ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

A SWARM BASED UNIFIED FRAMEWORK FOR MINING SEMANTICALLY OPTIMAL CONSISTENT PATTERNS ON CROSS-VIEW DATA

[1]Dr. N. Shanmuga Priya, [2]Mr. R. Srinithi, [3]Ms. V. Lavanya, [4]Mr. R. Vivin Richard

[1]Associate Professor and Head, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[2]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[3]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[4]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

ABSTRACT: -

Cross media retrieval plays an important role in identifying semantic consistency among the data that are being represented through different media. An appropriate framework is proposed for this process that initiates with Isomorphic Relevant Redundant Transformation (IRRT), which linearly renovates various heterogeneous low-level feature spaces to a top-level redundant feature-isomorphic space without any dimensionality reduction, i.e., without data loss. Furthermore, in addition to the semantic consistency among the isomorphic illustrations in the mid-level space, a new hybrid method is suggested which is the hybrid of Genetic Algorithm and Correlation-based Joint Feature Learning (CJFL) model to extract a unique optimal high-level semantic subspace shared across the feature-isomorphic data. Accordingly, the Semantic Consistent Pattern (SCP) for cross-view data shall be obtained. The main intention of this research is to instantaneously optimize the parameters and feature subset without corrupting the accuracy. While comparing the frameworks using MAP score by keeping Normalized correlation as the distance metric, it is confirmed that the proposed framework forms the better classifier system than the existing framework. Extremely, the proposed effective framework has its extensive applications in the field of the information retrieval system that works on cross view data.

Keywords: *Cross-View, Cross-Media, Shared Feature Subspace Learning, Heterogeneous Data, Dimensionality Reduction, Semantic Consistent Pattern, Genetic Algorithm.*

1. INTRODUCTION

The unimaginable growth of the IT makes cross-view data generally available in the real world. Cross-View data defined as the multiple similar small-scale or aggregate entities symbolized with different forms, backgrounds or modalities. Based on the consistent pattern generated with the complementary information from different views for those applications. The representations of different views can be integrated into a definite Semantically Consistent Patterns (SCP) covers overall complementary information from the rest and then the resulting consistent representation will be more constructive for entirely using the complementarily among different views. However, it is a challenging task to mine the SCP for cross-view data.

Firstly different views span heterogeneous low-level feature spaces; there is no clear correspondence among the cross-view representations. For an instance, the co-occurring image and text in a web page deliver the same semantic concept from the perspectives of vision and writing, respectively. So it is hard to measure the relationship between them based on their own heterogeneous representations directly. Therefore, to relate different views, an issue to be first addressed is to figure out a mid-level feature-isomorphic space, in which the corresponding information from different views will be completely embedded. In the interim, the isomorphic representation in the mid-level space, it can be expected as illustrated in Fig. 1 that it is mainly composed of requisite, redundant, and noisy components, respectively [1].

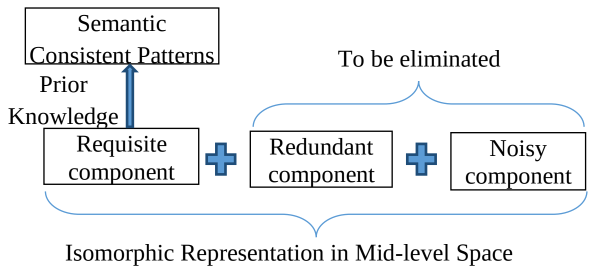

Fig 1: - Components of Isomorphic Representation in Mid-level Space

The requisite component refers to the complimentary information among isomorphic representations that is necessary for building the Semantically Consistent Patterns (SCP) with prior knowledge. Contrasting to the requisite component, the latter two refer to non-requisite information.

The redundant component takes high relativity with the requisite component, whereas the noisy one takes no dependence with both the requisite and redundant components. Hence, another issue to be dealt with mining the SCP is to extract at least a unique high-level semantic subspace shared through the feature-isomorphic data. Consequently, the requisite component can be well conserved without the redundant and noisy components being remained.

