine is sensitive to noisy labels. Meanwhile, it cannot capture the specificity of label and the one-by-one learning fails to learn label correlation.	Demonstrated the effectiveness of intuitionistic fuzzy set in filtering noisy labels and improved the feature representation in a unified framework by cascading the clustering with the group-based label manifold regularization.
Yin et al. [183]	Fuzzy rough set Covering rough set	Feature selection Feature interactivity	How to incorporate the characteristics of multi-label to define fuzzy β covering neighborhood relation and the corresponding uncertainty measure remains unsolved.	Extended the robustness of fuzzy β covering by considering multi-neighborhood on multi-label fuzzy β covering decision by fusing the relevance, redundancy, interactivity, and complementarity information between features and the fuzzy β covering decision.
Sun et al. [184]	Neighborhood rough set	Feature selection Noisy label Instance similarity	Traditional granular ball models cannot cope with the partial multi-label learning and incur secondary errors with anomalous splitting on granular structure.	Enhanced the robustness of granular ball against noisy label by combining with fuzzy mutual information and optimizing a fuzzy membership degree-based objective function. A granular ball partial multi-label system is thus established.
• Qian et al. [185]	Local rough set	Feature selection Missing label Label correlation	Incomplete label distribution cannot be effectively handled, and local rough set shows superiority on learning incomplete labels.	Extended the capability of local rough set on multi-label feature selection based on customized neighborhood relation and cost-sensitive approximation operators.
• Qian et al. [186]	Neighborhood rough set	Feature selection Missing label Instance similarity	Incomplete labels and dimensionality disaster appear simultaneously, imposing difficulty for label distribution learning.	Developed a new significance measure called neighborhood-tolerance discrimination index by leveraging neighborhood-tolerance relation on neighborhood discrimination index.
∘ Yin et al. [187]	Fuzzy rough set	Feature selection Noisy label Label ambiguity	While the label distribution implies crisp assumption on relation between instances and relevant labels may not hold, the relative solutions are limited and the setting of fuzzy neighborhood radius is empirical.	Optimized the robustness of fuzzy neighborhood granules by introducing representative samples identification as a subproblem and extended the β-precision fuzzy rough set in describing semantics of noisy multi-label.
∘ Yin et al. [188]	Fuzzy rough set	Feature selection Label ambiguity	Shortage of solutions for noisy feature-based feature selection in label enhancement and insufficient consideration of evaluation perspectives on uncertainty measure.	Improved the robustness of feature selection criterion for label enhancement by extending ∘ β-precision fuzzy dependency and fuzzy entropy measures for uncertainty quantification among natural neighbors-based instance evaluation coefficient from both the algebraic and information perspectives.
∘ Deng et al. [189]	Fuzzy rough set	Label ambiguity Missing label Topic-based label correlation	Low-rank assumption is prone to overfitting, while non-negative matrix factorization learns from parts-based representation with strong interpretability.	Introduced non-negative matrix factorization into fuzzy rough set to recover missing label and provides interpretability for the relationships between latent features and labels.
• Li et al. [190]	Linear regression	Label ambiguity Noisy label Instance similarity	Noisy labels in label distribution learning degenerate the classifier and are difficult to locate, which are feasible by three-way decision.	Developed a sequential three-way noisy instance detector before the label distribution learning, and impose reconstruction constraints on the feature weights for noisy instances.
∘ Sun et al. [191]	Fuzzy rough set Neighborhood rough set	Feature selection Noisy label Instance similarity	Hyperparameter settings in fuzzy neighborhood rough set incurs information loss in granulation, and noisy labels degenerate the label enhancement.	Constructed a label enhancement-based mutual information by combining approximation accuracy from algebraic perspective and fuzzy neighborhood entropy from information perspective.

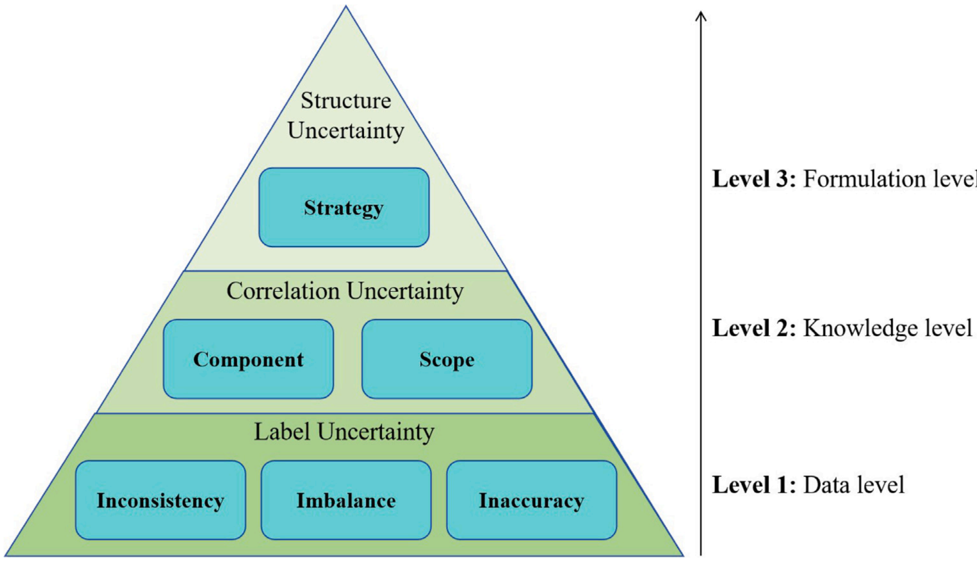

Fig. 7. Three-level uncertainty in multi-label classification.

fuzzy rough sets, Pawlak rough set, multi-granulation rough sets, Dempster-Shafer theory, and formal concept analysis. A common characteristic is that they share a pairwise approximation structure to the targeted concept.

2. It is shown that fuzzy rough sets (26/70) and neighborhood rough sets (25/70) are the most preferred methods for model adaptations in developing three-way-based multi-label classifiers.
3. The feature selection (43/70), label correlation (22/70), and label ambiguity (20/70) as the top three focused uncertainties. The emphasis on feature selection and label ambiguity stems from the straightforward extension of rough set, while the label correlation comes from the characteristics of multi-label.
4. Although topic-based label correlation seems more rationale in human cognition, the relevant studies are limited (5/22), which means the uncertainty analysis on topic generation should be highlighted.
5. Instance similarity (13/70) should not be restricted in feature space only, and the feature interactivity should be further considered in developing feature selection solutions.

4. Structural uncertainty analysis for three-way-based multi-label classification

The generalized relationship between different concepts incurs multifacted uncertainty in employing three-way decision on multi-label classification. To systematically minimize uncertainty, the essential uncertainty hierarchically is intensively discussed firstly. As the data-driven nature of uncertainty is beyond the assumption of traditional three-way-based model, a three-way-based solution called Multi-Label Sequential Decision-theoretic Three-Way Decisions (ML-SD3WD) is established by combining decision-theoretic rough set with sequential three-way decision accordingly. Furthermore, some potential research directions on implementing the ML-SD3WD model are discussed.

4.1. Uncertainty factors

Taking a bottom-up perspective, the uncertainty factors are summarized in three levels, namely label uncertainty (data level), correlation uncertainty (knowledge level), and structure uncertainty (formulation level) (see Fig. 7).

4.1.1. Level 1(Data level): Label uncertainty

Label uncertainty is the foundation in three levels and concerns the uncertainty incurred by the distribution of observable label association. Concretely, it involves three kinds of uncertainty, i.e., inconsistency, imbalance, and inaccuracy.

The inconsistency of labels refers to the case where the similar features of two instances do not guarantee the similarity on some labels. The inconsistency degrees are widespread in well-known benchmarks like Mulan,[3] where the degree of multi-classification of quality is measure-dependent and fluctuates in a wide range. Although inconsistency also applies to single-label-based classification, the inconsistency degree in the multi-label case is a distribution rather than a fixed value. Such a phenomenon has a far-reaching influence on data complexity [192], discretization [193], and label propagation [194], where some additional procedures are required for the reduction of information loss.

The imbalance of labels represents the imbalanced distribution of positive over negative classes from perspectives of both instance and label [195]. Given that the dimensionality of label space represents the interests of humans, the count of relevant labels for a particular instance is limited when compared with the label count. Furthermore, these relevant labels are less likely to share label association on a given instance. Consequently, the evaluation metrics reward the performance that achieves better discrimination on positive class, and many techniques (e.g., cost-sensitive penalty function [196], subproblem aggregation [197], variational distribution learning [198]) are exclusively developed to address this issue. Since the imbalance closely relates to sparsity, it is a critical indicator of data characteristics. It corresponds to *cardinality* for instance-based imbalance and *density* for instance-based and label-based imbalance.

The inaccuracy of labels refers to the untrustworthy annotations. Compared to the constant influx of data, maintaining accurate annotations manually at a large scale is costly. Meanwhile, factors including the saliency and similarity of labels can also degrade the annotation precision if volunteers are recruited. The classifier is thus degenerated if it is sensitive to the label variations. To alleviate the performance

[3] https://mulan.sourceforge.net/datasets-mlc.html

degeneration, some techniques (e.g., low-rank-based sparse reconstruction [199], pre-training-based pseudo-label generation [200], low-rank-based and full-rank-based matrix factorization [201]) on learning underlying statistical label distribution are employed.

4.1.2. Level 2(Knowledge level): Correlation uncertainty

The correlation uncertainty refers to the unknown label correlation and can deteriorate if the label uncertainty is processed inappropriately. By referring to knowledge, they can be learned from label association. One main reason why divide-and-conquer strategies fail in multi-label cases is that the latent label correlations constitute a covering rather than a partition on label space. In most cases, practitioners have virtually no knowledge available on label correlation due to the scarcity of supervision. Fortunately, the approximation on label correlation seems feasible by combining the component and scope.

The component of label correlation concerns what kinds of l_j are correlated with the given l_i, where $1 \leqslant i, j \leqslant q$. Although the order of label correlation may vary, the constraint on the feature side serves as an additional module, which regularizes the similarity of coefficients. Recently, the adaptive discovery of label correlation reports promising results by introducing techniques like variational Ref. [202], spectral graph [203], and graph learning [204].

The scope of label correlation concerns what kinds of conditions are required to support the label correlation between l_i and other label(s). The learning procedure is completed by following instance-based similarity measures in either observable or latent features. Compared with the global label correlation assumption, the topic-based local label correlation [205–207] assumption is a more generalized case. Practically, the local label correlation is mainly generated via partition-based clustering, where the cluster count is specified empirically.

4.1.3. Level 3(Formulation level): Structure uncertainty

The structure refers to the organization of uncertainty-aware modules in the objective function. To maintain the generalization and discrimination, a group of solutions formulates the multi-label classification into a single loss function, where the globally or locally label correlation constraint and the latent representation learning on both features and labels are explicitly introduced as regularization terms. However, the performance improvement is mainly from the two-way-based parameter optimization with unified knowledge granularity. The absence of an adaptive granulation mechanism implies the insufficiency in deducing an uncertainty-aware model.

4.2. ML-SD3WD model

Based on the investigations of uncertainty factors in the last section, it is worth mentioning how three-way decision gradually reduce the uncertainty of multi-label classification and the corresponding challenges here.

