International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 11 | Nov - 2021 ISSN: 2582-3930

A Survey on Features Extraction Using Dimensionality Reduction Techniques

Dr. Vinod Sharma

Department of Computer science & Engineering , SCE, Bhopal, India

----------***----------

Abstract:-Many of today's leading companies, such as Google, Facebook and Uber, Zomato Ola make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies. Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. In each part of a second the world generates an unprecedented volume of data. As data has become a crucial component of businesses and organizations across all industries, it is essential to process, analyze, and visualize it appropriately to extract meaningful insights from large datasets. The more data we produce every second, the more challenging it is to analyze and visualize it to draw valid inferences.

Introduction:-Data representation constantly increasing and expanding exponentially, and it is important for us to find useful information from this massive data. The overall process of analyzing data to find understandable and useful information is called data mining. Statistical and machine reasoning methods face a formidable problem when dealing with such high-dimensional data, and normally the number of input variables is reduced before a data mining algorithm can be successfully applied. The dimensionality reduction can be made in two different ways: by only keeping the most relevant variables from the original dataset (this technique is called feature selection) or by exploiting the redundancy of the input data and by finding a smaller set of new variables, each being a combination of the input variables, containing basically the same information as the input variables (this technique is called dimensionality reduction).
This is where Dimensionality Reduction comes into picture. Much of the data is highly redundant and can be efficiently brought down to a much smaller number of variables without a significant loss of information . The mathematical

Procedures making possible this reduction are called dimensionality reduction techniques; they have widely been developed by fields like Statistics or Machine Learning, and are currently a hot research topic. In this Paper we eliminates noise and redundant features**.**

Keywords: Dimensionality reduction, component , features, variance, supervised, unsupervised

Objectives:- Space required to store the data is reduced as the number of dimensions comes down. Less dimensions lead to less computation/training time.It reduces the time and storage space required. It helps Remove multi-collinearity which improves the interpretation of the parameters of the machine learning model. It becomes easier to visualize the data when reduced to very low dimensions such as 2D or 3D. Some algorithms do not perform well when we have a large dimensions. Feature reduction leads to the need for fewer resources to complete computations or tasks. Less computation time and less storage capacity needed means the computer can do more work. During machine learning, feature reduction removes multicollinearity resulting in improvement of the machine learning model in use.According to wikipedia, “feature selection is the process of selecting a subset of relevant features for use in model construction” or in other words, the selection of the most important features.In normal circumstances, domain knowledge plays an important role and we could select features we feel would be the most important.

Research Methodology:-

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches:supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. The type of

algorithm data scientists choose to use depends on what type of data they want to predict.

Supervised learning: In this type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.

Unsupervised learning: This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.

Semi-supervised learning: This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.

Reinforcement learning: Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

.In reference to the ML following approach are categorized:

1.Dimensionality Reduction:

In simple words, **dimensionality reduction** refers to the technique of reducing the dimension of a data feature set. Usually, machine learning datasets (feature set) contain hundreds of columns (i.e., features) or an array of points, creating a massive sphere in a three-dimensional space. By applying **dimensionality reduction**, you can decrease or bring down the number of columns to quantifiable counts, thereby transforming the three-dimensional sphere into a two-dimensional object (circle).

The higher is the number of features or factors (a.k.a. variables) in a feature set, the more difficult it becomes to visualize the training set and work on it. Another vital point to consider is that most of the variables are often correlated. So, if you think every variable within the feature set, you will include many redundant factors in the training set.

Furthermore, the more variables you have at hand, the higher will be the number of samples to represent all the possible combinations of feature values in the example. When the number of variables increases, the model will become more complex, thereby increasing the likelihood of overfitting. When you train an ML model on a large dataset containing many features, it is bound to be dependent on the training data. This will result in an overfitted model that fails to perform well on real data.

The primary aim of dimensionality reduction is to avoid overfitting. A training data with considerably lesser features will ensure that your model remains simple – it will make smaller assumptions.

Dimensionality Reduction Techniques

Dimensionality reduction refers to techniques for reducing the number of input variables in training data. When dealing with high dimensional data, it is often useful to reduce the dimensionality by projecting the data to a lower dimensional subspace which captures the “essence” of the data.