There are two models are proposed namely, Isomorphic Relevant Redundant Information (IRRT) and Correlation-based Joint Feature Learning (CJFL). Using this technique, a unique method of retrieving (mining) semantically consistent patterns for cross view data, IRRT model is used to linearly map multiple heterogeneous low-level feature spaces to high dimensional redundant feature spaces, to build mid-level isomorphic feature space. Due to some redundant data, noise still remains in it. In order to remove redundant information and noise from mid- level feature space, Hybrid GA-CJFL algorithm is used. The GA is used to get the optimal feature space by fine-tuning CJFL method without the degradation in accuracy.

The remaining of this paper is organised as follows: the background of the research work is described in section 2. Section 3 completely covers the technical details of the semantically mining the optimal consistent pattern using Hybrid GA-CJFL method. The experimental results and analysis is described in section 4 and the final conclusion is presented in section 5.

2. BACKGROUND STUDY

Some of the associated works are described as follows. To handle the heterogeneity across different views some traditional techniques based on statistical analysis has been proposed earlier, such as Canonical Correlation Analysis (CCA) and Partial Least Squares (PLS) to identify the low dimensional information of both views simultaneously, by increasing the correlation and covariance as discussed by Sun et al [2] and World [3] respectively. But, among them CCA is majorly used for cross-media retrieval according to Thompson [4], by extracting features in multi-view problem. It's useful for classification and clustering of multi-view data in accordance with Hardoon et al. [5] and Chaudhri et al [6]. CCA is merely equivalent to Linear Discriminant Analysis (LDA) as explained by Sun et al [2] but without class label. Different views obtained through

CCA uses a common representation. Hardoon et al [5] had used Kernel CCA to know what kind of representation is possible for images and its associated text.

Chaudhri et al. [6] had discussed an idea of projecting multiple views of data into a low dimension view among them. In addition to it, CCA usage has been prolonged to cross-view classification and retrieval by Sharma et al. Still, CCA has failure as it may not extract useful descriptors for general classification according to Thompson. Whereas, Kernel CCA changes its inability by changing the illustrations non-linearly to the high dimension feature subspace. Some of the other transformation methods like View Transformation Model (VTM) using Support Vector Regression (SVR) as explained by Kusakunniran et al. [7] where relativity motion is regressed and gait is recognized under multiple view angles, Multi-view Active Appearance Model (MAAM) as discussed by Ramnath et al., Multi-View Face Detector (MVFD), Vector Boosting algorithm (VB) as studied by Hung et al for multi-view classification [8].

As noise is present in feature-isomorphic space that's point to incorrect dealing of existing features of data. The features of cross-view data has shared subspace among them which consist of co-occurring features of those data which leads to appreciable noise elimination and it has been proved. The different types of shared subspace learning algorithms are available to capture the semantic consistency among the representations of data. In those algorithms, some of the shared subspace learning algorithms may involve multi-task learning as explained by Ando and Zhang, a framework for learning predictive structures from a set of multiple tasks and data that are unlabelled, also in accordance with Argyriou et al. [9], also as per Hui who dealt with multitask clustering, and Fei and Huan dealt with multi-label [9-11] and multi-class classification, and matrix – factorization.

`As proposed by Ando and Zhang [12-16] the Alternating Structure Optimization (ASO) algorithm to learn the predictive structure from multiple tasks in multi-task learning. Yet, it is non-convex so not assured of finding globally optimal solution. ASO also have its enhanced version, improved ASO (iASO), which is not convex one. But then again, there are also some convex framework called Convex Multi-Task Feature Learning (CMTFL) [8]. When no separate training dataset is used for entire tasks then multi-task and multi-label learning looks equivalent.

3. PROPOSED GA-CJFL FOR SEMANTICALLY MINING OPTIMAL CONSISTENT PATTERNS

The main objectives of this work are described in this section. A new framework that is used to extract optimal pattern for cross-view data is a two-step process. One is to map the low-level heterogeneous feature space to common high-level feature space. Second is used for making use of the linear transformation for classification with swarm method. Isomorphic Relevant Redundant Transformation (IRRT) is specifically used for linearly mapping multiple heterogeneous feature space to the homogeneous data view. A hybrid GA-CJFL known as shared subspace learning algorithm is used to determine the transformation that helps to classify data. The thorough explanation of proposed method is discussed in further section.

