ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 8(S9), 292–299, May 2015* DOI: 10.17485/ijst/2015/v8iS9/65576

Cross Media Data Retrieval based on Semantic Consistency

G. T. Anand[1*], V. Harish Kumar[1], T. T. Manikandan[1], T. Renuga Devi[2] and A. Umamakeswari[2]

[1]Computer Science and Engineering, School of Computing, SASTRA University, Thirumalaisamudhram, Thanjavur 613402, TamilNadu, India; gt.anand1994@gmail.com, abilash545@gmail.com, harish1150@gmail.com

[2]School of Computing, SASTRA University, Thirumalaisamudhram, Thanjavur 613402, TamilNadu, India; renugadevi@cse.sastra.edu, aum@cse.sastra.edu

Abstract

Cross media retrieval plays a vital role in finding semantic consistency among the data that are represented through different media. A framework is proposed for this process that starts with Isomorphic Relevant Redundant Transformation (IRRT), which linearly transforms different heterogeneous low-level feature spaces to a high-level redundant feature-isomorphic space without any dimensionality reduction, i.e., without data loss. Then, transforming these data with same dimensionality, with the help of Convex relaxed Alternating Structure Optimization (cASO). Consequently, the SCP for cross-view data can be obtained. On 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. So the proposed effective framework has its broad applications in the field of the information retrieval system that works on cross view data.

Keywords: Cross-view Data, Cross-Media Retrieval, Semantic Consistency, Semantic Consistent Pattern (SCP), Shared Feature Subspace Learning

1. Introduction

Nowadays, most of the data-processing work is done on cross-view data and on cross-media. Cross-view data maintains semantic consistency with different way of representation. An image coupled with the text representation of an object as shown in Figure 1 is an example of cross view data.

As the first step towards mining SCP, mapping of the heterogeneous low-level feature spaces of the dataset (where no correspondence can be found explicitly) onto the feature-isomorphic space is done. In the feature-isomorphic space the correlated different views of data is to be constructed. Our next step is to build SCP from the feature-isomorphic space by finding a transformation that specifies the mapping of a data-view to another. Thus, creation of this transformation will be helpful in the process of finding SCP.

This paper is organized as: Section-1 provides Introduction and some contributions to this work. Section-2 provides a brief overview of some related works of this paper. Section-3 provides a detailed view of proposed method. Section-4 deals with the working of the proposed method and its performance evaluation. Section-5 summarizes the method and also seeded some future works for this paper.

1.1 Main Contribution

The main contributions for this paper are described as follows:

1. A framework that is used to mine a pattern for cross-view data is a two step process. One is mapping the low-level heterogeneous feature space to common high-level feature space. Second is making use of the linear transformation for classification.

*Author for correspondence

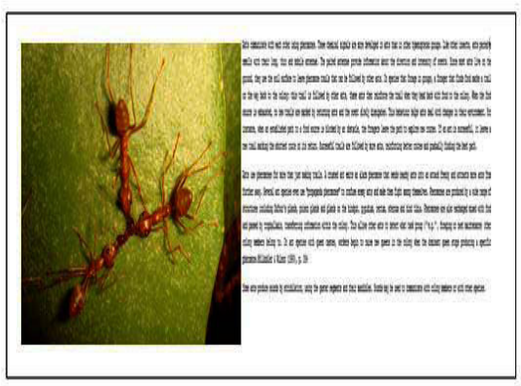

Figure 1. Image and its associated text.

2. Isomorphic Relevant Redundant Transformation (IRRT) is used for linearly mapping multiple heterogeneous feature space to the homogeneous data view.
3. A shared subspace learning algorithm namely, Convex relaxed Alternating Structure Optimization (cASO), to determine the transformation that helps to classify data.

2. Survey on Existing Frameworks

Some of the related works are described as follows.

In order to handle the heterogeneity across different views some classical techniques based on statistical analysis has been proposed, such as Canonical Correlation Analysis (CCA) and Partial Least Squares (PLS) to find the low dimensional information of both views simultaneously, by increasing the correlation and covariance as discussed by Sun et al[5] and Wold[6] respectively.