Dimensionality reduction techniques can be categorized into two broad categories:

1. Feature selection
2. Feature Extraction

1. Feature selection

The feature selection method aims to find a subset of the input variables (that are most relevant) from the original dataset. Feature selection includes three strategies, namely:

1. Filter strategy
2. Wrapper strategy
3. Embedded strategy

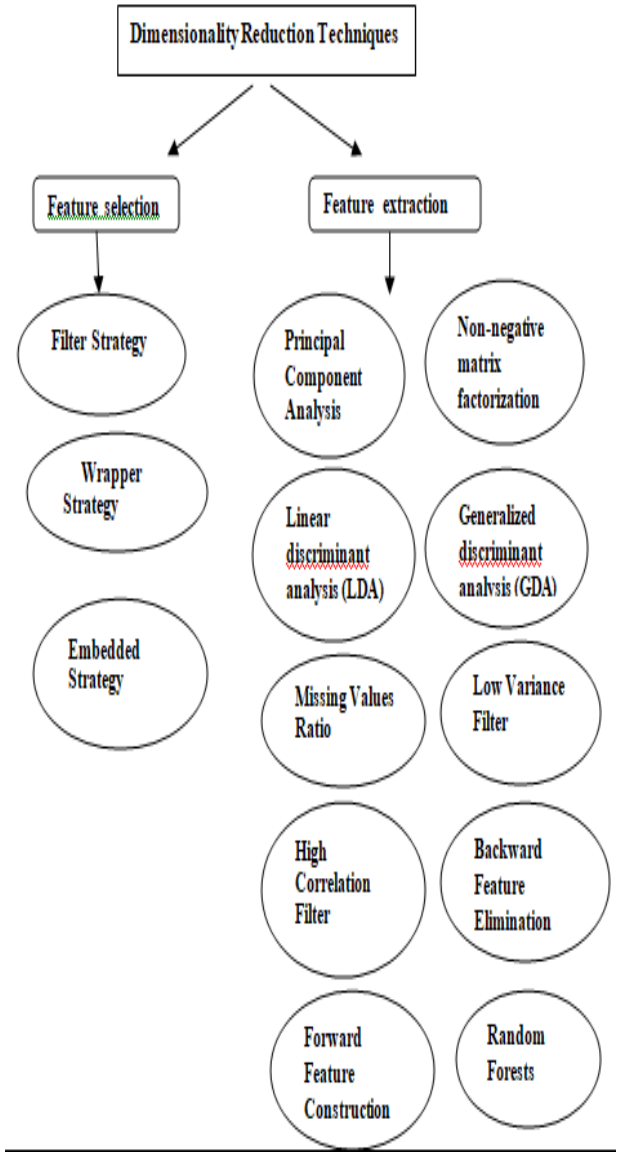

Figure 1:- Dimensionality Reduction techniques

2. Feature extraction

Feature extraction, a.k.a, feature projection, converts the data from the high-dimensional space to one with lesser dimensions. This data transformation may either be linear or it may be nonlinear as well. This technique finds a smaller set of new variables, each of which is a combination of input variables (containing the same information as the input variables).

Without further ado, let's dive into a detailed discussion of a few commonly used dimensionality reduction techniques'

2.1 Principal Component Analysis (PCA)

Principal Component Analysis is one of the leading linear techniques of dimensionality reduction. This method performs a direct mapping of the data to a lesser dimensional space in a way that maximizes the variance of the data in the low-dimensional representation.

Essentially, it is a statistical procedure that orthogonally converts the '*n*' coordinates of a dataset into a new set of *n* coordinates, known as the principal components. This conversion results in the creation of the first principal component having the maximum variance. Each succeeding principal component bears the highest possible variance, under the condition that it is orthogonal (not correlated) to the preceding components.

The PCA conversion is sensitive to the relative scaling of the original variables. Thus, the data column ranges must first be normalized before implementing the PCA method. Another thing to remember is that using the PCA approach will make your dataset lose its interpretability. So, if interpretability is crucial to your analysis, PCA is not the right dimensionality reduction method for your project.

2.2. Non-negative matrix factorization (NMF)

NMF breaks down a non-negative matrix into the product of two non-negative ones. This is what makes the NMF method a valuable tool in areas that are primarily concerned with non-negative signals (for instance, astronomy). The multiplicative update rule by Lee & Seung improved the NMF technique by – including uncertainties, considering missing data and parallel computation, and sequential construction.

These inclusions contributed to making the NMF approach stable and linear. Unlike PCA, NMF does not eliminate the mean of the matrices, thereby creating unphysical non-negative fluxes. Thus, NMF can preserve more information than the PCA method.