4.2.1. Pipeline

The supervision strength is a crucial point in problem formulation. We assume that the labels are represented in the logical style, and no hierarchical structure on labels exists. This means the classification concerns whether an instance is associated with a particular label. A three-way-based projection from features to labels is developed while reducing the multifaceted uncertainty. Following the perception of humans, the projection should be endowed with the following functionalities:

- It should be capable of capturing ambiguous topics to realize comprehension capability.
- It should be capable of making uncertain-aware decisions. In other words, it should proceed rapidly for salient labels and delay the decision otherwise.
- It should be capable of learning distribution-based relative label importance from logical labels based on well-defined uncertainty measures.

The three requirements are completed by formulating a general framework called multi-label sequential decision-theoretic three-way decision (ML-SD3WD). The main motivation is determining the instance-label pairs by highlighting interactions on uncertain topics. As shown in Fig. 8, the decision-theoretic rough set with the sequential three-way decision is integrated to resolve the subproblems of topic generation, label assignment, and label enhancement hierarchically. The prefix three-way on the three essential procedures emphasizes the operations of uncertainty. The corresponding results of "Three-way Topic Generation", "Three-way Label Assignment", and "Three-way Label Enhancement" are "Correlation Confidence Distribution", "Label Confidence Distribution", and "Enhanced Label Confidence Distribution", which gradually minimizes the size and degree of instance-label uncertainty in the label matrix for unseen instances. The label confidence distribution is both topic-dependent and label-dependent. The instances with uncertainty labels in current procedures will be handled in forthcoming steps by leveraging sophisticated label correlation. For comprehension, we illustrate the accumulation of trustworthy distribution on predicted labels from light gray to dark gray. Since the ML-SD3WD model emphasizes the learning on uncertainty instance-label pairs and preserves those instance-label pairs with trustworthy classification, one can infer that the trustworthy degree is monotonically increasing as the sequential three-way-based procedures continue.

The TAO framework displays different structures in developing modules "Three-way Topic Generation", "Three-way Label Assignment", and "Three-way Label Enhancement" (see dashed circles in Fig. 8). The rationality on such differences is given below:

- For the subtask topic generation, the triading is required for the generation of uncertain topics. Since this procedure concerns the characterization of topic roughness only, the acting and optimizing operations are unnecessary. A representative triading procedure occurs between two topics, where the uncertain denotes the boundary region.
- For the subtask label assignment, the acting is required to assign the label relevancy to instances. Here, neither triading nor optimizing is required as the processing of uncertain topics is highlighted. A representative acting procedure occurs between two topics for label assignment, where ensuring the label relevancy directly for confident labels and fuse labels otherwise.
- For the subtask label enhancement, triading and acting are required to confirm the label relevancy, where triading is introduced to identify the candidate instances for label enhancement and the acting on trustworthy and untrustworthy instance-label pairs are different. Meanwhile, it is relevant to the optimizing as it is the last procedure in ML-SD3WD. The enhanced labels correspond to the instances with high misclassification risk.

Formally speaking, the ML-SD3WD framework is denoted as a triple $\left(\mathbf{G}, \mathbf{W}, \left(\lambda_j^{*\cdot}(i), \boldsymbol{\gamma}_i\right)\right)$, where $\mathbf{G}$ denotes the topic-based representation, $\mathbf{W}$ denotes the topic-aware projection from features to labels in appropriate space, $\lambda_j^{*\cdot}(i)$ denotes the misclassification cost w.r.t. ith topic in the jth procedure ($j = 0, 1, 2$), $\boldsymbol{\gamma}_i$ denotes the threshold of determining enhanced labels for the ith topic. The S3WD and DTRS are executed iteratively by the tuple $\left(Des_j(x), \lambda_j^{*\cdot}(i)\right)$, where j decreases from 2 to 0.

4.2.2. Challenges

In what follows, some potential research directions on implementing the ML-SD3WD framework are discussed.

1. **Fine-grained three-way representation:** The three-way on single-label classification case only concerns the components

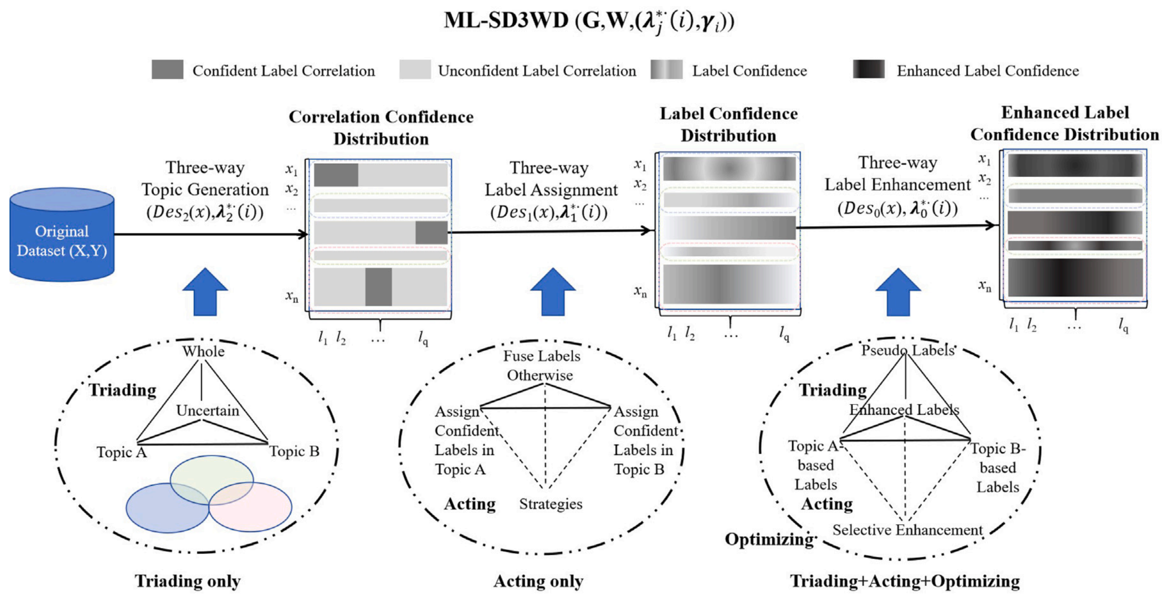

Fig. 8. Pipeline of ML-SD3WD Model.

of instance subsets. This stems from the assumption that concepts are at least mutually exclusive. However, inconsistency is a widespread issue in multi-label classification, which implies meaningless to directly discuss the triading across labels ($POS_B^{(\alpha,\beta)}(X)$, $BND_B^{(\alpha,\beta)}(X)$, $NEG_B^{(\alpha,\beta)}(X)$). Therefore, how to represent fine-grained three-way regions under the context of multi-label classification ($POS_{B_i}^{(\alpha_i,\beta_i)}(X_i)$, $BND_{B_i}^{(\alpha_i,\beta_i)}(X_i)$, $NEG_{B_i}^{(\alpha_i,\beta_i)}(X_i)$) remains an open issue.

2. **Feature augmentation:** The original features emphasize how an instance is generated rather than providing discriminative information on determining label association. A gap between representation and classification exists. Meanwhile, the underlying correlation is crucial in measuring instance-based similarity. An alternative solution is to learn some label-dependent features from the original features and concatenate them for topic generation. However, these features should be compact in size to avoid over-fitting. How to develop an effective strategy for learning augmented features remains an open issue.
3. **Robust label-specific features:** The label-specific features concern both the characteristics and correlations of the label. However, as the ever-increasing interests arise, the dimensionality of labels becomes higher, which incurs the phenomenon of long-tail labels. In other words, the prior probability for a particular label or combination may be negligible, while the misclassification on such labels can be intolerable (e.g., diagnosis for rare diseases). Therefore, the robustness of label-specific features can be risky. A feasible solution is to develop some novel uncertainty measures with stronger discrimination on candidate features. Although Zentropy [208,209], with the nature of multi-granularity, shows the potential in single-label classification, the effectiveness requires further demonstration.
4. **Topic generation:** Existing topic generation strategies mainly take a two-way-based clustering and neglect the possibility of overlapping instances. From the perspective of clusters, the partition-based operation may be suboptimal and the related bias can be propagated. Since similar topics share similar labels, such oversimplification in multi-label cases may incur more information loss. Therefore, a three-way-based topic generation strategy is preferred. Three-way clustering has shown its superiority in scenarios like community detection and recommendation. However, since the relationship among labels is uncertain, some adaptations to three-way clustering should be considered. How to incorporate inconsistency in both instances and labels into the similarity metric remains an open issue.
5. **Topic-based label correlation:** The locality of label correlation is more likely than the global label correlation if the influence of topics should be considered. The exploitation of topic-based label correlation can regularize the sparsity of label-specific features. However, the topic-based label correlations are still uncertain even if the topics are available. Two challenges are raised: (1) How to adaptively learn the components of topic-based label correlation without explicit supervision information, and (2) How to deduce a robust topic-based label correlation with limited instances. The first issue concerns the simultaneous optimizations on the label correlation order and the correlated labels. The second issue concerns the exploitation of the roughness of topics.
6. **Adaptive loss function:** A significant trend in extending DTRS from single-label to multi-label is the increasing number of loss functions. In the ML-SD3WD model, the cost-sensitive may be considered for topic generation, label assignment, and label enhancement, which implies more loss functions. A worse matter is the exponential combinations of parameters from different misclassification costs, which brings in many issues like sophisticated optimization and over-fitting. In this case, how to define the quantitative relationship with the least hyperparameters remains an open issue.
7. **Cross-topic label aggregation:** For non-salient instance-label pairs, the label assignment cannot be directly determined by a single topic. Instead, they are associated with multiple topics, with the relative importance of each topic determined by the similarity between the instances and the relevant topics. Based on these multi-perspective weights and label correlation information, label aggregation strategies can be designed, leveraging collective decision-making and evidence fusion techniques. However, several challenges deserve discussion, including the

metric-learning on the similarity between instances and topics, the conflict analysis from different perspectives, and the extension of evidential theory and group decision for label aggregation.

8. **Selective label enhancement:** Label enhancement is a technique that learns the label distribution from logical labels by exploiting similarities locally. By referring to selective label enhancement, the label enhancement concerns instances with untrustworthy label predictions only. This raises two questions: (1) How to define the untrustworthy label predictions, and (2) How to exploit the trustworthy label predictions without retraining. The first issue concerns the uncertainty measure, and one possible solution is the ensemble of a multi-granulation rough set and a decision-theoretic rough set. The second issue concerns knowledge representation, and one feasible solution is knowledge distillation.
9. **The trade-off between efficiency and effectiveness:** The ML-SD3WD model can sequentially improve the instance-label pairs with uncertainty. In contrast, the efficiency of ML-SD3WD stems from the distribution of trustworthy instance-label pairs. The two goals are contradictive and may encounter challenges in large-scale multi-label datasets (i.e., the number of instances exceeds 5000, and there are more than 100 labels). How to maintain the satisfying scalability of ML-SD3WD remains challenging.

5. Applications of 3WD multi-label classification

Three-way multi-label classification shows superiority over problems with limited precision and/or models with insufficient interpretability. In this section, five representative applications are offered to explain how three-way decision solves the domain-specific uncertainty.