But, among them CCA is mostly used for cross-media retrieval according to Thompson[7], by extracting features in multi-view problem. CCA can also be useful for classification and clustering of multi-view data in accordance with Hardoon et al.[8] and Chaudhri et al[9]. CCA is equivalent to Linear Discriminant Analysis (LDA) as explained by Sun et al.[5] but without class label.

A common representation for different views has been obtained through CCA. Hardoon et al.[7] had used Kernel CCA to learn a common representation for images and its associated text. Chaudhri et al.[8] had discussed about projecting multiple views of data into a low dimension view among them. Further, CCA usage extended to cross-view classification and retrieval by Sharma et al[9].

However, CCA has inability as it may not extract useful descriptors for classification according to Thompson[7]. But, Kernel CCA alters its inability by transforming the representations 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.[11] where motion is regressed and gait is recognized under multiple view angles, Multi-view Active Appearance Model (MAAM) as discussed by Ramnath et al.[12], Multi-View Face Detector (MVFD), Vector Boosting algorithm (VB) as studied by Hung et al.[13] for multi-view classification.

Since, noise is present in feature-isomorphic space that leads to incorrect dealing of features of data. But, it is proved that features of cross-view data have shared subspace among them which consist of co-occurring features of those data which leads to appreciable noise elimination.

There are various shared subspace learning algorithms to capture the semantic consistency among the representations of data. Some of the shared subspace learning algorithms involve multi-task learning as explained by Ando and Zhang[14], a framework for learning predictive structures from a set of multiple tasks and data that are unlabeled, also in accordance with Argyriou et al.[15], also as per Huy[16] who dealt with multitask clustering, and Fei and Huan dealt with multi-label[17] and multi-class[19] classification, and matrix – factorization[21–24].

In multi-task learning, Ando and Zhang[14] had proposed the Alternating Structure Optimization (ASO) algorithm[13] to learn the predictive structure from multiple tasks. However, it is non-convex so not guaranteed of finding globally optimal solution. ASO also have its improved version, improved ASO (iASO), which is not convex. But, there are also some convex framework called Convex Multi-Task Feature Learning (CMTFL)[14].

When no separate training dataset is used for entire tasks then multi-task and multi-label learning looks equivalent. Tsoumakas et al. proposed least squares loss based shared subspace learning, called Shared-Subspace for Multi-Label Classification (SSMLC)[18], to exploit correlation, is compared with ASO and with Shared Structure in Multi-Class Classification (SSMCC) and proved its superiority.

Besides, there is an unsupervised Joint Shared Non-negative Matrix Factorization (JSNMF)[21], to find shared subspace which is proved to be inferior to ASO.

3. Mining Semantic Consistency

A better framework to mine semantic consistent pattern as follows:

3.1 Overview

A mid-level representation is constructed to build semantic shared subspace from low-level feature space. Many linear transformations are needed to reduce dimensionality conflicts between different views of data. It is also observed here that the redundant feature-isomorphic space also contains correlated representations. Thus, in this space correlation information has been retained.

Further to proceed, the mapping from any representation to another has to be established, since two representations have been mapped to same dimensional space. For visualization refer Figure 2. If the linear transformation has been found, then our semantic pattern mining has been completed successfully.

3.2 Construction of Feature-isomorphic Space

In proposed framework, Isomorphic Relevant Redundant Transformation (IRRT) is used to construct this feature-isomorphic space from low-level feature spaces of different views. Here CCA[5], a classical method is also used, but its dimensionality reduction problem had increased its limitations. This construction of feature-isomorphic space can be generally formulated as:

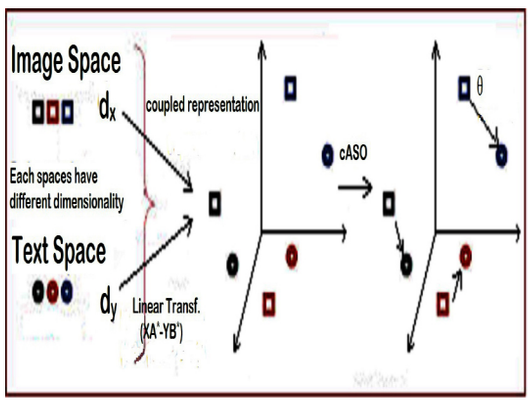

Figure 2. Overview of the framework.

$$\min_{A,B} ||XA - YB||_F^2$$
$$\text{s.t. } A^T X^T XA = I \text{ and } B^T Y^T BY = I \quad (1)$$

where $A \in R^{dxxp}$, $B \in R^{dyxp}$, and $p \in \{1,\ldots,\min(d_x,d_y)\}$. The above equation means that the point $(Y^*B)_{i,p}$ is projected to the point $(X^*A)_{i,p}$ for i=1,…,n. The above projection is clearly explained in Convex optimization[3]. Then, feature-isomorphic space representation for the given dataset is obtained, by using the optimal values of A^* and B^*. Thus, the transformation from the initial dataset representation to feature-isomorphic space representation is expressed as:

$$\mu_X = A^{*T} X \text{ and } \mu_Y = B^{*T} Y \quad (2)$$

Eq.(2) can be integrated to:

$$\mu = \frac{(\mu_X + \mu_Y)}{2} \quad (3)$$

Here p is limited range of values in CCA. Since, p have its maximum value as $\min(d_x,d_y)$ in CCA, the resulting space have its dimension as p. So the dimension of the resulting view lies in the range of $[1, \min(d_x,d_y)]$. This leads to loss of complementary 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 constraints for this mapping, and with no loss in complementary information. This proves IRRT overcomes dimensionality reduction problem. The resulting optimization problem:

$$\min_{A,B} ||XA - YB||_F^2$$
$$\Psi 1: \text{s.t. } ||XA||_* \le \varepsilon \text{ and } ||YB||_* \le \gamma \quad (4)$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$ are positive parameters to control the effect of dimensionality reduction. In "Mining Semantically Consistent Pattern for Cross-view Data" by Zhang, Lei and et al.[1] it is proved that IRRT is simply not an extension of CCA. In IRRT $p \gg \max(d_x,d_y)$ 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 IRRT, the optimization problem has been converted to a convex-relaxed one i.e., iteratively the problem converges to global optimal solution unlike CCA, which is converted to eigenvalue problem, a non-convex method.

In accordance with Lemma 1 of Zhang, Lei and et al.[1], parameters used here satisfies

$$||X||_*||A||_* \le \varepsilon \text{ and } ||Y||_*||B||_* \le \gamma \quad (5)$$

$$||A||_* \le \varepsilon / ||X||_* \text{ and } ||B||_* \le \gamma / ||Y||_* \quad (6)$$

Thus, **Ψ1** can be reformulated as:

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

$$\textbf{Ψ2: s.t. } ||A||_* \le \varepsilon / ||X||_* \text{ and } ||B||_* \le \gamma / ||Y||_* \quad (7)$$

According to Zhang, Lei and et al[1], in the efficient solver of **Ψ2**, the usage of Accelerated Projected Gradient (APG) algorithm for Euclidean projection of any given point is clearly presented in Algorithm 1.

Algorithm 1: Efficient Projection on Trace Norm Constraints (EPTNC)

Input: s, G

Output: z^*

1. By singular value decomposition of s, one can obtain resulting matrices (U_s,E_s,V_s^T).
2. Let p=|H|, where H={i ∈ 1,…,n| σ_i>0}and σ_i is the i^{th} singular value of s.
3. Declare $\theta = (\sum_{i=1}^{p} \sigma_i - m)/p$.
4. Set t_i=max{ $\sigma_i - \theta$,0}
5. Set E_{z^*}=diag(σ_1 ,…,σ_p,0).
6. Compute $z^*=U_s E_{z^*} V_s^T$..