Sequential NMF is characterized by a stable component base during construction and a linear modeling process. This makes it the perfect tool in astronomy. Sequential NMF can preserve the flux in the direct imaging of circumstellar structures in astronomy, such as detecting exoplanets and direct imaging of circumstellar disks.

2.3. Linear discriminant analysis (LDA)

The linear discriminant analysis is a generalization of Fisher's linear discriminant method that is widely applied in statistics, pattern recognition, and machine learning. The LDA technique aims to find a linear combination of features that

can characterize or differentiate between two or more classes of objects. LDA represents data in a way that maximizes class separability. While objects belonging to the same class are juxtaposed via projection, objects from different classes are arranged far apart. LDA is also closely related to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data.Geometric anomalities in high dimension lead to the well-known curse of dimensionality. Nevertheless, proper utilization of concentration of measure phenomena can make computation easier.

2.4. Generalized discriminant analysis (GDA)

The generalized discriminant analysis is a nonlinear discriminant analysis that leverages the kernel function operator. Its underlying theory matches very closely to that of support vector machines (SVM), such that the GDA technique helps to map the input vectors into high-dimensional feature space. Just like the LDA approach, GDA also seeks to find a projection for variables in a lower-dimensional space by maximizing the ratio of between-class scatters to within-class scatter.

2.5. Missing Values Ratio

When you explore a given dataset, you might find that there are some missing values in the dataset. The first step in dealing with missing values is to identify the reason behind them. Accordingly, you can then impute the missing values or drop them altogether by using the befitting methods. This approach is perfect for situations when there are a few missing values.

However, what to do when there are too many missing values, say, over 50%? In such situations, you can set a threshold value and use the missing values ratio method. The higher the threshold value, the more aggressive will be the dimensionality reduction. If the percentage of missing values in a variable exceeds the threshold, you can drop the variable.

Generally, data columns having numerous missing values hardly contain useful information. So, you can remove all the data columns having missing values higher than the set threshold.

2.6. Low Variance Filter

Just as you use the missing values ratio method for missing variables, so for constant variables, there's the low variance filter technique. When a dataset has constant variables, it is not possible to improve the model's performance. Why? Because it has zero variance.

In this method also, you can set a threshold value to wean out all the constant variables. So, all the data columns with variance lower than the threshold value will be eliminated. However, one thing you must remember about the low variance filter method is that variance is range dependent. Thus, normalization is a must before implementing this dimensionality reduction technique.

2.7. High Correlation Filter

If a dataset consists of data columns having a lot of similar patterns/trends, these data columns are highly likely to contain identical information. Also, dimensions that depict a higher correlation can adversely impact the model's performance. In such an instance, one of those variables is enough to feed the ML model.

For such situations, it's best to use the Pearson correlation matrix to identify the variables showing a high correlation. Once they are identified, you can select one of them using VIF (Variance Inflation Factor). You can remove all the variables having a higher value (VIF > 5). In this approach, you have to calculate the correlation coefficient between numerical columns (Pearson's Product Moment Coefficient) and between nominal columns (Pearson's chi-square value). Here, all the pairs of columns having a correlation coefficient higher than the set threshold will be reduced to 1.

Since correlation is scale-sensitive, you must perform column normalization.

2.8. Backward Feature Elimination

In the backward feature elimination technique, you have to begin with all 'n' dimensions. Thus, at a given iteration, you can train a specific classification algorithm is trained on *n* input features. Now, you have to remove one input feature at a time and train the same model on *n-1* input variables *n* times. Then you remove the input variable whose elimination generates the smallest increase in the error rate, which leaves behind *n-1* input features. Further, you repeat the classification using *n-2* features, and this continues till no other variable can be removed.

Each iteration *(k)* creates a model trained on n-k features having an error rate of *e(k)*. Following this, you must select the maximum bearable error rate to define the smallest number of features needed to reach that classification performance with the given ML algorithm.

2.9. Forward Feature Construction

The forward feature construction is the opposite of the backward feature elimination method. In the forward feature construction method, you begin with one feature and continue to progress by adding one feature at a time (this is the variable that results in the greatest boost in performance).

Both forward feature construction and backward feature elimination are time and computation-intensive. These methods are best suited for datasets that already have a low number of input columns.

