Heterogeneous data sources

1. 1Introduction

This is a very important research study which is meant to evaluate the effects of learning from heterogeneous data. It is relevant to study this topic because there has been a problem of learning representation of human face (Basso et al, 2003). To achieve this we will have to use neuron network which will help us collect relevant dataset, convert it into homogenous format, use machine learning and finally ensure that new data is developed into representation when it has the same structure as the original training data (Blanz and Vetter, 2000). For us to achieve greater result, we will ensure that we adopt active appearance model where data should remain in 2D images of face. They are them marked and then learning of representation is done using PCA to ensure that the shapes and appearance information are remaining normal. We will have to follow the above process in order to produce new 2D face images. The importance of using this approach is because it is able to produce high quality data and it also has multiple inputs which can be observed at different resolutions.

1. 2 Data Collection

It is important to collect data from different sources for this research (Schroff, Florian, et al, 2011). This will ensure that we have comprehensive data that with various features to meet the requirement of different purposes. We will use users profile in the establishment of recommendation systems or we will also use a model that can support the application of chronological behavior of users and social networks to deduce their interest on similar products (Abney, 2007). We will ensure that we use many assorted

data sources to help us in building strong learning models to use in this study (de Almeida Freitas, 2014). In this case this will be named as framework assorted learning and we proposed to collect data such as no overlapping features, occasions and different networks.

1. 3 Transformation of data into homogeneous Format

To transform data into homogeneous format, we will ensure that we use an overall optimization structure and also formulate an equivalent learning model that comes from gradient boosting (Wallace, Roy, et al, 2011). This will be done to reduce empirical losses which usually occur during data transformation into homogeneous format (Gao *et al,* 2009). We will achieve this by the introduction of two different constraints which includes the fact that there must be a consensus in the proposed overlapping instances and the prediction in connected instances must be graphed effectively (Turney, 2000). We will use stochastic gradient to find the solution for objective function and we will also ensure that we design weighting strategy to put more focus on useful data sources but ignore others (Eldardiry and Neville, 2011). We will initially ensure that suggested strategy is able to produce more accurate data. The use of this approach will enable us to be more successful that when ordinary concatenation information sources are used as commonly used in movies ratings.

1. 4 Application of Machine Learning

In this research, we also plan to use 3D morphable model to support our work in modeling and animation. We will also ensure that we only use standardized 3D morphable model and interactive learning algorithm. We

will achieve greater result when we add arithmetic model a noise (Blanz and Vetter, 2002). The importance of the use of these models is that we will be able to make good use of missing data as it is a common phenomenon when data is collected through 3D generation system (Wissner-Gross, 2016). We will be forced to use 3D morphable because current models are more complex but are more efficient. At the end of our analysis, we will ensure that comparison analysis is done between 3D morphable model and traditional model (Basso *et al* , 2003). This will ensure that we used more reliable sources that ensure that the best quality results are achieved.

1. 5 Significance of this Research study

This research is very important in the identification of how heterogeneous data affect machine learning. It also promotes adaptive learning where the learners acquire the experience on how to execute a given task based on the data given for training. Finally this research is significant because it enables the learner to create self –organization.

Bibliography

Abney, Steven. Semisupervised learning for computational linguistics. CRC Press, 2007.

Basso, T. et al. *Reanimating faces in images and video* . In Proc. of Eurographics 2003, 2003.

Blanz, V and Vetter, T. Reconstructing the complete 3d shape of faces from partial information. it - *Information Technology,* 44(6): 295–302, January 2002.

de Almeida Freitas, Fernando, et al. " Grammatical Facial Expressions Recognition with Machine Learning." FLAIRS Conference. 2014.

Eldardiry H and Neville J, " Across-model collective ensemble classification," in AAAI, 2011

Gao, W. et al " *Heterogeneous source consensus learning via decision propagation and negotiation,* " in KDD, pp. 339–348, 2009.

Lingala, Mounika, et al. " Fuzzy logic color detection: Blue areas in melanoma dermoscopy images." Computerized Medical Imaging and Graphics 38. 5 (2014): 403-410.

Maes, Chris, et al. " Feature detection on 3D face surfaces for pose normalisation and recognition." Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on. IEEE, 2010.

Schroff, Florian, et al. " Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

Turney, Peter. " Types of cost in inductive concept learning." (2000).

Wallace, Roy, et al. " Inter-session variability modelling and joint factor analysis for face authentication." Biometrics (IJCB), 2011 International Joint Conference on. IEEE, 2011.

Wissner-Gross, A. Edge. com. Retrieved 8 January2016.

Zliobaite, Indre, et al. " Active learning with evolving streaming data."

Machine Learning and Knowledge Discovery in Databases. Springer Berlin

Heidelberg, 2011. 597-612.