Approximation that leads to better result as follows as in Convex optimization[3]:

$$f_{\gamma,S}(Z)=f(S)+\langle\nabla f(S),Z-S\rangle+(1/2)$$
$$*\langle\nabla f(S)(Z-S),Z-S\rangle+\gamma||Z-S||_F^2/2 \quad (8)$$

which is second order approximation that fasters the convergence to globally optimal solution.

3.3 Semantic Consistent Pattern Mining

Generally, ASO, a semi-supervised learning algorithm, uses a number of Auxiliary Problems (AP) with unlabelled dataset and extracts common structural parameter of APs to improve the performances of Target Problems (TP). As it helps in learning predictors, represented as:

$$f_l(x)= u_l^T x = w_l^T x + v_l^T \theta x$$

where θ is the structure parameter to be determined, which is an *hxd* matrix with ortho-normal rows, i.e., $\theta x \theta^T = I$ and u_l,w_l, and v_l are weight vectors for high-dimensional

Algorithm 2: Isomorphic Relevant Redundant Transformation (IRRT)

Input: $Z_0=[A_{Z0}\ B_{Z0}]$, f(•), γ_1, G, t_0=1, max-iter

Output: Z^*

1. Set $f_{\gamma,S}(Z)=f(S)+\langle\nabla f(S),Z-S\rangle+\gamma||Z-S||_F^2/2$
2. $A_{Z1} = A_{Z0}$ and $B_{Z1} = B_{Z0}$.
3. for i = 1,2,…,max-iter do
4. Set $a_i=(t_{i-1}- 1)/t_{i-1}$.
5. Compute $A_{Si} =(1+a_i) A_{Zi} - a_i A_{Zi-1}$
6. Compute $B_{Si} =(1+a_i) B_{Zi} - a_i B_{Zi-1}$
7. Set $S_i=[A_{Si}\ B_{Si}]$.
8. Compute $\nabla_{AS} f(A_S)$ and $_{BS}f(B_S)$
9. while (true)
10. Compute $A_S'=A_{Si} - \nabla_{AS} f(A_S)/ \gamma_i$.
11. Compute $B_S'=B_{Si} - \nabla_{BS} f(B_S)/ \gamma_i$.
12. Compute $[A_{Zi+1}]$=EPTNC(A_S',G)
13. Compute $[B_{Zi+1}]$=EPTNC(B_S',G)
14. Set $Z_{i+1}=[A_{Zi+1}\ B_{Zi+1}]$
15. if $f(Z_{i+1}) \le f_{\gamma,S}(Z_{i+1})$, then break;
16. else update $\gamma_i = \gamma_i \times 2$.
17. end-if
18. end-while
19. Update $t_i=(1+\sqrt{(1+4t_{i-1}^2)})/2$ and $\gamma_{i+1} = \gamma_i$
20. end-for
21. Set $Z^*=Z_{i+1}$.

of full-featured one and shared low-dimensional space. iASO (improved ASO) can be formulated as:

$$\min_{\{u_\ell,v_\ell\},\theta\theta^T=I} \sum_{\ell=1}^{m} \left(\frac{1}{n_\ell} \sum_{i=1}^{n_\ell} L(u_\ell^T x_i^\ell, y_i^\ell) + g_\ell(u_\ell, v_\ell, \theta)\right) \quad (9)$$

where $g_l(u_l,v_l,\theta)$ is the regularization function, which controls task relatedness, can be expressed as:

$$g_\ell(u_\ell, v_\ell, \theta) = \alpha||u_\ell - \theta^T v_\ell||^2 + \beta||u_\ell||^2 \quad (10)$$

Here, the formulation reduces to the ASO algorithm by setting β = 0 in Eq. (10); and it reduces to m independent Support Vector Machines (SVM) by setting α = 0.