5.1. Concept-cognitive learning

Concept cognitive learning simulates the human process of understanding concepts, capturing the relationships and similarities among unknown concepts. This inherent ability to recognize and handle multiple related concepts naturally aligns with the multi-label classification paradigm. Yuan et al. [210] proposed an incremental method for updating the progressive fuzzy three-way concept. The renewed three-way concept is characterized by membership and non-membership simultaneously, providing a more compact concept description when new objects arrive. Guo et al. [211] constructed a three-way-based fuzzy-granular structure for incremental concept-cognitive learning, By introducing the big concept priority principle, the number of pseudo-concepts is significantly reduced. Zhang et al. [212] defined a three-way-based two-way concept-cognitive learning to efficiently deduce necessary and sufficient three-way concepts. Based on the incremental mechanism of object and attribute variation, the necessary and sufficient three-way concepts can be adaptively updated. Masuyama et al. [213] developed a continual learning framework for updating multi-label concepts by updating adaptive resonance theory-based clusters. The label probability is estimated via the distribution of representative instances and is robust to scaling differences in features. Li et al. [214] presented a novel concept drift detector to acquire high-dimensional multi-label concepts. The category of concept drift is determined by the differences in the relationship between selected features and labels. Ma et al. [215] established a transferable generative framework by incorporating the advantages of semantic embedding autoencoders, feature transformation networks, and feature generation networks for multi-label zero-shot learning. With the knowledge pre-trained from single-label images, the discrimination on dominant and minor object categories is significantly improved.

5.2. Emotion detection

The unprecedented growth of interconnections in cyberspace flourishes the development of network events by sharing versatile emotions. Precision emotion detection is raised for the imperative requirements of reputation safeguarding including nationality, industry, and personality. Subhashini et al. [216] pointed out that the boundary and intensity of emotions tend to be ambiguous and can be simultaneously associated with a given textual comment. Ren and Wang [217] adopted a transformation-based strategy on document-level multi-label emotion detection. The decision-theoretic rough set determines the positive and negative regions based on the intensity of sentiment words, while the boundary region is determined by the proportion of sentence-level sentiment polarity. Yang et al. [218] extended the sequential three-way decision model to deal with the problem of dynamic emotion updating. With the temporal-spatial multi-granularity structure, the corresponding discrimination is significantly improved for sentence-level emotion recognition as compared with the included networks. Chen et al. [219] developed a three-way feature generation method for sentence-level emotion recognition. The optimal feature representation is with a hierarchical structure determined by a fuzzy quotient space. Liu et al. [220] introduced an intuitionistic fuzzy set on a large-scale group decision-making model for landslide treatment selection. To evaluate the public's sentiment towards expert opinions, the decision-theoretic fashion of three-way decision ranks the trustworthy alternative in the sense of misclassification risk. Subhashini et al. [221] constructed a three-way decision model for customer preference analysis. With the low statistical correlation between fuzzy features and semantic features, the three-way opinion classification determines certain parts (i.e., positive region and negative region) via fuzzy features, while determining the uncertain parts (i.e., boundary region) from the deep learning of semantic features.

5.3. Medical diagnosis

Organ dysfunction is a terrible threat to health, making medical diagnosis becomes an issue with the property of cost-sensitive and interpretability-aware. Although many kinds of signals are automatically captured, how to precisely interpret the essential evidence in an explainable manner remains challenging. Wang et al. [222] proposed a cost-sensitive sequential three-way decision-making method to deal with multi-attribute group decision-making problems for emergency diagnosis. For each tripartition, the costs in the decision process and decision result are considered for a single doctor, providing an elemental cost distribution for group decision-making. Yue et al. [223] introduced evidential theory into deep convolutional neural networks to assist image-based medical decision-making. The deep convolution neural networks reclassify risky predictions by weighing the ratio of unknown evidence to conflicting evidence. Yin et al. [224] developed an inexact decision-theoretic fashion of a three-way reasoning model for medical diagnosis. With the strong indicators of increased belief/disbelief in diseases, all possible diseases are categorized into three groups for an arbitrary patient. Pham et al. [225] constructed a deep convolutional neural network for the chest X-ray-based thoracic disease diagnosis. By combining conditional training and label smoothing techniques, the model integrates with disease hierarchy and alleviates the potential bias from weak-supervised labels. Ashfaq et al. [226] designed a novel deterministic approach to optimize the evidential distribution in deep neural networks for individual healthcare. An additional layer was added to the project to transform from evidential space to semantic label embedding space, addressing uncertainties from out-of-distribution instances. Barandas et al. [227] explored the robustness of uncertainty quantification measures for cardiology diagnosis. The standardized measures can effectively quantify the trustworthy degree of the model and control the subjectivity of threshold-based calibration.

5.4. Recommender systems

Precise promotion requires a comprehensive understanding of not only the users and commodities but also the latent relationships. Recommender systems serve as a platform to bridge the two parties. While the trajectories of determining the preference for commodities are tractable, it is still uncertain precisely which set of commodities should be recommended to the user at the correct moment without compromising personal privacy. Zhang and Min [228] customized the decision-theoretic rough set to construct a three-way recommender system. In particular, the decision cost part is generalized to misclassification cost and teacher cost, while the original conditional probability is determined by random forest. Wu et al. [229] introduced a shadowed set to conduct three-way movie recommendations. To enhance the effectiveness of personalized recommendations, an instance-level neighborhood rough set measures the relative boundary of faithful user-based preference decisions, and a shadowed set deduces the formulation of tripartition with the principle of uncertainty invariance. Ye et al. [230] formulated the movie recommendation problem as a temporal-spatial sequential three-way decision model. A multi-stage cost-sensitive model is proposed to capture variations of three-way recommendations by updating recurrent neural networks sequentially. Bogaert et al. [231] evaluated the performance of representative multi-label classifiers for cross-selling in financial services. They conclude that classifier chains' binary relevance with Adaboost or random forest is a promising benchmark. Zhao et al. [232] developed a multi-behavior interactions-based model for streaming recommending cases. The instant and long-term user preferences are subsequently characterized by a multi-behavior learning module and attentive memory network and fused for performance boosting via a gate mechanism afterward. Perez et al. [233] constructed a semantic projection from the photos of items to the similarity of user preference from the interactions of social networks for the cold start recommendation problem. With an encoding of the general-purpose convolutional neural network, the preference similarity on semantic features is gradually reached via a cascade structure by minimizing the cross-entropy loss.

5.5. Social network analysis

Social network analysis studies the relationships and interactions in networks, often facing uncertainties in conflict and consensus judgments. Multi-label classification can facilitate behavior understanding by predicting multiple attributes or roles of these entities simultaneously while accounting for the inherent uncertainties in these relationships. This integration enhances the depth and accuracy of social network insights by leveraging the complex and interconnected nature of the data, particularly in contexts where conflict and consensus are critical. Liang et al. [234] explored the group consensus reaching in intersections between decision-makers and participants. The minimum adjustment consensus is obtained by measuring the maximum closeness degree between inconsistent decision-makers and the most influential one. Liang et al. [235] investigated large-scale group consensus mechanisms by simultaneously considering the frequent connections within groups and limited connections among subnetworks. With the optimized network topology, the three-way rules characterizing the competition relationship on social roles and interest conflicts are thus deduced. Xiao et al. [236] extended the consensus-reaching process to the multi-scale information system. The optimal scale combination induced by the three-way decision improved the robustness of the mutual trust relation. Shi et al. [237] presented a multi-label graph convolutional network to capture the uncertain many-to-many relationship between humans and interests. A unified framework is constructed to aggregate the intra-label interaction and node label properties, represented by two Siamese graph convolutional networks. Wen et al. [238] developed a multi-label learning framework to leverage multi-source information for user profiling in online games. With feature embedding and multi-relational graph embedding, the multi-source information is represented, and the underlying correlations are deduced using two variational autoencoders with disentangled latent spaces. Guo et al. [239] established a federated graph learning model for community detection with the requirement of privacy-preserving. To prevent the leakage of the sensitive user profile, the vertex degrees and label weights are concealed by a label perturbation strategy and homomorphic encryption, respectively.

6. Discussions

The ML-SD3WD model primarily focuses on enhancing the capability of uncertainty processing during the data analysis stage, leveraging the principles of three-way decision-making. While the model has shown significant improvements in handling uncertainty, it primarily addresses the higher-level aspects of data analysis. The underlying uncertainties associated with the quality of labels and the robustness of features, which are crucial for the initial stages of data preparation, are not fully covered. This section aims to discuss these advanced uncertainty issues and their potential solutions, providing a comprehensive view of the challenges and opportunities in real-world applications.

6.1. Limitations

Although the ML-SD3WD model substantially improves the capability of uncertainty processing, practitioners should be prudent in directly developing three-way-based multi-label solutions if the available information is inadequate. Insufficient information may come from labels or conditions or both, which is caused by factors including annotation cost, collection cost, inconsistent labeling practices, device failures, and knowledge absences. In this section, these issues are summarized as low-quality labels and primitive features.

1. Low-quality labels: Low-quality labels, characterized by uncertainties in label correlations or insufficient reliable label information, pose significant challenges to the application of three-way decision in multi-label learning. Compared with automatically generated data, label annotations are primarily determined by experts or volunteers. While expert labeling offers high accuracy, considerable expenditure is a bottleneck in collecting large-scale, high-quality labels. As a result, the volume of labeled data is often limited, which is called few-shot learning in the machine learning community. In contrast, crowdsourcing labeling is cost-effective and scalable, however, it suffers from varying levels of annotator expertise, resulting in inaccurate labels. Crowdsourcing can also lead to issues of incompleteness, noise, and inconsistency across the label space. The presence of flawed labels in the observable label space makes it difficult to learn the latent label correlations, thereby challenging the recognition of uncertain instance-label pairs for three-way-based models. Therefore, scholars should carefully identify the labeling type in their applications and make appropriate adjustments to the pipeline of three-way-based models.
2. Primitive features: Primitive features refer to modality-dependent but task-irrelevant representations. For example, in image processing, the values of RGB channels represent objects in different figures, and the ordered combinations of pixels denote varying scenarios. Standardized data processing reduces the cost of data preparation. However, artificially annotated labels often emphasize the priority of wide-range conditions, which differ from primitive features in terms of granularity. This discrepancy between cognitive representation and task-relevant information can degrade the generalization of three-way-based models if they are directly employed. Therefore, scholars should carefully examine the semantics of available

conditions and develop appropriate transformations to mitigate the side effects of the information gap.

6.2. Solutions

The main idea to address these limitations lies in finding a knowledge base that is beneficial for solving the problem. From the perspective of the source of the knowledge base, it generally falls into two categories: external knowledge associations and internal knowledge representations. Specifically, solutions can be constructed from the following aspects.