3.1. Construction of Feature-isomorphic Space

In proposed framework, Isomorphic Relevant Redundant Transformation (IRRT) is used to build the feature isomorphic space from low-level feature spaces of different views. At this juncture CCA, a classical method is also involved and its dimensionality reduction problem had increased its limitations. This construction of feature-isomorphic space can be generally expressed as:

$$\min_{A,B} \|XA - YB\|_F^2$$

s.t $A^T X^T XA = I$ and $B^T Y^T BY = I$

Where $A \in R^{dxxp}, B \in R^{dyxp}$ and $p \in \{1, \dots, \min(d_x, d_y)\}$. In above equation that the point $(Y * B)_{i,p}$is projected to the point$(X * A)_{i,p}$ for $i = 1, \dots, n$. Then, feature-isomorphic space representation for the given dataset is obtained, by using the optimal values of A^*andB^*. Thus, the required transformation from the initial dataset representation to feature-isomorphic space representations is expressed as:

$\mu_X = A^{*T}X$ and $\mu_Y = B^{*T}Y$

The above equation can be integrated to:

$$\mu = \frac{(\mu_X + \mu_Y)}{2}$$

Here p is limited range of values in CCA. Since, p has its maximum value as$\min(\mathrm{d_x}, \mathrm{d_y})$ in CCA, the resultant space have its new dimension as p. So the dimension of the resulting view lies in the range of$[1, \min(\mathrm{d_x}, \mathrm{d_y})]$. This leads to loss of corresponding information due to the reduction of dimension of a view to that of another. This is called dimensionality reduction problem. But IRRT uses trace norm limitations for this mapping, and with no loss in complementary information. This proves IRRT overcomes dimensionality reduction problem. The resulting optimization problem:

$$\min_{\mathrm{A,B}} \|\mathrm{XA} - \mathrm{YB}\|_{\mathrm{F}}^2$$

Lemma 1: s.t $\|\mathrm{XA}\|_* \le \varepsilon$ and $\|\mathrm{YB}\|_* \le \gamma$

Where e and γ are positive parameters to control the effect of dimensionality reduction. In [17] it is proved that IRRT is simply not an extension of CCA. In IRRT $\mathrm{p} \gg \max(\mathrm{d_x}, \mathrm{d_y})$ is also achieved. Also unlike CCA which uses orthogonal constraints for the projection to feature isomorphic space, IRRT uses low rank constraints in order to achieve linear mapping from multiple views to feature-isomorphic space. In accordance with Lemma 1 of Zhang, Lei and et al. parameters used here satisfies

$\|\mathrm{X}\|_*\|\mathrm{A}\|_* \le \varepsilon$ and $\|\mathrm{Y}\|_*\|\mathrm{B}\|_* \le \gamma$

$\|\mathrm{A}\|_* \le \frac{\varepsilon}{\|\mathrm{X}\|_*}$ and $\|\mathrm{B}\|_* \le \frac{\gamma}{\|\mathrm{Y}\|_*}$

Thus lemma 1 can be formulated as

$$\min_{\mathrm{A,B}} \|\mathrm{XA} - \mathrm{YB}\|_{\mathrm{F}}^2$$

Lemma 2: s.t $\|\mathrm{A}\|_* \le \frac{\varepsilon}{\|\mathrm{X}\|_*}$ and $\|\mathrm{B}\|_* \le \frac{\gamma}{\|\mathrm{Y}\|_*}$

The details are given in Algorithm 1 in [17]. According to Zhang et al [17] in the efficient solver of Lemma 2, the usage of Accelerated Projected Gradient (APG) algorithm for Euclidean projection of any given point is clearly presented in Algorithm 2 [17].