As discussions by Chen, Jianhui and et al.[2] about the non-convexity of the above formulation, now it is able to use the convex converted one by relaxing its feasible

domain into a convex set. And thus, it is changed to convex-relaxed ASO (cASO) formulation as:

$$\min_{U,M,\{t_\ell\}} \sum_{\ell=1}^{m}\left(\frac{1}{n_\ell}\sum_{i=1}^{n_\ell} L(u_\ell^T x_i^\ell, y_i^\ell)\right) + \alpha\eta(1+\eta)\sum_{\ell=1}^{m} t_\ell$$

$$s.t. \begin{pmatrix} \eta I + M & u_\ell \\ u_\ell^T & t_\ell \end{pmatrix} \geq 0, \forall \ell \in N_m,$$

$$tr(M) = h, M \leq I, M \in S_+^d \quad (11)$$

Now, it is implicitly meant that the solution obtained using this formulation in globally optimal solution, given that loss function L is convex. Here M= $\theta^T\theta$. It has two optimization variables U and M and cASO is similar to the block coordinate descent method.

The efficient value of U for the given M can be computed by:

$$\min_U \sum_{\ell=1}^{m}\left(\frac{1}{n_\ell}\sum_{i=1}^{n_l} L(u_\ell^T x_i^\ell, y_i^\ell) + \hat{g}(u_\ell)\right) \quad (12)$$

for the given $\hat{g} = \alpha\eta(1+\eta)tr(u_\ell^T(\eta I + M)^{-1}u_\ell)$. And the optimal M for given U can be found by:

$$\min_M \left(tr(U^T(\eta I + M)^{-1}U)\right)$$

$$s.t.\ tr(M) = h, M \leq I, M \in S_+^d \quad (13)$$

But the optimal approach to find M is discussed by Chen, Jianhui and et al.[2] as follows:

Let U of R^{dxm} having SVD of $P_1EP_2^T$, where P_1 and P_2 are orthogonal and rank(U)=q, for q≤m≤d.

$$E=diag(\sigma_{1,\ldots,}\ \sigma_m) \in R^{dxm}$$

The updating steps of M are listed as follows:

$$\min_{\{\gamma_i\}_{i=1}^q} \sum_{i=1}^{q} \frac{\sigma_i^2}{\eta + \gamma_i}$$

$$s.t. \sum_{i=1}^{q} \gamma_i = h, 0 \leq \gamma_i \leq 1, \forall i \in N_q$$

$$svd\ computation: M = Q\Lambda Q^T$$

$$\Lambda = diag(\lambda_1, \ldots, \lambda_d), where\ Q \in R^{d\times d}$$

$$iteratively:$$

$$\min_{Q,A} tr((\eta I + \Lambda)^{-1}Q^T P_1 E E^T P_1^T Q)$$

$$s.t. QQ^T = Q^T Q = I, \sum_{i=1}^{d} \lambda_i = h, \forall \lambda_i \geq 0$$

This follows convex optimization problem by constructing θ, using first h columns of matrix P_1.

Algorithm 3: Convex-relaxed Alternating Structure Optimization (cASO)

Input: X, Y, α, β, h

Output: U, V and θ

1. Initialize M with respect to constraints of (11).
2. while(true)
3. Compute U using equation (12).
4. Find SVD of $U=P_1EP_2^T$.
5. Update M via above steps.
6. end-while
7. Construct θ using top h eigenvectors of M.
8. Construct V= θ U.
9. Return.

By finding θ, the structural parameter, it is easy to find the semantic consistent pattern by:

$$\tau_{x_i} = \theta^{*T} A^{*T} x_i \ \ and \ \ \tau_{y_i} = \theta^{*T} B^{*T} y_i \quad (12)$$

$$\tau_i = (\tau_{x_i} + \tau_{y_i})/2 \quad (13)$$

Now τ_i represents Semantic Consistent Patterns (SCP) for each data sample. Since τ_{x_i} and τ_{y_i} are in same shared feature space.

As proposed by Khusro et al.[25], semantic web is used to bring the data to open standard format on web and make software applications to retrieve snippets of structured data from HTML pages by analyzing text documents for data which is mixed with the surrounding text. This proposed framework is helpful for those applications.

4. Experimental Result

An experiment is conducted based on the idea presented in this paper. The following section provides a detailed process involved in this experiment.