1. Pre-training and fine-tuning [240]: This technique is recommended for addressing the issue of limited and potentially low-quality labels incurred by expert labeling. By incorporating external knowledge trained on large-scale, high-quality datasets, the pre-training technique can provide a robust initial representation. This initial representation can be further fine-tuned using a smaller amount of labeled data specific to the task. Fine-tuning the limited but discriminative annotations plays a regularization term in the problem formulation, thereby preventing overfitting and enhancing the model's generalization ability.
2. Transfer learning [241]: This technique is recommended for addressing low-quality labels incurred by expert labeling. Both pre-training and transfer learning leverage external knowledge, but they differ in the characteristics of the external datasets. Pre-training requires a large-scale, task-independent dataset, thus incurring substantial computational overhead. In contrast, transfer learning utilizes a smaller dataset that is more closely related to the target task, making it more computationally efficient. The relationship between the source and target datasets in transfer learning is stronger than in pre-training, ensuring better performance with fewer resources.
3. Weakly supervised learning [242]: This technique is recommended for addressing low-quality labels incurred by crowdsourcing labeling. By leveraging the internal knowledge of latent label representation, weakly supervised learning transforms the flawed label supervision as a distribution over the instance-label pairs. Depending on whether the oracle is available, the strategy can be further categorized as active learning and semi-supervised learning. In the active learning setting, a small proportion of informative and representative instance-label pairs are consecutively annotated by experts, and the learning becomes effective when the label correlation is gradually approximated. In contrast, the semi-supervised estimates the maximal posterior probability of instance-label pair based on cluster and manifold assumption.
4. Deep learning [243]: This technique is recommended for addressing primitive features. By constructing multiple processing layers with functionalities such as convolution, pooling, and full connection, deep learning models can deduce internal structures with multiple levels of abstraction. For example, in image processing, the network progressively captures features such as edges, textures, and shapes. These features approximate the semantics of engineered features, and the variations in high-level features are more closely related to label associations. Consequently, the information gap is reduced, and three-way-based learning on hierarchical features becomes more effective.
5. Broad learning [244]: This technique is recommended for addressing primitive features. Both deep learning and broad learning exploit internal knowledge, but they differ in model structure. Unlike deep learning, which is time-consuming in searching for optimal solutions for millions of parameters across multiple processing layers, broad learning trains a flat network with only three layers: feature mapping, enhancement, and output. By reducing redundancy in group-based representation, broad learning can efficiently extract high-level features, including edges and textures. Consequently, the three-way-based learning on these high-level features becomes more effective.

It is worth mentioning that the three-way decision can reduce the uncertainty incurred by these techniques for discriminative labels and desirable features, especially when uncertainty is recognized. Therefore, the functionality of three-way decision in multi-label classification holds significant promise for knowledge discovery in various applications.

7. Conclusions

Multi-label classification is an advanced concept cognition task with the challenges originating from the generalized uncertain relationships among different concepts. Three-way decision can deduce a desirable solution by sequentially reducing the label correlation uncertainty. The decision-making fashion is a breakthrough of the two-way-based formulation, where the latter takes subjective assumptions of label correlation. Additionally, the structure of three-way decision is uncertainty-driven and can be collaborative with different granularity, which offers a promising schema for dealing with diversified kinds of uncertainty.

A systematic review of the developments of three-way decision from the perspectives of model adaptation is conducted. Through the categorization of three-way multi-label classification and investigation of uncertainty factors, it is evident that extensions on a particular three-way model may be suboptimal for multi-label cases. Based on the stratified uncertainty, a cascade triading-acting-optimizing framework called ML-SD3WD is developed, where topic generation, label assignment, and label enhancement are three essential procedures for correlation confidence distribution, label confidence distribution, and enhanced label confidence distribution, respectively. Furthermore, some uncertainty-related issues of fine-grained three-way representation, feature augmentation, robust label-specific features, topic generation, topic-based label correlation, adaptive loss function, cross-topic label aggregation, selective label enhancement, and trade-off between efficiency and effectiveness, are identified in customizing the ML-SD3WD model. The variations of these issues may substantially enrich the context of the three-way decision. Finally, the discussed limitations and solutions of three-way multi-label classification are conducive to enhancing uncertainty-driven application solutions.

CRediT authorship contribution statement

Yuanjian Zhang: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Tianna Zhao:** Writing – review & editing, Investigation, Conceptualization. **Duoqian Miao:** Supervision, Funding acquisition. **Yiyu Yao:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China "Key Special Project on Cyberspace Security Governance" under the Grant 2022YFB3104702, National Natural Science Foundation of China under the Grant 62376198.

Data availability

No data was used for the research described in the article.

References

[1] M.L. Zhang, Z.H. Zhou, A review on multi-label learning algorithms, IEEE Trans. Knowl. Data Eng. 26 (8) (2014) 1819–1837.
[2] W. Siblini, P. Kuntz, F. Meyer, A review on dimensionality reduction for multi-label classification, IEEE Trans. Knowl. Data Eng. 33 (3) (2021) 839–857.
[3] W.W. Liu, H.B. Wang, X.B. Shen, I.W. Tsang, The emerging trends of multi-label learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7955–7974.
[4] Y.J. Li, Y.N. Gao, B.Z. Chen, Z. Zhang, G.M. Lu, D. Zhang, Self-supervised exclusive-inclusive interactive learning for multi-label facial expression recognition in the wild, IEEE Trans. Circuits Syst. Video Technol. 32 (5) (2022) 3190–3202.
[5] X. Kang, X.F. Shi, Y.N. Wu, F.J. Ren, Active learning with complementary sampling for instructing class-biased multi-label text emotion classification, IEEE Trans. Affect. Comput. 14 (1) (2023) 523–536.
[6] H. Zhou, C.Y. Zhang, Y. Luo, C.P. Hu, W.J. Zhang, Thinking inside uncertainty: Interest moment perception for diverse temporal grounding, IEEE Trans. Circuits Syst. Video Technol. 32 (10) (2022) 7190–7203.
[7] M. Monfort, B.W. Pan, K. Ramakrishnan, A. Andonian, B.A. McNamara, A. Lascelles, Q.F. Fan, D. Gutfreund, R.S. Feris, A. Oliva, Multi-moments in time: Learning and interpreting models for multi-action video understanding, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2022) 9434–9445.
[8] Y.W. Li, Z.M. Zhang, F. Zhou, Y.T. Xing, J.Q. Li, C.Y. Liu, Multi-label classification of arrhythmia for long-term electrocardiogram signals with feature learning, IEEE Trans. Instrum. Meas. 70 (2021) 2512611.
[9] L. Yi, L. Zhang, X.Y. Xu, J.X. Guo, Multi-label softmax networks for pulmonary nodule classification using unbalanced and dependent categories, IEEE Trans. Med. Imaging 42 (1) (2023) 317–328.
[10] G. Holste, Y.L. Zhou, S. Wang, A. Jaiswal, M.Q. Lin, S. Zhuge, Y.Z. Yang, D. Kim, T.H. Nguyen-Mau, M.T. Tran, J. Jeong, W. Park, J. Ryu, F. Hong, A. Verma, Y. Yamagishi, C. Kim, H. Seo, M. Kang, L.A. Celi, Z.Y. Lu, R.M. Summers, G. Shih, Z.Y. Wang, Y.F. Peng, Towards. long tailed, Multi-label disease classification from chest X-ray: Overview of the CXR-LT challenge, Med. Image Anal. 97 (2024) 103224.
[11] M.L. Zhang, Z.H. Zhou, ML-KNN: A lazy learning approach to multi-label learning, Pattern Recognit. 40 (7) (2007) 2038–2048.
[12] Y. Yu, W. Pedrycz, D.Q. Miao, Multi-label classification by exploiting label correlations, Expert Syst. Appl. 41 (2014) 2989–3004.
[13] M.L. Zhang, L. Wu, LIFT: Multi-label learning with label-specific features, IEEE Trans. Pattern Anal. Mach. Intell. 37 (1) (2015) 107–120.
[14] G. Madjarov, D. Gjorgjevikj, S. Deroski, Two stage architecture for multi-label learning, Pattern Recognit. 45 (3) (2012) 1019–1034.
[15] S.J. Huang, J. Rong, Z.H. Zhou, Active learning by querying informative and representative examples, IEEE Trans. Pattern Anal. Mach. Intell. 36 (10) (2014) 1936–1949.
[16] A. Alali, M. Kubat, Prudent: a pruned and confident stacking approach for multi-label classification, IEEE Trans. Knowl. Data Eng. 27 (9) (2015) 2480–2493.
[17] J. Huang, G.R. Li, Q.M. Huang, X.D. Wu, Learning label-specific features and class-dependent labels for multi-label classification, IEEE Trans. Knowl. Data Eng. 28 (12) (2016) 3309–3323.
[18] L. Rokach, A. Schclar, E. Itach, Ensemble methods for multi-label classification, Expert Syst. Appl. 41 (16) (2014) 7507–7523.
[19] S. Nazmi, X.Y. Yan, A. Homaifar, E. Doucette, Evolving multi-label classification rules by exploiting high-order label correlations, Neurocomoputing 417 (2020) 176–186.
[20] R. Wang, S. Kwong, X. Wang, Y.H. Jia, Active k-labelsets ensemble for multi-label classification, Pattern Recognit. 109 (2020) 107583.
[21] Y.Z. Zou, X.G. Hu, P.P. Li, Y.H. Ge, Learning shared and non-redundant label-specific features for partial multi-label classification, Inform. Sci. 656 (2024) 119917.
[22] Y.H. Li, L. Hu, W.F. Gao, Label correlations variation for robust multi-label feature selection, Inform. Sci. 609 (2022) 1075–1097.
[23] Z. Zhang, Z. Zhang, J.L. Yao, L. Liu, J.Y. Li, G.Q. Wu, X.D. Wu, Multi-label feature selection via adaptive label correlation estimation, ACM Trans. Knowl. Discov. Data 17 (9) (2023) 3604560.
[24] H.X. Li, C. Zhang, X.Y. Jia, Y. Gao, C.L. Chen, Adaptive label correlation based asymmetric discrete hashing for cross-modal retrieval, IEEE Trans. Knowl. Data Eng. 35 (2) (2023) 1185–1199.
[25] X.Y. Jia, S.S. Zhu, W.W. Li, Joint label-specific features and correlation information for multi-label learning, J. Comput. Sci. Tech. 35 (2) (2020) 247–258.
[26] J.H. Dai, J.L. Chen, Y. Liu, H. Hu, Novel multi-label feature selection via label symmetric uncertainty correlation learning and feature redundancy evaluation, Knowl.-Based Syst. 207 (2020) 106342.
[27] J. Zhang, C.D. Li, D.L. Cao, Y.J. Lin, S.Z. Su, L. Dai, S.Z. Li, Multi-label learning with label-specific features by resolving label correlations, Knowl.-Based Syst. 159 (2018) 148–157.
[28] C.Q. Zhang, Z.W. Yu, H.Z. Fu, P.F. Zhu, L. Chen, Q.H. Hu, Hybrid noise-oriented multilabel learning, IEEE Trans. Cybern. 50 (6) (2020) 2837–2850.
[29] X.Y. Zhang, F. Min, G.J. Song, H. Yu, LSTC: When label-specific features meet third-order label correlations, Inform. Sci. 632 (2023) 617–636.
[30] S.P. Xu, X.B. Yang, H.L. Yu, D.J. Yu, J.Y. Yang, E.C.C. Tsang, Multi-label learning with label-specific feature reduction, Knowl.-Based Syst. 104 (2016) 52–61.
[31] Z.B. Yu, M.L. Zhang, Multi-label classification with label-specific feature generation: A wrapped approach, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2022) 5199–5210.
[32] Y.J. Lin, Q.H. Hu, J.H. Liu, X.Q. Zhu, X.D. Wu, MULFE: Multi-label learning via label-specific feature space ensemble, ACM Trans. Knowl. Discov. Data 16 (1) (2022) 5.
[33] Y. Zhu, J.T. Kwok, Z.H. Zhou, Multi-label learning with global and local label correlation, IEEE Trans. Knowl. Data Eng. 30 (6) (2018) 1081–1094.
[34] J.H. Ma, B.C.Y, Chiu, T.W.S. Chow, Multilabel classification with group-based mapping: A framework with local feature selection and local label correlation, IEEE Trans. Cybern. 52 (6) (2022) 4596–4610.
[35] Y. Yang, H.M. Chen, Y. Mi, C. Luo, S.J. Hong, T.R. Li, Multi-label feature selection based on stable label relevance and label-specific features, Inform. Sci. 648 (2023) 119525.
[36] Z. Zhang, L. Liu, J.Y. Li, X.D. Wu, Integrating global and local feature selection for multi-label learning, ACM Trans. Knowl. Discov. Data 17 (1) (2023) 4.
[37] W. Weng, B.W. Wei, W. Ke, Y.L. Fan, J.B. Wang, Y.W. Li, Learning label-specific features with global and local label correlation for multi-label classification, Appl. Intell. 53 (3) (2023) 3017–3033.
[38] X. Geng, Label distribution learning, IEEE Trans. Knowl. Data Enigneering 28 (7) (2016) 1734–1748.
[39] N. Xu, Y.P. Liu, X. Geng, Label enhancement for label distribution learning, 33(4), 2021, pp. 1632–1643.
[40] J. Hobbs, Granularity, in: Proceedings of the 9th International Joint Conference on Artificial Intelligence, 1985, pp. 432–435.
[41] J.D. Qin, L. Martinez, W. Pedrycz, X.Y. Ma, Y.Y. Liang, An overview of granular computing in decision-making: Extensions, applications, and challenges, Inf. Fusion 98 (2023) 101833.
[42] D.D. Guo, W.H. Xu, W.P. Ding, Y.Y. Yao, X.Z. Wang, W. Pedrycz, Y.H. Qian, Concept-cognitive learning survey: Mining and fusing knowledge from data, Inf. Fusion 109 (2024) 102426.
[43] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[44] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[45] K.T. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems 20 (1) (1986) 87–96.
[46] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[47] B. Ganter, R. Wile, Formal Concept Analysis: Mathematical Foundations, Springer, Berlin, Germany, 1999.
[48] W. Pedrycz, Shadowed sets: Representing and processing fuzzy sets, IEEE Trans. Syst. Man, Cybern. Part B: Cybern. 28 (1) (1998) 103–109.
[49] F. Li, D.Q. Miao, W. Pedrycz, Granular multi-label feature selection based on mutual information, Pattern Recognit. 67 (2017) 410–423.
[50] X.Y. Che, D.G. Chen, J.S. Mi, A novel approach for learning label correlation with application to feature selection of multi-label data, Inform. Sci. 512 (2020) 795–812.
[51] W.B. Qian, Y.S. Xiong, J. Yang, W.H. Shu, Feature selection for label distribution learning via feature similarity and label correlation, Inform. Sci. 582 (2022) 38–59.
[52] A.H. Tan, X.W. Ji, J.Y. Liang, Y.Z. Tao, W.Z. Wu, W. Pedrycz, Weak multi-label learning with missing labels via instance granular discrimination, Inform. Sci. 594 (2022) 200–216.
[53] S.D. Lin, C.X. Wang, Y. Mao, Y.J. Lin, Feature selection for label distribution learning under feature weight view, Int. J. Mach. Learn. Cybern. 15 (5) (2024) 1827–1840.
[54] Z.W. Bian, Q. Chang, J. Wang, W. Pedrycz, N.R. Pal, Takagi–sugeno-kang fuzzy systems for high dimensional multilabel classification, IEEE Trans. Fuzzy Syst. 32 (6) (2024) 3790–3804.
[55] Q. Liu, M.J. Cai, Q.G. Li, C.Q. Huang, Multi-label feature selection based on adaptive label enhancement and class-imbalance-aware fuzzy information entropy, Internat. J. Approx. Reason. 176 (2025) 109320.
[56] T.N. Zhao, Y.J. Zhang, D.Q. Miao, Granular correlation-based label-specific feature augmentation for multi-label classification, Inform. Sci. 689 (2025) 121473.
[57] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inform. Sci. 180 (3) (2010) 341–353.
[58] C. Gao, Y.Y. Yao, Actionable strategies in three-way decisions, Knowl.-Based Syst. 133 (2017) 141–155.