3.2. Mining the Semantic Consistency among Isomorphic Representations

In this section, a new shared subspace learning algorithm, called Correlation-based Joint Feature Learning (CJFL) model, to mine the semantic consistency among isomorphic representations, and display show to solve the CJFL model. Here Genetic Algorithm is blended with CJFL to produce optimal feature space by tuning the essential parameters of CJFL. By manipulating the correlations diagonally isomorphic representations, CJFL could extract a unique high-level semantically shared subspace. At this subspace the requisite component will be maintained to a large extent without the redundant and noisy information being remained. Similarly, the SCP for cross-view data can be obtained. Specifically, let (A^*, B^*) be the optimal solutions of the problem Lemma 2. Then we have the sets of isomorphic relevant redundant representations $J = \{a_i = A^{*T}x_i\}_{i=1}^n$ and $R = \{b_i = B^{*T}y_i\}_{i=1}^n$. Let C_X^t and C_Y^t be the sample set of t-th class from J and R, respectively. It can be defined

$$S_X^t = \{(a_i, a_j) | a_i, a_j \in C_X^t, i \ne j\},$$

$$S_Y^t = \{(b_i, b) | b_i, b_j \in C_Y^t, i \ne j\}$$

$$D_X^{tk} = \{(a_i, a_j) | a_i \in C_X^t \wedge a_j \in C_X^k, i \ne j, t \ne k\}$$

$$D_Y^{tk} = \{(b_i, b_j) | b_i \in C_Y^t \wedge b_j \in C_X^k, i \ne j, t \ne k\}$$

Let $S_X = U_t S_X^t$ and Let $S_X = U_t S_Y^t$

$\mathcal{D}_X = U_t U_k \mathcal{D}_X^{tk}$ and $\mathcal{D}_Y = U_t U_k \mathcal{D}_Y^{tk}$

Obviously, each pair of data from $S_X$or $S_Y$is semantically similar to each other and the one from $\mathcal{D}_X$or$\mathcal{D}_Y$is semantically dissimilar to each other.

And $J_{\mathcal{D}} + R_{\mathcal{D}}$is a joint between-class scatter matrix from both J and R. To simultaneously minimize the within-class distance while maximizing the between class distance, it is straightforward to formulate the above problem as a trace ratio optimization problem:

$$\Omega_1: \max_{\Theta^T\Theta=I} \frac{tr(\Theta^T(J_S + R_S)\Theta)}{tr(\Theta^T(J_{\mathcal{D}} + R_{\mathcal{D}})\Theta)}$$

Where the orthogonal constraint for Θ is used to eliminate the redundant information in the mid-level space, which takes high relativity with the requisite component. Unlike the scatter matrices in LAD[18], SSP [19], and TROP [20], both the joint within class and between-class scatter matrices $J_S + R_S$and$J_{\mathcal{D}} + R_{\mathcal{D}}$make a full use of the identity of sample distributions from different views in the mid-level feature-isomorphic space. On the other hand, the complementarily across isomorphic representations should be well preserved. Thus, we can re define the formulation

$$\Omega_2: \max_{\Theta^T\Theta=I} \frac{tr(\Theta^T(J_{\mathcal{D}} + R_{\mathcal{D}})\Theta)}{tr(\Theta^T(J_S + R_S)\Theta) + \alpha\|J\Theta - R\Theta\|_F^2}$$

Where the term $\|J\Theta - R\Theta\|_F^2$denotes the correlation based residual to avoid violating the intrinsic structure of the coupled representations, the regularization term $\|\Theta\|_F^2$controls the complexity of the model, and α, β > 0 are the regularization parameters.

The optimal $\Theta^*$for the problem of $\Omega_2$can be obtained by maximizing the following trace difference problem:

$$\Theta^* = \arg\max_{\Theta^T\Theta=} tr\left(\Theta^T\left(J_{\mathcal{D}} + R_{\mathcal{D}} - \eta_t(J_S + R_S + \alpha(J^TJ - 2J^TR + R^TR) + \beta I)\right)\Theta\right)$$

Where $\eta_t$the trace ratio value of the t-th iteration is, hence, $\Theta^*$is composed of the Eigen vectors corresponding to the k largest eigenvalues of the matrix$J_{\mathcal{D}} + R_{\mathcal{D}} - \eta_t(J_S + R_S + \alpha(J^TJ - 2J^TR + R^TR) + \beta I)$.Here, the iterative algorithm proposed in [20] to solve the problem in above equation. The details are given in Algorithm 1.