4.1 Dataset

This experiment is conducted on a publicly available cross-media or cross-view dataset, called Wikipedia featured article collection[4], the statistical information about the dataset is available in Table 1.

Table 1. Feature sets descriptions

Dataset	Feature Set	Total Attributes	Total Labels	Total Samples
Wikipedia Feature Article	Image representation(v_x)	128	10	2866
	Text representation(v_y)	130	10	2866

It consists of 2866 pairs of text and image representations, so called featured article collection. Text representation consist of at least 70 words of text belongs to the corresponding image. Text-view is constructed by TF-IDF encoding method. In original dataset, there are 29 labels but for the experiment case only 10 labels are considered randomly.

4.2 Experimental Setup

As a preprocessing task, all data has to be normalized to unity. The dataset is split into training and test sets with the proportion of 4:1. The experiment has to be repeated five times in order to obtain average performance.

For better analyses, the presented framework has been evaluated by tuning the experiment using 5-fold i.e., classes were folded to 5, based on the AUC (area under the ROC curve) scores. The ROC curve, plotting true positives rate against false positives rate, represents the performance of a binary classifier by varying its discrimination threshold. In this experiment, AUC is calculated from ROC curve. For retrieval, the metric used called Normalized Correlation. As explained in Convex Optimization[3] Euclidean distance problem only treat lengths as free variables that doesn't appear in any part of problem and this distance is invariant not only under orthogonal transformation, but also under translations. So the Euclidean distance metric is not used. The detailed information about the experiment has been tabulated in Table 2 (for IRRT) and Table 3 (for IRRT and cASO).

The better results can be obtained by choosing a good value to the parameters. So, in IRRT, parameters $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$ obtain their best values as any one of the set $\{10^i \mid i = -4,-3,-2,-1, 0, 1, 2, 3, 4\}$, and the dimensionality parameter p can have any one value on the set $\{2^i \times 100 \mid i = 1, 2, 3, 4, 5\}$, and in cASO, regularization parameter η will get its value as 10^{-3}($\alpha = 1$ and β as accordingly).

The parameter sensitivity of IRRT has been visualized in Figure 3. Parameter sensitivity is analyzed so that the best result is obtained at the most sensitive point of the algorithm. For cASO, parameter sensitivity analysis has been done and reported in Figure 4.

Table 2. Map score of IRRT

Attributes of performance evaluation	Image to text retrieval	Text to image retrieval
Map Score	0.1530	0.1357
Map Score per class	0.0559	0.0497
	0.1588	0.1428
	0.2272	0.1982
	0.1334	0.1225
	0.1337	0.1278
	0.1424	0.1187
	0.1003	0.0809
	0.0888	0.1044
	0.1329	0.1087
	0.2083	0.1828
Total classes	10	10
Distance metric	Normalized Correlation	Normalized Correlation
RP	0.1322	0.1310
Precision at k	0.1302	0.1222

Table 3. Map score of IRRT and CASO

Attributes of performance evaluation	Image to text retrieval	Text to image retrieval
Map Score	0.1548	0.1364
Map Score per class	0.1539	0.0531
	0.1656	0.1493
	0.2377	0.1935
	0.1285	0.1236
	0.1367	0.1365
	0.1506	0.1224
	0.2969	0.0805
	0.0888	0.1048
	0.1352	0.1137
	0.1993	0.1733
Total classes	10	10
Distance metric	Normalized Correlation	Normalized Correlation
RP	0.1454	0.1296
Precision at k	0.1313	0.1287

Along with the parameter sensitivity, computational time for IRRT and cASO has also been visualized for different size of dataset and plotted a graph as shown in Figure 5.