[59] Y.Y. Yao, Three-way decision and granular computing, Internat. J. Approx. Reason. 103 (2018) 107–123.
[60] Y.Y. Yao, The dao of three-way decision and three-world thinking, Internat. J. Approx. Reason. 162 (2023) 109032.
[61] L.W.Q. Suo, H. Yang, Q.Y. Li, H.L. Yang, Y.Y. Yao, A review of three-way decision: Triadic understanding, Organ. Perspect. Int. J. Approx. Reason. 173 (2024) 109268.
[62] C.M. Jiang, D.D. Guo, Y. Duan, Y. Liu, Strategy selection under entropy measures in movement-based three-way decision, Internat. J. Approx. Reason. 119 (2020) 280–291.
[63] C.M. Jiang, D.D. Guo, R.Y. Xu, Measuring the outcome of movement-based three-way decision using proportional utility functions, Appl. Intell. 51 (12) (2021) 8598–8612.
[64] C.M. Jiang, R.Y. Xu, P.X. Wang, Measuring effectiveness of movement-based three-way decision using fuzzy Markov model, Internat. J. Approx. Reason. 152 (2023) 456–469.
[65] D.D. Guo, C.M. Jiang, P. Wu, Three-way decision based on confidence level change in rough set, Internat. J. Approx. Reason. 143 (2022) 57–77.
[66] C.M. Jiang, D.D. Guo, Y. Duan, Measure effectiveness of change-based three-way decision using utility theory, Cogn. Comput. 14 (3) (2022) 1009–1018.
[67] D.D. Guo, C.M. Jiang, R.X. Sheng, S.S. Liu, A novel outcome evaluation model of three-way decision: A change viewpoint, Inform. Sci. 607 (2022) 1089–1110.
[68] J.F. Xu, Y.J. Zhang, D.Q. Miao, Three-way confusion matrix for classification: A measure driven view, Inform. Sci. 507 (2020) 772–794.
[69] L.J. Yin, Q.H. Zhang, F. Zhao, D. Liu, G.Y. Wang, Superiority of three-way decisions from the perspective of probability, Artif. Intell. Rev. 56 (2) (2023) 1263–1295.
[70] D.D. Guo, W.H. Xu, Y.H. Qian, W.P. Ding, M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[71] J.L. Duan, G.Y. Wang, X. Hu, D.Y. Xia, D. Wu, Mining multigranularity decision rules of concept cognition for knowledge graphs based on three-way decision, Inf. Process. Manage. 60 (4) (2023) 103365.
[72] J. Yang, J.C. Kuang, G.Y. Wang, Q.H. Zhang, Y.M. Liu, Q. Liu, D.Y. Xia, S. Li, X.Q. Wang, D. Wu, Adaptive three-way KNN classifier using density-based granular balls, Inform. Sci. 678 (2024) 120858.
[73] J. Yang, Z.Z. Liu, S.Y. Xia, G.Y. Wang, Q.H. Zhang, S. Li, T.H. Xu, 3WC-GBNRS++: A novel three-way classifier with granular-ball neighborhood rough sets based on uncertainty, IEEE Trans. Fuzzy Syst. 32 (8) (2024) 4376–4387.
[74] D.Y. Xia, G.Y. Wang, Q.H. Zhang, J. Yang, S.Y. Xia, Three-way approximations fusion with granular-ball computing to guide multigranularity fuzzy entropy for feature selection, IEEE Trans. Fuzzy Syst. 32 (10) (2024) 5963–5977.
[75] H.L. Yang, Y. Wang, Z.L. Guo, Three-way conflict analysis based on hybrid situation tables, Inform. Sci. 628 (2023) 522–541.
[76] G.M. Lang, W.P. Ding, D.Q. Miao, H. Fujita, Y.Y. Yao, Trisection-fusion and fusion-trisection methods of three-way conflict analysis with pythagorean fuzzy information, Appl. Soft Comput. 164 (2024) 111939.
[77] M.J. Hu, Three-way conflict analysis with preference-based conflict, Inform. Sci. 693 (2025) 121676.
[78] J.F. Xu, P. Xin, Y.J. Zhang, Three-way neighborhood based stream computing for incomplete hybrid information system, Knowl.-Based Syst. 262 (2023) 110232.
[79] Y.D. Huang, Y.J. Zhang, J.F. Xu, Incremental approaches for optimal scale selection in dynamic multi-scale set-valued decision tables, Int. J. Mach. Learn. Cybern. 14 (6) (2023) 2251–2270.
[80] X.F. Tan, C. Gao, J. Zhou, J.J. Wen, Three-way decision-based co-detection for outliers, Internat. J. Approx. Reason. 160 (2023) 108971.
[81] X.Y. Zhang, Z. Yuan, D.Q. Miao, Outlier detection using three-way neighborhood characteristic regions and corresponding fusion measurement, IEEE Trans. Knowl. Data Eng. 36 (5) (2024) 2082–2095.
[82] M.J. Du, J.Q. Zhao, J.R. Sun, Y.Q. Dong, M3 W: Multistep three-way clustering, IEEE Trans. Neural Netw. Learn. Syst. 35 (4) (2024) 5627–5640.
[83] Y.X. Chen, P. Zhu, Y.Y. Yao, An axiomatic framework for three-way clustering, Inform. Sci. 675 (2024) 120761.
[84] H.R. Ju, Y. Lu, W.P. Ding, J.X. Cao, X.B. Yang, Three-way evidence theory-based density peak clustering with the principle of justifiable granularity, Appl. Soft Comput. 152 (2024) 111217.
[85] Q.X. Zhang, J. Wu, D.Q. Miao, C.R. Zhao, Q. Zhang, Attentive multi-granularity perception network for person search, Inform. Sci. 681 (2024) 121191.
[86] Q.X. Zhang, D.Q. Miao, Q. Zhang, C.W. Wang, Y.P. Li, Zhang H.Y., Learning adaptive shift and task decoupling for discriminative one-step person search, Knowl.-Based Syst. 304 (2024) 112483.
[87] Z.J. Li, D.Q. Miao, Sequential end-to-end network for efficient person search, in: Proceedings of the 35th Association for the Advancement of Artificial Intelligence, 2021, pp. 2011–2019.
[88] J.F. Xu, Y.P. Wei, Y.X. Chen, A three-way trajectory privacy-preserving model based on multi-feature fusion, Appl. Soft Comput. 158 (2024) 111591.
[89] J. Qian, H.Y. Jiang, Y. Yu, H. Wang, D.Q. Miao, Multi-level personalized k-anonymity privacy-preserving model based on sequential three-way decisions, Expert Syst. Appl. 239 (2024) 122343.
[90] J. Qian, M.C. Zheng, Y. Yu, C.P. Zhou, D.Q. Miao, A dynamic anonymization privacy-preserving model based on hierarchical sequential three-way decisions, Inform. Sci. 686 (2025) 121316.
[91] Y.J. Zhang, D.Q. Miao, W. Pedrycz, T.N. Zhao, J.F. Xu, Y. Yu, Granular structure-based incremental updating for multi-label classification, Knowl.-Based Syst. 189 (2020) 105066.
[92] J.H. Liu, Y.J. Lin, J.X. Du, H.B. Zhang, Z.Y. Chen, J. Zhang, ASFS: A novel streaming feature selection for multi-label data based on neighborhood rough set, Appl. Intell. 53 (2) (2023) 1707–1724.
[93] D.Y. Deng, T. Chen, Z.X. Deng, K.Y. Liu, P.F. Zhang, Dynamic feature selection based on F-fuzzy rough set for label distribution learning, Int. J. Fuzzy Syst. 26 (8) (2024) 2688–2706.
[94] Y.J. Zhang, T.N. Zhao, D.Q. Miao, W. Pedrycz, Granular multilabel batch active learning with pairwise label correlation, IEEE Trans. Syst. Man, Cybernetics: Syst. 52 (5) (2022) 3079–3091.
[95] C.M. Zhu, D.J. Cao, S.P. Guo, R.G. Zhou, L. Wei, Y.L. Dong, D.Q. Miao, Weak-label-based global and local multi-view multi-label learning with three-way clustering, Int. J. Mach. Learn. Cybern. 13 (5) (2022) 1337–1354.
[96] W.B. Qian, Y.Q. Tu, J. Qian, W.H. Shu, Partial multi-label learning via three-way decision-based tri-training, Knowl.-Based Syst. 276 (2023) 110743.
[97] T.Y. Yin, H.M. Chen, Z.H. Wang, K.Y. Liu, Z. Yuan, S.-J. Horng, T.R. Li, Feature selection for multilabel classification with missing labels via multi-scale fusion fuzzy uncertainty measures, Pattern Recognit. 154 (2024) 110580.
[98] Y.Y. Yao, S.K.M. Wong, P. Lingras, A decision-theoretic rough set model, in: Z.W. Ras, M. Zemankova, M.L. Emrichm (Eds.), Methodologies for Intelligent Systems, 1990, pp. 17–25.
[99] Y.Y. Yao, S.K.M. Wong, A decision theoretic framework for approximating concepts, Int. J. Man-Mach. Stud. 37 (1992) 793–809.
[100] M.H. DeGroot, Optimal Statistical Decisions, McGraw-Hill, New York, 1970.
[101] R.O. Duda, P.E. Hart, Pattern Classification and Scence Analysis, Wiley, New York, 1973.
[102] R.L. Winkler, Introduction To Bayesian Inference and Decision, Holt, Rinehart and Winston, Inc., New York, 1972.
[103] D. Ślęzak, Rough sets and Bayes factor, in: J.F. Peters, A. Skowron (Eds.), Transactions on Rough Sets III, in: Lecture Notes in Computer Science, vol. 3400, 2005, pp. 202–229.
[104] B. Zhou, Multi-class decision-theoretic rough sets, Internat. J. Approx. Reason. 55 (2014) 211–224.
[105] X.Y. Jia, W.W. Li, L. Shang, A multiphase cost-sensitive learning method based on the multiclass three-way decision-theoretic rough set model, Inform. Sci. 485 (2019) 248–262.
[106] D.J. Ye, D.C. Liang, T. Li, S.J. Liang, Multi-class decision-making method for decision-theoretic rough sets based on the constructive covering algorithm, IEEE Access 8 (2020) 57833–57848.
[107] Y.Y. Yao, X.F. Deng, Sequential three-way decisions with probabilistic rough sets, in: Proceedings of the 10th IEEE International Conference on Cognitive Informatics and Cognitive Computing, 2011, pp. 120–125.
[108] Y.Y. Yao, Granular computing and sequential three-way decisions, in: P. Lingras, M. Wolski, C. Cornelis, S. Mitra, P.e. Wasilewski (Eds.), RSKT 2013, in: LNCS (LNAI), vol. 8171, 2013, pp. 16–27.
[109] Y.Y. Yao, Granular computing: Basic issues and possible solutions, in: Proceedings of the 5th Joint Conference on Information Sciences, Vol. 1, 2000, pp. 186–189.
[110] A.V. Savchenko, Fast multi-class recognition of piecewise regular objects based on sequential three-way decisions and granular computing, Knowl.-Based Syst. 91 (2016) 252–262.
[111] J. Qian, C.H. Liu, X.D. Yue, Multigranulation sequential three-way decisions based on multiple thresholds, Internat. J. Approx. Reason. 105 (2019) 396–416.
[112] X. Yang, T.R. Li, H. Fujita, D. Liu, A sequential three-way approach to multi-class decision, Internat. J. Approx. Reason. 104 (2019) 108–125.
[113] Y. Xu, J.X. Tang, X.S. Wang, Three sequential multi-class three-way decision models, Inform. Sci. 537 (2020) 62–90.
[114] J. Qian, C.H. Liu, D.Q. Miao, X.D. Yue, Sequential three-way decisions via multi-granularity, Inform. Sci. 507 (2020) 606–629.
[115] W.B. Qian, Y.Y. Zhou, J. Qian, Y.L. Wang, Cost-sensitive sequential three-way decision for information system with fuzzy decision, Internat. J. Approx. Reason. 149 (2022) 85–103.
[116] W.B. Chen, Q.H. Zhang, Y.Y. Dai, Sequential multi-class three-way decisions based on cost-sensitive learning, Internat. J. Approx. Reason. 146 (2022) 47–61.
[117] T.N. Zhao, Y.J. Zhang, D.Q. Miao, H.Y. Zhang, Multi-granular labels with three-way decisions for multi-label classification, Int. J. Mach. Learn. Cybern. 14 (2023) 3737–3752.
[118] X.Y. Che, D.G. Chen, J. Deng, J.S. Mi, Exploiting local label correlation from sample perspective for multi-label classification via three-way decision theory, Appl. Soft Comput. 149 (2023) 110950.
[119] Y. Yu, M. Wan, J. Qian, D.Q. Miao, Z.Q. Zhang, P.F. Zhao, Feature selection for multi-label learning based on variable-degree multi-granulation decision-theoretic rough sets, Internat. J. Approx. Reason. 169 (2024) 109181.
[120] T.N. Zhao, Y.J. Zhang, D.Q. Miao, W. Pedrycz, Selective label enhancement for multi-label classification based on three-way decisions, Internat. J. Approx. Reason. 150 (2022) 172–187.

