Muhammad Saifullah, bscs2019-16

Muhammad Shehzad Ali Khan, bscs2019-68

Muhammad Zohaiab Aslam Khan, bscs2019-57

Fuzziness based semi-supervised multimodal learning for patient’s activity recognition using RGBDT videos

Muhammed J.A. Patwary, Weipeng Cao, Xi-Zhao Wang, Mohammad Ahsanul Haque

Table of Content::

ALL images are taken from same article under discussion,

Acknowledgement:

We would like to express gratitude to Dr. Malik Jahan, for having us with this wonderful project of FSSL-PAR. We Muhmmad Zohaib Aslam Khan, Saifullah Khan, Muhammad Shehzad Ali Khan as group contribute our best in this work.

Abstract:

Today Patient activity recognition is a very hot topic, many models are being worked out for it. Due to Patient's condition, it is not possible to get a good database, so Haque et al. from the Hammel Neurorehabilitation and Research Center (HNRC), Denmark arrange a database. Volunteers act to make the database, for the database two variances of the dataset are considered, first one volunteers are covered with a blanket and vice versa. When a large amount of summarized data(labeled) and domain knowledge is available, supervised learning performs very well, but in this problem we have both these challenges. So we use Fuzziness semi-supervised learning for patient activity recognition(FSSL-PAR).

Introduction:

Wang at el in 2015, proposed the relation between Fuzziness and classification, the classifier with high fuzziness in the output vector has more probability of misclassification and lower fuzziness indicates less probability of misclassification. So We use SSL with fuzziness to get significant supervised data(labeled) to make further progress in unsupervised learning. There are three main goals of this work:

1. Proposed model to investigate PA.
2. Unlike traditional approaches, use multimodal data instead of unimodal data.
3. Find relation between input labeled data and efficiency of algorithm

The detailed technique of body motion analysis:

Body motion tracking:

After removing noise, for categorization of 8 PA categories, they detect both local and global motion with intensity , the reason behind it was movement of different organs can be thought as the same like shifting of arm and twisting wrist. The technique of activity recognition consists of two techniche GFT and LKT. GFT is used to identify the pixels and LKT is used to track those pixels in frame.

Let two consecutive frames Kand L, K(X)=(x,y) and L(x,y) represent magnitude of intensities at (x,y) coordinates of the frame. Pixels at (x,y) can be represented as [Px,Py]t and features can be tracked by tracking this in successive frames through the window of size **Wx x Wy** from K to L frame.

ALL images are taken from same article under discussion,

$$f_{GFT}(\Theta) = \sum_{x=p_x}^{p_x+\omega_x} \sum_{x=p_y}^{p_y+\omega_y} (I(x) - J(x+\Theta))^2$$

Here { I(x)-j(x+Θ) } represent the difference between K and L frames of **Wx x Wy** sized window. First they applied GFT to full **ROI** but there was a problem that it loses pixels with low textures,so they divide the **ROI** into 20 grids and apply GFT on each and then KLT. A motion map of higher intensity is produced in **ROI.** It contains 3 aspects, 1.vedio frames, 2.traced feature points and 3. Motion's intensity of the second feature over the first feature.

Different feature extraction techniques for classification of activity:

Different techniques are used to attain features.

1. Principal component analysis (PCA)
2. First Difference of the motion map (FDiff)
3. Primitive Radon Features (PRadon)
4. Radon distance Features (DiffRaon)

Rationality of FSSL:

In this practice data distribution is unknown. So given data set D={(x1,y1)....(xn,yn)} $R^d x R^m$ where N is the number of training samples, d are dimensions and m are classes. The objective of the ML algorithm is to minimize error.

$$\mathbf{f}_N = arg\ min \frac{1}{N} \sum_{i=1}^{N} \ell(y_i, f(x_i))$$

Where l is an error function (real -predicted).According to VM dimensions theory as N approaches to affinity, f_N can be reached to a good hypothesis.

$$arg\ min R_n(f) = \mathbf{f}_N \rightarrow \mathbf{f}^* = inf\ R(f) \quad s.t.\ N \rightarrow \infty.$$

This inspired us to use more data, to increase the probability of high performance of algorithms.

Methodology of FSSL:

The given data is randomly divided into training set X^{tr} and test set X^{te}. X^{tr} is subdivided into labeled data X^{tr} (L) and unlabeled data X^{tr}(U). Xtr(L) is used in initial classifiers like **KNN** and **ELM**. Then X^{tr}(U) is used in the initial classifier to get fuzzy vectors, on the basis of magnitude of fuzziness vectors are divided into three vectors with Low, High and Medium fuzziness. Only vectors with low fuzziness are added in the original training set. New labeled data is used again with initial labeled data. This iterative work goes on until good efficiency is reached.

ALL images are taken from same article under discussion,

Non-iterative neural network with single hidden layer(ELM):

There are 3 essential parts in this algorithm. 1.input layer, 2.outputlyer and 3. Hidden layer. Firstly we select random wi and bias where i = 1,2,3,...P used between input and hidden layer, secondly on the basis of wi and bi , hidden layer output matrix is calculated. Final weights are obtained by this equ $\mathbf{B = (H^T H)^{-1} H^T T}$. In this **experiment 1** we randomly assigned values to Q,n and m where Q is no of nodes of hidden layer, n and m are two integers where n<m. We change the values of Q to check the change in node change efficiency of the algorithm.

Fuzzy KNN algorithm:

In the **second experiment** we use fuzzy KNN, we change the value of k to check on which value of k our algorithm is performing well.

Without the fusion of the features.