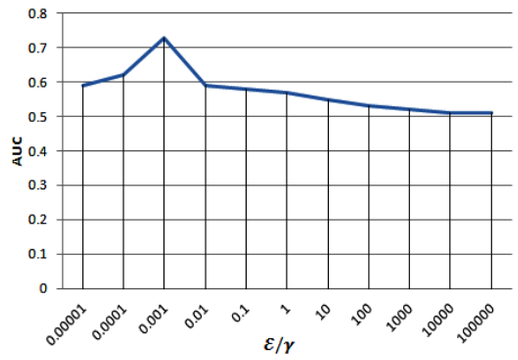

Figure 3. Parameter sensitivity of IRRT.

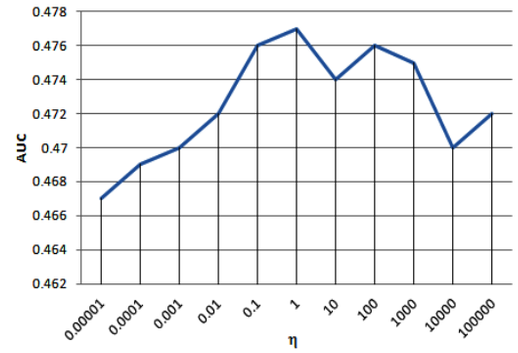

Figure 4. Parameter sensitivity of cASO.

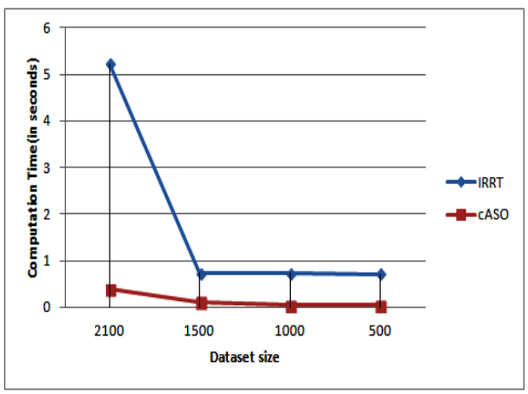

Figure 5. Computation time (in seconds) of IRRT and cASO algorithm.

Besides these analyses, as mentioned[3], a classification figure about the dataset is also attached. A classification problem can be solved, for a classifier, by two ways of discrimination, based on the resulting classifiers. If the classifier has to be linear then, Linear Discrimination Alternative, Robust Linear Discrimination, Support Vector Classifier or Approximate Linear discrimination via Logistic modeling can be used. Else for nonlinear classifier the choices left are Quadratic Discrimination and Polynomial Discrimination.

The classification figure mathematically describes the convex-optimization problem related to the dataset.

The Figure 6 represents the mathematical description about classification, where two set of points are classified by an affine function or linear discrimination function F, whose zero-level set (a line) separates them.

From the above description a fact, two set of points can be linearly discriminated, if only if their convex hulls do not intersect, is revealed.

Figure 7 represents the classification figure of the dataset that was transformed by the proposed framework

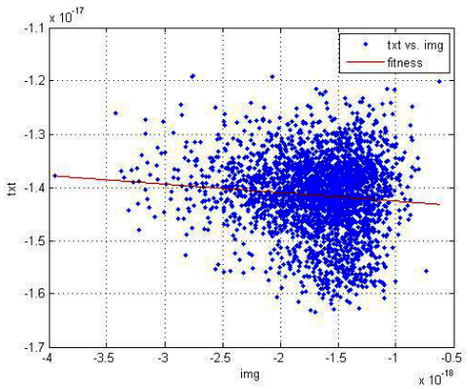

Figure 6. Linear Discrimination problem representation for Wikipedia dataset (result of IRRT).

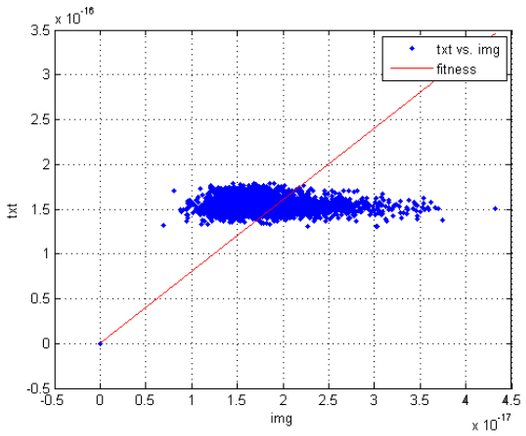

Figure 7. Linear Discrimination problem representation for Wikipedia dataset (result of IRRT and CASO).

where one can understand that the transformation will make a view of dataset to come closer to other, i.e., this is the resulting feature space of this experiment on the above mentioned dataset.