Input: an arbitrary columnly orthogonal matrix Θ_0, the matrices $J, R, J_{\mathcal{D}}, R_{\mathcal{D}}, J_S, R_S$ and positive integer h, two positive numbers α, β and max-iter

Output:Θ^*

1. For t=0,1,2,…,max-iterate do
2. Compute η_t
3. Perform Eigen-decomposition of the matrix $J_{\mathcal{D}} + R_{\mathcal{D}} - \eta_t(J_S + R_S + \alpha(J^TJ - 2J^TR + R^TR) + \beta I)$
4. Θ_{t+1} is given by the column vectors of the matrix P corresponding to the h largest eigen value
5. End for
6. Set optimal $\Theta^* = \Theta_{t+1}$

In algorithm 1, Let (A^*, B^*) be the optimal solution of the problem Lemma 2 and Θ^*be the optimal one of the problemΩ_2. Then, for the i-th couple of heterogeneous representations(x_i, y_i), here can obtain their own isomorphic relevant representations with the optimalA^*, B^*, and Θ^*as follows:

$ךּ_{x_i} = \Theta^{*T} A^{*T} x_i$ and$ךּ_{y_i} = \Theta^{*T} B^{*T} y_i$. In addition, we can exploit the consistent representation $ךּ_i$ of different views, i.e., the Semantically Consistent Patterns (SCP) for the cross-view data in the semantically shared subspace based on $ךּ_{x_i}$ and $ךּ_{y_i}$:

$$\daleth_i = \frac{(\daleth_{x_i} + \daleth_{y_i})}{2}$$

3.3. GA-Based Optimization of CJFL

GA is pragmatic to the optimization of CJFL parameters. The individual of GA consists of two CJFL parameters: J and R. The fitness of GA is calculated by $fitness = W_A \times CFJL_{accuracy} + W_F \times F_i \; \forall i = 1, .., n_f$, where $W_A$weight of the mining accuracy is, W_F is the weight of number of features (n_f).The proposed GA-Based Optimization of CJFL algorithm is given in Algorithm 4 and illustrated in Figure 2.

Input: an arbitrary columnly orthogonal matrixΘ_0, the matrices $J, R, J_{\mathcal{D}}, \mathrm{R}_{\mathcal{D}}, J_S, \mathrm{R}_S$ and positive integer h, two positive numbers α, β and max-iteration, number of population

Output: Optimal Θ^*

1. Create initial population and perform crossover and mutation
2. Fitness Evaluation
3. If stopping criteria is satisfied then
4. Display result of optimal feature space and optimal J, R
5. Else go to step 2
6. For t=0,1,2,…,max-itera do
7. Compute η_t
8. Perform Eigen-decomposition of the matrix $J_{\mathcal{D}} + \mathrm{R}_{\mathcal{D}} - \eta_t(J_S + \mathrm{R}_S + \alpha(J^TJ - 2J^TR + R^TR) + \beta I)$
9. Θ_{t+1} is given by the column vectors of the matrix P corresponding to the h largest eigen value
10. End for
11. Set optimal $\Theta^* = \Theta_{t+1}$

4. RESULTS & DISCUSSIONS

In this section, we assess and analyze the effectiveness of the learned SCP by the proposed framework for cross-view data. Our experiments are closely conducted on three publicly available cross-view datasets, namely, UCI Multiple Features (UCI MFeat) [21], COREL 5K [22], and Wikipedia [23].

4.1. Accuracy Comparison

The proposed Hybrid GA-CJFL method is having the average correct rate of 97 whereas the SCP and CJFL method having 92% and 96% of accuracy results, respectively. The overall accuracy percentage details are shown in fig 2.