Features	Supervised learning algorithms						SSL(ELM)					SSL(KNN)				
	KNN	DT	NB	SVM	LDA	GLM	5%	10%	30%	50%	80%	5%	10%	30%	50%	80%
R-F1	35.82	42.47	**49.68**	49.29	30.14	45.96	45.19	46.24	46.08	47.16	**50.48**	46.54	48.50	48.89	50.16	**53.03**
R-F2	55.51	43.94	59.51	**60.00**	56.82	57.47	56.04	57.96	56.40	60.32	**60.45**	58.39	60.61	59.42	63.93	**65.77**
R-F3	68.25	61.45	12.50	**72.66**	66.74	71.57	69.42	69.40	71.43	72.41	**74.60**	70.87	72.69	75.11	76.32	**79.73**
R-F4	65.76	63.72	61.90	58.63	63.09	**68.02**	66.42	67.36	68.46	69.43	**70.44**	68.94	70.29	71.85	72.02	**74.60**
T-F1	36.21	32.65	**47.49**	44.83	33.38	45.56	45.40	45.64	46.46	**47.43**	**47.43**	44.75	44.92	48.99	49.90	**50.98**
T-F2	**56.57**	33.29	12.50	46.49	46.40	48.73	53.35	53.32	5H.39	57.35	**58.43**	56.92	55.99	57.91	60.92	**61.75**
T-F3	58.52	63.34	12.50	63.18	41.37	**64.10**	62.41	62.55	64.38	64.41	**67.39**	65.13	65.28	68.00	67.83	**70.16**
T-F4	63.11	58.13	49.43	32.72	59.05	**64.95**	62.47	63.40	63.46	66.47	**67.46**	65.89	66.97	66.13	69.30	**70.13**
D-F1	18.22	24.51	**29.19**	28.54	21.68	26.70	40.46	41.43	40.44	43.44	**44.38**	41.78	43.59	42.77	45.09	**46.60**
D-F2	28.26	22.97	32.20	**37.83**	32.21	34.31	38.35	38.36	39.41	41.42	**42.41**	39.46	40.28	41.38	43.50	**45.53**
D-F3	38.38	42.32	12.50	46.19	**50.81**	46.67	45.35	46.36	48.37	53.38	**54.41**	43.62	44.68	50.38	56.05	**58.83**
D-F4	41.37	37.65	33.67	**47.12**	44.26	45.24	47.42	47.40	48.40	51.47	**51.49**	48.79	49.77	50.62	54.64	**56.76**

Note: Here, R = RGB, T = Thermal and D = Depth.
F1 = FDiff, F2 = PCA, F3 = PRadon and F4 = DiffRadon.
For example, R-F1 means RGB-FDiff.

Table 2
After the fusion of the features.

Features	Supervised learning algorithms						SSL(ELM)					SSL(KNN)				
	KNN	DT	NB	SVM	LDA	GLM	5%	10%	30%	50%	80%	5%	10%	30%	50%	80%
R-D-F1	30.73	**53.13**	49.3	50.02	34.28	47.72	52.75	53.92	54.99	55.90	**58.98**	53.29	57.85	56.86	61.57	**62.60**
R-D-F2	52.33	48.28	**63.01**	51.83	48.99	53.77	63.92	63.99	64.31	65.92	**66.38**	65.66	66.72	66.66	67.35	**68.37**
R-D-F3	59.2	59.34	12.5	**72.29**	70.3	70.26	74.33	74.28	77.68	76.28	**80.63**	77.36	77.58	78.27	80.93	**81.07**
R-D-F4	60.4	60.9	61.39	55.84	65.89	**70.20**	71.39	72.97	75.11	76.30	**78.31**	76.39	78.29	79.03	80.78	**82.49**
R-T-F1	39.7	44.22	**54.97**	49.29	40.32	50.1	54.95	55.90	55.99	56.90	**58.20**	56.69	57.40	58.63	61.74	**63.87**
R-T-F2	61.19	45.48	12.5	60.82	57.48	**62.62**	62.92	61.99	64.31	63.92	**65.25**	56.01	60.11	61.83	62.11	**64.75**
R-T-F3	69.38	60.39	12.5	74.93	71.24	**77.68**	75.61	78.68	68.00	78.93	**80.26**	70.58	75.93	78.21	80.46	**84.53**
R-T-F4	61.81	67.89	66.04	55.74	68.72	**72.41**	70.19	71.97	72.13	73.30	**75.13**	71.81	71.25	74.95	76.44	**78.68**
D-T-F1	27.81	37.33	42.81	**48.08**	31.17	42.62	48.75	49.92	49.99	49.90	**51.98**	51.10	53.62	55.56	**57.27**	56.85
D-T-F2	42.16	35.19	12.5	46.17	40.61	**47.03**	56.21	55.10	57.10	60.39	**61.18**	54.90	54.55	58.83	62.29	**64.14**
D-T-F3	50.5	54.17	12.5	59.4	51.95	**61.12**	65.43	65.28	67.00	67.83	**69.16**	63.11	64.98	68.22	69.50	**72.31**
D-T-F4	53.91	52.91	49.18	40.98	59.25	**63.96**	66.09	67.17	66.53	69.50	**70.73**	63.41	64.12	64.38	66.32	**68.47**

Note: Here, R = RGB, T = Thermal and D = Depth.
F1 = FDiff, F2 = PCA, F3 = PRadon and F4 = DiffRadon.
For example, R-D-F1 means RGB-Depth-FDiff.

Conclusions:

- ☐ FSSL strategy is performing very prominent Over supervised learning strategies so it is a worthy strategy to work on this problem.
- ☐ Fusion of Multimodal features are very efficient in classification, illustrated from above table.
- ☐ From the above table it is very clear that labeled data is in positive direct relation with accuracy of algorithm.

FSSL is very handi in problems like this when Supervised Learning algo can't be used because of low annotation data. In future this can be compared with other SSL arts and can be used in other deep learning domains.

ALL images are taken from same article under discussion,