[121] B. Yu, H.J. Xie, Y. Fu, Z.S. Xu, Three-way graph convolutional network for multi-label classification in multi-label information system, Appl. Soft Comput. 161 (2024) 111767.
[122] Y. Yu, W. Pedrycz, D.Q. Miao, Neighborhood rough sets based multi-label classification for automatic image annotation, Internat. J. Approx. Reason. 54 (2013) 1373–1387.
[123] H. Li, D.Y. Li, Y.H. Zhai, S.G. Wang, J. Zhang, A novel attribute reduction approach for multi-label data based on rough set theory, Inform. Sci. 367 (2016) 827–847.
[124] Y.J. Zhang, D.Q. Miao, Z.F. Zhang, J.F. Xu, S. Luo, A three-way selective ensemble model for multi-label classification, Internat. J. Approx. Reason. 103 (2018) 394–413.
[125] M.S. Liang, J.S. Mi, T. Feng, Optimal granulation selection for multi-label data based on multi-granulation rough sets, Granul. Comput. 4 (2019) 323–335.
[126] X.D. Fan, Q. Chen, Z.J. Qiao, C.Z. Wang, M.Y. Ten, Attribute reduction for multi-label classification based on labels of positive region, Soft Comput. 24 (18) (2020) 14039–14049.
[127] X. Wen, D.Y. Li, C. Zhang, Y.H. Zhai, A weighted ML-*k*NN based on discernibility of attributes to heterogeneous sample pairs, Inf. Process. Manage. 59 (5) (2022) 103053.
[128] Y.L. Wu, J.H. Liu, X.H. Yu, Y.J. Lin, S.Z. Li, Neighborhood rough set based multi-label feature selection with label correlation, Concurr. Comput. Pr. Exp. 34 (22) (2022) e7162.
[129] X.D. Fan, X.Y. Chen, C.Z. Wang, Y. Wang, Y. Zhang, Margin attribute reductions for multi-label classification, Appl. Intell. 52 (6) (2022) 6079–6092.
[130] Z.H. Wang, X.P. Zhang, The granulation attribute reduction of multi-label data, Appl. Intell. 53 (2023) 19266–19284.
[131] X.B. Liu, H.M. Zheng, W.X. Chen, L.Y. Xia, J.H. Dai, A novel multi-label feature selection method based on knowledge consistency-independence index, Inform. Sci. 677 (2024) 120870.
[132] Z.X. Deng, T.R. Li, D.Y. Deng, K.Y. Liu, P.F. Zhang, S.M. Zhang, Z.P. Luo, Feature selection for label distribution learning using dual-similarity based neighborhood fuzzy entropy, Inform. Sci. 615 (2022) 385–404.
[133] Y. Kou, G.P. Lin, Y.H. Qian, S.J. Liao, A novel multi-label feature selection method with association rules and rough set, Inform. Sci. 624 (2023) 299–323.
[134] J.H. Liu, Y.J. Lin, W.P. Ding, H.B. Zhang, C. Wang, J.X. Du, Multi-label feature selection based on label distribution and neighborhood rough set, Neurocomputing 524 (2023) 142–157.
[135] W.B. Qian, X.D. Long, Y.L. Wang, Y.H. Xie, Multi-label feature selection based on label distribution and feature complementarity, Appl. Soft Comput. 90 (2020) 106167.
[136] W.B. Qian, J.T. Huang, Y.L. Wang, Y.H. Xie, Label distribution feature selection for multi-label classification with rough set, Internat. J. Approx. Reason. 128 (2021) 32–55.
[137] W.B. Qian, F.K. Xu, J. Qian, W.H. Shu, W.P. Ding, Multi-label feature selection based on rough granular-ball and label distribution, Inform. Sci. 650 (2023) 119698.
[138] W.B. Qian, W.Y. Ruan, Y.H. Li, J.T. Huang, Granular ball-based label enhancement for dimensionality reduction in multi-label data, Appl. Intell. 53 (2023) 24008–24033.
[139] J.H. Liu, Y.J. Lin, Y.W. Li, W. Weng, S.X. Wu, Online multi-label streaming feature selection based on neighborhood rough set, Pattern Recognit. 84 (2018) 273–287.
[140] W.L. Chen, X. Sun, Dynamic multi-label feature selection algorithm based on label importance and label correlation, Int. J. Mach. Learn. Cybern. 15 (2024) 3379–3396.
[141] X.W. Lu, W.B. Qian, S.M. Dai, J.T. Huang, Label distribution feature selection based on hierarchical structure and neighborhood granularity, Inf. Fusion 112 (2024) 102588.
[142] S. Vluymans, C. Cornelis, F. Herrera, Y. Saeys, Multi-label classification using a fuzzy rough neighborhood consensus, Inform. Sci. 433 (2018) 96–114.
[143] Y.W. Li, Y.J. Lin, J.H. Liu, W. Weng, Z.K. Shi, S.X. Wu, Feature selection for multi-label learning based on kernelized fuzzy rough sets, Neurocomputing 318 (2018) 271–286.
[144] A.H. Tan, J.Y. Liang, W.Z. Wu, J. Zhang, L. Sun, C. Chen, Fuzzy rough discrimination and label weighting for multi-label feature selection, Neurocomputing 465 (2021) 128–140.
[145] W.B. Qian, C.Z. Xiong, Y.L. Wang, A ranking-based feature selection for multi-label classification with fuzzy relative discernibility, Appl. Soft Comput. 102 (2021) 106995.
[146] X.Y. Che, D.G. Chen, J.S. Mi, Label correlation in multi-label classification using local attribute reductions with fuzzy rough sets, Fuzzy Sets and Systems 426 (2022) 121–144.
[147] J.C. Xu, K.L. Shen, L. Sun, Multi-label feature selection based on fuzzy neighborhood rough sets, Complex Intell. Syst. 8 (3) (2022) 2105–2129.
[148] E.L. Yao, D.Y. Li, Y.H. Zhai, C. Zhang, Multilabel feature selection based on relative discernibility pair matrix, IEEE Trans. Fuzzy Syst. 30 (7) (2022) 2388–2401.
[149] X.Y. Che, D.G. Chen, J.S. Mi, Learning instance-level label correlation distribution for multilabel classification with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 31 (8) (2023) 2871–2884.
[150] H.B. Zhong, P. Zhang, G.X. Liu, Multi-label feature selection via redundancy of the selected feature set, Appl. Intell. 53 (9) (2023) 11073–11091.
[151] T.Y. Yin, H.M. Chen, Z. Yuan, J.H. Wan, K.Y. Liu, S.J. Horng, T.R. Li, A robust multilabel feature selection approach based on graph structure considering fuzzy dependency and feature interaction, IEEE Trans. Fuzzy Syst. 31 (12) (2023) 4516–4528.
[152] J.F. Jiang, X.Y. Zhang, Z. Yuan, Multi-label feature selection using self-information in divergence-based fuzzy neighborhood rough sets, Pattern Recognit. 155 (2024) 110684.
[153] J.H. Dai, W.Y. Huang, C.C. Zhang, J. Liu, Multi-label feature selection by strongly relevant label gain and label mutual aid, Pattern Recognit. 145 (2024) 109945.
[154] J.H. Dai, Q. Liu, W.X. Chen, C.C. Zhang, Multilabel feature selection based on fuzzy mutual information and orthogonal regression, IEEE Trans. Fuzzy Syst. 32 (9) (2024) 5136–5148.
[155] J.M. Liu, K. Huang, C. Chen, J. Mao, An oversampling algorithm of multi-label data based on cluster-specific samples and fuzzy rough set theory, Complex Intell. Syst. 10 (5) (2024) 6267–6282.
[156] W.B. Qian, F.K. Xu, J.T. Huang, J. Qian, A novel granular ball computing-based fuzzy rough set for feature selection in label distribution learning, Knowl.-Based Syst. 278 (2023) 110898.
[157] M.J. Cai, M. Yan, P. Wang, F. Xu, Multi-label feature selection based on fuzzy rough sets with metric learning and label enhancement, Internat. J. Approx. Reason. 168 (2024) 109149.
[158] Z.X. Deng, T.R. Li, D.Y. Deng, K.Y. Liu, Z.P. Luo, P.F. Zhang, Feature selection for handling label ambiguity using weighted label-fuzzy relevancy and redundancy, IEEE Trans. Fuzzy Syst. 32 (8) (2024) 4436–4447.
[159] L. Sun, Y.S. Chen, W.P. Ding, J.C. Xu, LEFSA: label enhancement-based feature selection with adaptive neighborhood via ant colony optimization for multilabel learning, Int. J. Mach. Learn. Cybern. 15 (2) (2024) 533–558.
[160] H. Li, Z.J. Wang, Multi-label-specific features learning algorithm based on label importance and fuzzy rough set, Int. J. Fuzzy Syst. (2024) http://dx.doi.org/10.1007/s40815-024-01776-2.
[161] J.H. Dai, Z.Y. Wang, W.Y. Huang, Novel multi-label feature selection via label enhancement and relative maximal discernibility pairs, Int. J. Mach. Learn. Cybern. 15 (8) (2024) 3237–3253.
[162] T. Denoeux, Z. Youne, F. Abdallah, Representing uncertainty on set-valued variables using belief functions, Artificial Intelligence 174 (7–8) (2010) 479–499.
[163] M.H. Masson, S. Destercke, T. Denoeux, Modelling and predicting partial orders from pairwise belief functions, Soft Comput. 20 (3) (2016) 939–950.
[164] Z. Mehravaran, J. Hamidzadeh, R. Monsefi, Feature selection based on correlation label and b-r belief function(FSCLBF) in multi-label data, Soft Comput. 28 (2) (2024) 1445–1457.
[165] J. Chen, Z.W. Du, X. Sun, S. Zhao, Y.P. Zhang, Granul. Comput. 6 (1) (2021) 59–68.
[166] G. Nápoles, M. Bello, Y. Salgueiro, Long-term cognitive network-based architecture for multi-label classification, Neural Netw. 140 (2021) 39–48.
[167] C. Cui, H. Huo, T. Fang, Deep hashing with multi-central ranking loss for multi-label image retrieval, IEEE Signal Process. Lett. 30 (2023) 135–139.
[168] J.M. Wu, E.C.C. Tsang, W.H. Xu, C.L. Zhang, L.Z. Yang, Correlation concept-cognitive learning model for multi-label classification, Knowl.-Based Syst. 290 (2024) 111566.
[169] L. Li, J.Z. Xu, Multi-label category enhancement fusion distillation based on variational estimation, Knowl.-Based Syst. 300 (2024) 112092.
[170] B. Yu, H.J. Xie, M.J. Cai, W.P. Ding, Mg-gcn: multi-granularity graph convolutional neural network for multi-label classification in multi-label information system, IEEE Trans. Emerg. Top. Comput. Intell. 8 (1) (2024) 288–299.
[171] C. Peng, H.B. Wang, J. Wang, L.D. Shou, K. Chen, G. Chen, C. Yao, Learning label-adaptive representation for large-scale multi-label text classification, IEEE-ACM Trans Audio Speech Language Process. 32 (2024) 2630–2640.
[172] Y.Z. Zou, X.G. Hu, P.P. Li, Gradient-based multi-label feature selection considering three-way variable interaction, Pattern Recognit. 145 (2024) 109900.
[173] X. Wei, J.B. Huang, R. Zhao, H. Yu, Z. Xu, Multi-label text classification model based on multi-level constraint augmentation and label association attention, ACM Trans. Asian Low-Resource Language Inf. Process. 23 (1) (2024) 14.
[174] B. Liu, A. Zhou, B.K. Wei, J. Wang, G. Tsoumakas, Oversampling multi-label data based on natural neighbor and label correlation, Expert Syst. Appl. 259 (2025) 125257.
[175] W.B. Qian, C.Z. Xiong, Y.H. Qian, Y.L. Wang, Label enhancement-based feature selection via fuzzy neighborhood discrimination index, Knowl.-Based Syst. 250 (2022) 109119.
[176] L. Sun, T.Y. Yin, W.P. Ding, Y.H. Qian, J.C. Xu, Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy, IEEE Trans. Fuzzy Syst. 30 (5) (2022) 1197–1211.
[177] S.P. Liang, Z. Liu, D.L. You, W.W. Pan, Online multi-label stream feature selection based on neighborhood rough set with missing labels, Pattern Anal. Appl. 25 (4) (2022) 1025–1039.