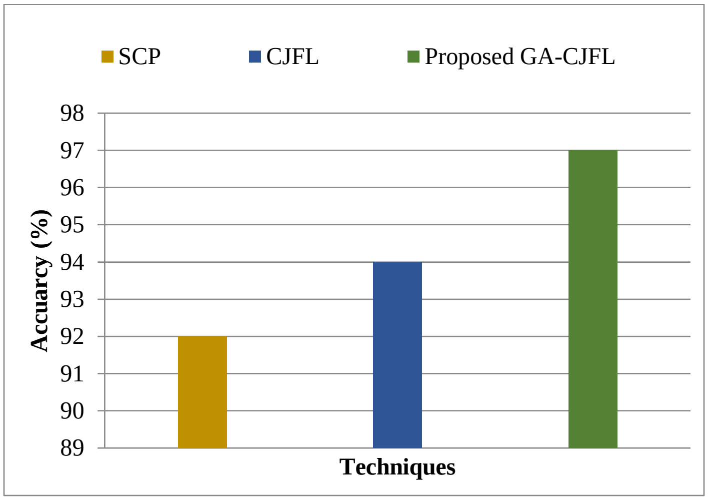

Fig 4.1: - The Overall Accuracy Percentage Comparative Result

4.2. Precision Comparison

The Fig.3 shows the precision comparison result of existing SCP, CJFL and proposed GA-CJFL algorithm. From the Fig.3, it is well known that the proposed system works better than existing system with the high precision result of 92%. The existing system has accuracy result which is less than the proposed GA-CJFL. The reason is that the proposed system has high convergence rate than the CJFL algorithm.

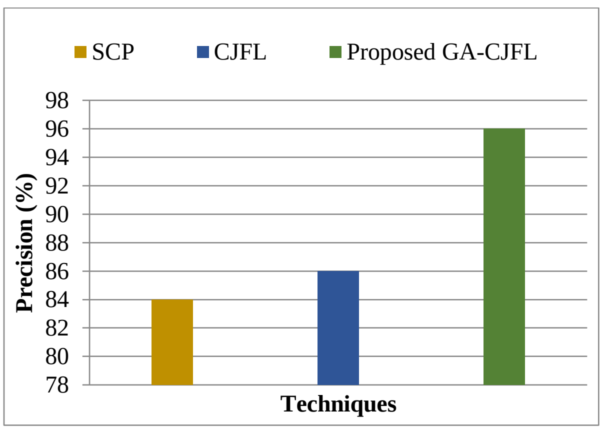

Fig 4.2: - The Overall Precision Percentage Comparative Result

4.3. Recall Rate Comparison

The Fig.4 shows the recall comparison result of existing SCP, CJFL and proposed GA-CJFL algorithm. From the Fig.4, it is obvious that the proposed system has high recall rate of 0.94 which is higher than the existing algorithms. The reason is that the proposed system has less execution time than the existing algorithms.

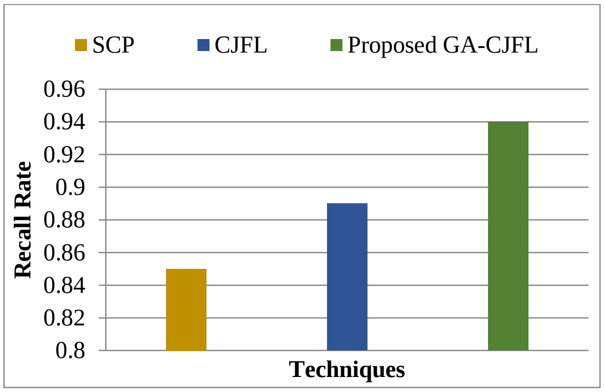

Fig 4.3: - The Overall Recall Rate Comparative Result

4.4. ROC Comparison

The Fig.5 shows the ROC comparison result of existing SCP, CJFL and proposed GA-CJFL algorithm. From the Fig.5, it is obvious that the proposed system works better than the existing algorithms. When the input is increases based on the false positive rate the true positive rate also increases. The reason is that the proposed system has high optimal classification than the existing algorithms.