Since alternative frameworks are difficult to implement, however it also gives a good result. But cASO achieves a better global optimum result without any complex calculations.

5. Conclusion

In this framework, IRRT and cASO had played a major role to map low-level heterogeneous 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 convex relaxation is used to project feasible domain set onto the convex set in cASO. Advancements in Convex optimization and approximations will be helpful to enhance the proposed framework in order to gain better results.

6. References

1. Zhang L, Zhao Y, Zhu Z, Wei S, Wu X. Mining semantically consistent patterns for cross-view data. IEEE Transactions on Knowledge and Data Engineering. 2014; 26(11):2745–58.
2. Chen J, Tang L, Liu J, Ye J. A convex formulation for learning shared structures from multiple tasks. Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM; 2009. p. 137–44.
3. Boyd S, Vandenberghe L. Convex Optimization. New York: Cambridge University Press; 2009.
4. Wikipedia Featured articles. Available from: http://en.wikipedia.org/wiki/Wikipedia:Featured_articles
5. Sun L, Ji S, Ye J. Canonical correlation analysis for mult-ilabel classification: A least-squares formulation, extensions, and analysis. IEEE Trans Pattern Anal Mach Intell. 2011; 33(1):194–200.
6. Wold H. Partial least squares. Encyclopedia of statistical sciences: Wiley Online Library; 2006.
7. Thompson B. Canonical correlation analysis. Encyclopedia of statistics in behavioral science: Wiley Online Library; 2005.
8. Hardoon D, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. Neural Comput. 2004; 16(12):2639–64.
9. 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.
10. Sharma A, Kumar A, Daume H, Jacobs DW. Generalized mul-tiview analysis: A discriminative latent space. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2012. p. 2160–7.
11. Kusakunniran W, Wu Q, Zhang J, Li H. Gait recognition under various viewing angles based on correlated motion regression. IEEE Trans Circ Syst Video Tech. 2012; 22(6):966–80.
12. Ramnath K, Koterba S, Baker S, Matthews I, Hu C, Xiao J, Cohn J, Kanade T. Multi-view AAM fitting and construction. Int J Comput Vis. 2008 Feb; 76(2):183–204.
13. 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.
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. Argyriou A, Evgeniou T, Pontil M. Convex multitask feature learning. Mach Learn. 2008; 73(3):243–72.
16. 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.
17. Fei H, Huan J. Structured feature selection and task relationship inference for multi-task learning. Knowl Inform Syst. 2013; 35(2):345–64.
18. Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data. Data mining and Knowledge Discovery handbook. US: Springer; 2010. p. 667–85.
19. Kong X, Ng M, Zhou Z. Transductive multi-label learning via label set propagation. IEEE Trans Knowl Data Eng. 2013; 25(3):704–19.
20. Candes EJ, Recht B. Exact matrix completion via convex optimization. Found Comput Math. 2009; 9(6):717–72.
21. Gupta SK, Phung D, Adams B, Venkatesh S. Regularized nonnegative shared subspace learning. Data Min Knowl Discov. 2011; 26(1):57–97.
22. 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.
23. Zhang ZY, Li T, Ding C. Non-negative tri-factor tensor decomposition with applications. Knowl Inform Syst. 2013; 34(2):243–65.
24. Wang YX, Zhang YJ. Nonnegative matrix factorization: A comprehensive review. IEEE Trans Knowl Data Eng. 2013; 25(6):1336–53.
25. Khusro S, Jabeen F, Mashwani SR, Alam I. Linked open data: towards the realization of semantic web - a review. Indian Journal of Science and Technology. 2014; 7(6):745–64.