[178] W.H. Shu, Y.C. Hu, W.B. Qian, Multi-label feature selection for missing labels by granular-ball based mutual information, Appl. Intell. 54 (2024) 12589–12612.
[179] C.Q. Gull, J. Aguilar, A semi-supervised learning algorithm for multi-label classification and multi-assignment clustering problems based on a multivariate data analysis, Eng. Appl. Artif. Intell. 137 (2024) 109189.
[180] Z.M. Liu, J.H. Li, X. Zhang, X.Z. Wang, Multi-level information fusion for missing multi-label learning based on stochastic concept clustering, Inf. Fusion 115 (2025) 102775.
[181] W.B. Qian, Y.Q. Tu, J.T. Huang, W.P. Ding, Partial multi-label learning via robust feature selection and relevance fusion optimization, Knowl.-Based Syst. 286 (2024) 111365.
[182] F. Li, Q. Ai, X.N. Li, W.H. Wang, Q.Y. Gao, F. Zhao, Intuitionistic fuzzy least squares MLTSVM for noisy label data using label-specific features and local label correlation, Expert. Syst. with Appl. 260 (2025) 125351.
[183] T.Y. Yin, H.M. Chen, J.H. Wan, P.F. Zhang, S.J. Horng, T.R. Li, Exploiting feature multi-correlations for multilabel feature selection in robust multi-neighborhood fuzzy β covering space, Inf. Fusion 104 (2024) 102150.
[184] L. Sun, Q.F. Zhang, W.P. Ding, T.X. Wang, J.C. Xu, FCPFS: Fuzzy granular ball clustering-based partial multilabel feature selection with fuzzy mutual information, IEEE Trans. Emerg. Top. Comput. Intell. (2024) http://dx.doi.org/10.1109/TETCI.2024.3399665.
[185] W.B. Qian, P.Dong, Y.L. Wang, S.M. Dai, J.T. Huang, Local rough set based feature selection for label distribution learning with incomplete labels, Int. J. Mach. Learn. Cybern. 13 (2022) 2345–2364.
[186] W.B. Qian, P. Dong, S.M. Dai, J.T. Huang, Y.L. Wang, Incomplete label distribution feature selection based on neighborhood-tolerance discrimination index, Appl. Soft Comput. 130 (2022) 109693.
[187] T.Y. Yin, H.M. Chen, T.R. Li, Z. Yuan, C. Luo, Robust feature selection using label enhancement and β-precision fuzzy rough sets for multilabel fuzzy decision system, Fuzzy Sets and Systems 461 (2023) 108462.
[188] T.Y. Yin, H.M. Chen, Z. Yuan, B.B. Sang, S.J. Horng, T.R. Li, C. Luo, LEFMIFS: Label enhancement and fuzzy mutual information for robust multilabel feature selection, Eng. Appl. Artif. Intell. 133 (2024) 108108.
[189] J. Deng, D.G. Chen, H. Wang, R.F. Shi, Matrix factorization algorithm for multi-label learning with missing labels based on fuzzy rough set, Fuzzy Sets and Systems 498 (2025) 109143.
[190] W.W. Li, Y.Q. Lu, L. Chen, X.Y. Jia, Label distribution learning with noisy labels via three-way decisions, Internat. J. Approx. Reason. 150 (2022) 19–34.
[191] L. Sun, W.J. Du, J.C. Xu, B.F. Chang, Noise-resistant fuzzy multineighbourhood rough set-based feature selection with label enhancement and its application for multilabel classification, Appl. Soft Comput. 167 (2024) 112284.
[192] M. Bello, G. Nápoles, K. Vanhoof, R. Bello, Data quality measures based on granular computing for multi-label classification, Inform. Sci. 560 (2021) 51–67.
[193] A. Cano, J.M. Luna, E.L. Gibaja, S. Ventura, LAIM discretization for multi-label data, Inform. Sci. 330 (2016) 370–384.
[194] A.H. Tan, J.Y. Liang, W.Z. Wu, J. Zhang, Semi-supervised partial multi-label classification via consistency learning, Pattern Recognit. 131 (2022) 108839.
[195] A.N. Tarekegn, M. Giacobini, K. Michalak, A review of methods for imbalanced multi-label classification, Pattern Recognit. 118 (2021) 107965.
[196] M.X. Ding, Y.L. Yang, Z.Q. Lan, Multi-label imbalanced classification based on assessments of cost and value, Appl. Intell. 48 (10) (2018) 3577–3590.
[197] M.L. Zhang, Y.K. Li, H. Yang, X.Y. Liu, Towards class-imbalance aware multi-label learning, IEEE Trans. Cybern. 52 (6) (2022) 4459–4471.
[198] X.Y. Zhao, Y.X. An, N. Xu, X. Geng, Variational continuous label distribution learning for multi-label text classification, IEEE Trans. Knowl. Data Eng. 36 (6) (2024) 2716–2729.
[199] X.W. Ji, A.H. Tan, W.Z. Wu, S.M. Gu, Multi-label classification with weak labels by learning label correlation and label regularization, Appl. Intell. 53 (17) (2023) 20110–20133.
[200] J.F. Zhao, R. Song, C.T. Yue, Z.X. Wang, H. Xu, Weak-PMLC: A large-scale framework for multi-label policy classification based on extremely weak supervision, Inf. Process. Manage. 60 (5) (2023) 103442.
[201] T.Q. Deng, Q.W. Jia, J.Y. Wang, H. Fujita, Transformed schatten-1 penalty based full-rank latent label learning for incomplete multi-label classification, Inform. Sci. 650 (2023) 119699.
[202] Y. Wang, T. Zhang, C.W. Zhou, Z. Cui, J. Yang, Instance-aware deep graph learning for multi-label classification, IEEE Trans. Multimed. 25 (2023) 90–99.
[203] Y.L. Fan, J.H. Liu, J.E. Tang, P.Z. Liu, Y.J. Lin, Y.Z. Du, Learning correlation information for multi-label feature selection, Pattern Recognit. 145 (2024) 109899.
[204] Z. Qin, H.M. Chen, Y. Mi, C. Luo, S.J. Horng, T.R. Li, Multi-label feature selection with adaptive graph learning and label information enhancement, Knowl.-Based Syst. 285 (2024) 111363.
[205] P. Zhang, W.F. Gao, J.C. Hu, Y.H. Li, Multi-label feature selection based on the division of label topics, Inform. Sci. 553 (2021) 129–153.
[206] J.H. Ma, T.W.S. Chow, Latent topic-aware multioutput learning, IEEE Trans. Syst. Man, Cybern. Syst. 53 (12) (2023) 7547–7559.
[207] S.W. Qin, H. Wu, L.H. Zhou, Y.J. Zhao, L. Zhang, TAE: Topic-aware encoder for large-scale multi-label text classification, Appl. Intell. 54 (8) (2024) 6269–6284.
[208] K.H. Yuan, D.Q. Miao, Y.Y. Yao, H.Y. Zhang, X.R. Zhao, Feature selection using zentropy-based uncertainty measure, IEEE Trans. Fuzzy Syst. 32 (4) (2024) 2246–2260.
[209] K.H. Yuan, D.Q. Miao, W. Pedrycz, W.P. Ding, H.Y. Zhang, Ze-HFS: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery, IEEE Trans. Knowl. Data Eng. 36 (11) (2024) 7326–7339.
[210] K.H. Yuan, W.H. Xu, W.T. Li, W.P. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inform. Sci. 584 (2022) 127–147.
[211] D.D. Guo, W.H. Xu, Y.H. Qian, W.P. Ding, Fuzzy-granular concept-cognitive learning via three-way decision: performance evaluation on dynamic knowledge discovery, IEEE Trans. Fuzzy Syst. 32 (3) (2024) 1409–1423.
[212] C.L. Zhang, E.C.C. Tsang, W.H. Xu, Y.D. Lin, L.Z. Yang, J.M. Wu, Dynamic updating variable precision three-way concept method based on two-way concept-cognitive learning in fuzzy formal contexts, Inform. Sci. 655 (2024) 119818.
[213] N. Masuyama, Y. Nojima, C.K. Loo, H. Ishibuchi, Multi-label classification via adaptive resonance theory-based clustering, IEEE Trans. Pattern Anal. Mach. Intell. 45 (7) (2023) 8696–8712.
[214] P.P. Li, H.X. Zhang, X.G. Hu, X.D. Wu, High-dimensional multi-label data stream classification with concept drifting detection, IEEE Trans. Knowl. Data Eng. 35 (8) (2023) 8085–8099.
[215] P.R. Ma, Z.Q. He, W. Ran, H. Lu, A transferable generative framework for multi-label zero-shot learning, IEEE Trans. Circuits Syst. Video Technol. 34 (5) (2024) 3409–3423.
[216] L.D.C.S. Subhashini, Y.F. Li, J.L. Zhang, A.S. Atukorale, Y.T. Wu, Mining and classifying customer reviews: A survey, Artif. Intell. Rev. 54 (8) (2021) 6343–6389.
[217] F.J. Ren, L. Wang, Sentiment analysis of text based on three-way decisions, J. Intell. Fuzzy Systems 33 (1) (2017) 245–254.
[218] X. Yang, Y.J. Li, Q.K. Li, D. Liu, T.R. Li, Temporal-spatial three-way granular computing for dynamic text sentiment classification, Inform. Sci. 596 (2022) 551–566.
[219] J. Chen, Y. Chen, Y.C. He, Y. Xu, S. Zhao, Y.P. Zhang, A classified feature representation three-way decision model for sentiment analysis, Appl. Intell. 52 (7) (2022) 7995–8007.
[220] F. Liu, Z.L. Zhou, J. Wu, C.X. Liu, Y. Liu, Selection of landslide treatment alternatives based on LSGDM method of TWD and IFS, Complex Intell. Syst. 10 (2024) 3041–3056.
[221] L.D.C.S. Subhashini, Y.F. Li, J.L. Zhang, A.S. Atukorale, Assessing the effectiveness of a three-way decision-making framework with multiple features in simulating human judgement of opinion classification, Inf. Process. Manage. 59 (2) (2022) 102823.
[222] Y. Wang, B.Z. Sun, X.R. Zhang, Q. Wang, BWM and multimoora-based multigranulation sequential three-way decision model for multi-attribute group decision-making problem, Internat. J. Approx. Reason. 125 (2020) 169–186.
[223] X.D. Yue, Y.F. Chen, B. Yuan, Y. Lv, Three-way image classification with evidential deep convolutional neural networks, Cogn. Comput. 14 (6) (2022) 2074–2086.
[224] L.J. Yin, Q.H. Zhang, C.Y. Wu, Q. Mou, Novel three-way decision model in medical diagnosis based on inexact reasoning, Eng. Appl. Artif. Intell. 130 (2024) 107731.
[225] H.H. Pham, T.T. Le, D.Q. Tran, D.T. Ngo, H. Nguyen, Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels, Neurocomputing 437 (2021) 186–194.
[226] A. Ashfaq, M. lingman, M. Sensoy, S. Nowaczyk, DEED: Deep evidential doctor, Artificial Intelligence 325 (2023) 104019.
[227] M. Barandas, L. Famiglini, A. Campagner, D. Folgado, R. Simao, F. Cabitza, H. Gamboa, Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram, Inf. Fusion 101 (2024) 101978.
[228] H.R. Zhang, F. Min, Three-way recommender systems based on random forests, Knowl.-Based Syst. 91 (2016) 275–286.
[229] C.Y. Wu, Q.H. Zhang, F. Zhao, Y.L. Cheng, G.Y. Wang, Three-way recommendation model based on shadowed set with uncertainty invariance, Internat. J. Approx. Reason. 135 (2021) 53–70.
[230] X.Q. Ye, D. Liu, A cost-sensitive temporal-spatial three-way recommendation with multi-granularity decision, Inform. Sci. 589 (2022) 670–689.
[231] M. Bogaert, J. Lootens, D. Van den Poel, M. Ballings, Evaluating multi-label classifiers and recommender systems in the financial service sector, European J. Oper. Res. 279 (2) (2019) 620–634.
[232] Y. Zhao, S.J. Wang, Y. Wang, H.W. Liu, Mbsrs: A multi-behavior streaming recommender system, Inform. Sci. 631 (2023) 145–163.
[233] P. Pérez-Núñez, J. Díez, O. Luaces, B. Remeseiro, A. Bahamonde, Users' photos of items can reveal their tastes in a recommender system, Inform. Sci. 642 (2023) 119227.
[234] D.C. Liang, Y.Y. Fu, Z.S. Xu, Three-way group consensus decision based on hierarchical social network consisting of decision makers and participants, Inform. Sci. 585 (2022) 289–312.