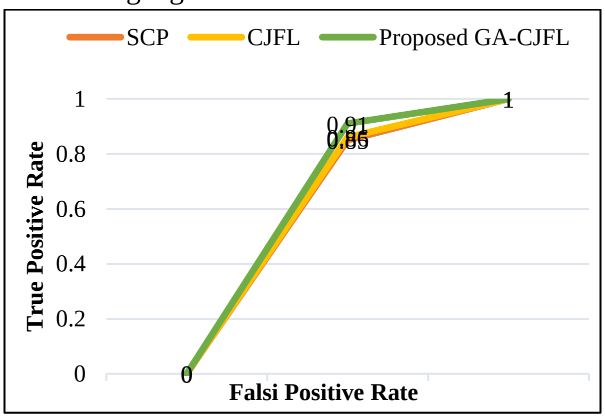

Fig 4.4: - ROC Comparison Result

5. CONCLUSION

In this framework, IRRT and GA-CFJL had played a major role to map optimal feature space to semantically shared space, where the semantic consistency has been mined. In this framework, second order approximation is used in IRRT method and GA is used to project feasible domain set onto the convex set in CFJL. Advancements in GA and approximations will be helpful to enhance the proposed framework in order to gain better results.

REFERENCES

[1] Xu, Y., Yin, Y., Wang, J., Wei, J., Liu, J., Yao, L., & Zhang, W. (2021). Unsupervised cross-view feature selection on incomplete data. Knowledge-Based Systems, 234, 107595.

[2] Hu, Z., Gutiérrez-Basulto, V., Xiang, Z., Li, R., & Pan, J. Z. (2024). COTET: Cross-view Optimal Transport for Knowledge Graph Entity Typing. arXiv preprint arXiv:2405.13602.

[3] Cheng, Qiang and Zhou, Hongbo and Cheng, Jie, “The FisherMarkov Selector Fast Selecting Maximally Separable Feature Subset for Multiclass Classification with Applications to High Dimensional Data,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.33, no.6, pp.1217-1233, 2011.

[4] Sun L, Ji S, Ye J. Canonical correlation analysis for multi-label classification: A least-squares formulation, extensions, and analysis. IEEE Trans Pattern Anal Mach Intell. 2011; 33(1):194–200.

[5] Wold H. Partial least squares. Encyclopedia of statistical sciences: Wiley Online Library; 2006.

[6] Thompson B. Canonical correlation analysis. Encyclopedia of statistics in behavioral science: Wiley Online Library; 2005.

[7] Hardoon D, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. Neural Comput. 2004; 16(12):2639–64.

[8] Chaudhuri K, Kakade SM, Livescu K, Sridharan K. Multi-view clustering via canonical correlation analysis. Proceedings of the 26th annual international conference on machine learning. ACM; 2009. p. 129–36.

[9] Kusakunniran W, Wu Q, Zhang J, Li H. Gait recognition under various viewing angles based on correlated motionregression. IEEE Trans CircSyst Video Tech. 2012; 22(6):966–80.

[10] Huang C, Ai H, Li Y, Lao S. High-performance rotation invariant multiview face detection. IEEE Trans Pattern Anal Mach Intell. 2007; 29(4):671–86.

[11] Argyriou A, Evgeniou T, Pontil M. Convex multitask feature learning. Mach Learn. 2008; 73(3):243–72.

[12] Huy TN, Shao H, Tong B, Suzuki E. A feature-free and parameter-light multi-task clustering framework. Knowl Inform Syst. 2013; 32(1):251–76.

[13] Fei H, Huan J. Structured feature selection and task relationship inference for multi-task learning. Knowl Inform Syst. 2013; 35(2):345–64.

[14] Ando RK, Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data. J Mach Learn Res. 2005; 6:1817–53.

[15] Gupta SK, Phung D, Adams B, Venkatesh S. Regularized nonnegative shared subspace learning. Data Min KnowlDiscov. 2011; 26(1):57–97.

[16] Ma H, Zhao W, Shi Z. A non-negative matrix factorization framework for semi-supervised document clustering with dual constraints. Knowl Inform Syst. 2013; 36(3):629–51.

[17] Zhang ZY, Li T, Ding C. Non-negative tri-factor tensor decomposition with applications. Knowl Inform Syst. 2013; 34(2):243–65.

[18] Wang YX, Zhang YJ. Nonnegative matrix factorization: A comprehensive review. IEEE Trans Knowl Data Eng. 2013; 25(6):1336–53.