[235] D.C. Liang, W.Y. Duan, Large-scale three-way group consensus decision considering individual competition behavior in social networks, Inform. Sci. 641 (2023) 119077.
[236] Y.B. Xiao, X.L. Ma, J.C.R. Alcantud, J.M. Zhan, A group consensus method based on social network and three-way decision under multi-scale information systems, Appl. Soft Comput. 162 (2024) 111824.
[237] M. Shi, Y.F. Tang, X.G. Zhu, J.X. Liu, Multi-label graph convolutional network representation learning, IEEE Trans. Big Data 8 (5) (2022) 1169–1181.
[238] X. Wen, S.W. Zhao, H.B. Wang, R.Z. Wu, M.H. Qu, T.L. Hu, G. Chen, J.R. Tao, C.J. Fan, Multi-source multi-label learning for user profiling in online games, IEEE Trans. Multimed. 25 (2023) 4135–4147.
[239] K. Guo, D.R. Chen, Q.Q. Huang, F. Li, C. Guo, D.J. Wu, X.M. Liu, K. Chen, Privacy-preserving multi-label propagation based on federated leraning, IEEE Trans. Netw. Sci. Eng. 11 (1) (2024) 886–899.
[240] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Conference of the North-American-Chapter of the Association-for-Computational-Linguistics-Human Language Technologies, 2019, pp. 4171–4186.
[241] F.Z. Zhuang, Z.Y. Qi, K.Y. Duan, D.B. Xi, Y.C. Zhu, H.S. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, Proc. IEEE 109 (1) (2021) 43–76.
[242] Z.H. Zhou, A brief introduction to weakly supervised learning, Natl. Sci. Rev. 5 (1) (2018) 44–53.
[243] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[244] C.L.P. Chen, Z.L. Liu, Broad learning system: An effective and efficient incremental learning system without the need for deep architecture, IEEE Trans. Neural Netw. Learn. Syst. 29 (1) (2018) 10–24.