2.10. Random Forests

Random forests are not only excellent classifiers but are also extremely useful for feature selection. In this dimensionality reduction approach, you have to carefully construct an extensive network of trees against a target attribute. For instance, you can create a large set (say, 2000) of shallow trees (say, having two levels), where each tree is trained on a minor fraction (3) of the total number of attributes.

The aim is to use each attribute's usage statistics to identify the most informative subset of features. If an attribute is found to be the best split, it usually contains an informative feature that is worthy of consideration. When you calculate the score of an attribute's usage statistics in the random forest in relation to other attributes, it gives you the most predictive attributes.

Conclusion and related work:-

Feature selection and extraction is an important part of any machine learning process. Here we study on several methods for feature selection and dimensionality reduction that can aid in improving model performance.With the implementation of above techniques there are many ways and researches are generated in research field. Techniques are beneficial in genome sequencing, extraction of irrelevant information,increase performance of relevant models etc..To conclude, when it comes to dimensionality reduction, no technique is the absolute best. Each has its quirks and advantages. Thus, the best way to implement dimensionality reduction techniques is to use systematic and controlled experiments to figure out which technique(s) works with your model and which delivers the best performance on a given dataset. Techniques are beneficial in terms of minimizing an error rate to optimize better result at high level performance.

Web References:-

[1]https://towardsdatascience.com/feature-selection-and-dimensionality-reduction-f488d1a035de

[2] https://searchenterpriseai.techtarget.com/ feature/

[3] https://www.upgrad.com/blog/top-dimensionality-reduction-techniques-for-machine-learning/

[4]"Pavan Vadapalli" Director of Engineering @ upGrad. Motivated to leverage technology to solve problems. Seasoned leader for startups and fast moving orgs. Working on solving problems of scale and long term technology.

References:-

[1] Ansam A. AbdulHussien "Comparison of Machine Learning Algorithms to Classify Web Pages" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 11, 2017.

[2] [Abdi2003] Abdi, H. Lewis-Beck, M.; Bryman, A. & Futing, T. (ed.) Encyclopedia for research methods for the social sciences Factor rotations in factor analyses Sage, 2003, 792-795

[3]. [Abdi2007] Abdi, H. Salkind, N. (ed.) Encyclopedia of measurements and statistics Singular value decomposition (SVD) and Generalized Singular Value Decomposition (GSVD) Sage Publications, 2007, 907-912.

[4] [Baccini1996] Baccini, A.; Besse, P. & Falguerolles, A L1-norm PCA and a heuristic approach. A. Ordinal and Symbolic Data Analysis. Diday, E.; Lechevalier, Y. & Opitz, O. (Eds.) Springer, 1996, 359-368.

[5] C.O.S. Sorzano1, † , J. Vargas1 , A. Pascual-Montano "A survey of dimensionality reduction techniques". 1 Natl. Centre for Biotechnology (CSIC) C/Darwin, 3. Campus Univ. Autónoma, 28049 Cantoblanco, Madrid, Spain {coss,jvargas,pascual}@cnb.csic.es Corresponding author.

[6] F.Shen, X.Luo and Yi.Chen, "Text Classification Dimension Reduction Algorithm for Chinese Web Page Based on Deep Learning", International Conference on Cyberspace Technology (CCT 2013), pp. 451 – 456, Beijing, China, 23 Nov. 2013.

[7] I. Anagnostopoulos, C. Anagnostopoulos, V. Loumos and E. Kayafas, "Classifying Web pages employing a probabilistic neural network", IEE Proceedings-Software, Vol.151, No.3, June 2004, PP.139 - 150.

[8]Kainen P.C. (1997) Utilizing geometric anomalies of high dimension: When complexity makes computation easier. In: Kárný M., Warwick K. (eds) Computer Intensive Methods in Control and Signal Processing: The Curse of Dimensionality, Springer, 1997, pp. 282–294.

[9] Martinez,A.M.;Kak,A. C. (2001). "PCA versus LDA" (PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence.*23(=2):228233.* doi*:*10.1109/34.908974

[10] M.S.Othman, L.M Yusuf and J. Salim, “Web classification using extraction and machine learning techniques”, In Information Technology (ITSim), 2010 International Symposium in Vol 2, PP. 765 -770, Kuala Lumpur, 15-17 June 2010.

[11] Roweis2000 Roweis, S. T. & Saul, L. K. Nonlinear Dimensionality Reduction by Locally Linear Embedding Science, 2000, 290, 2323-2326 131.

[12] Rubinstein2010 Rubinstein, R.; Bruckstein, A. M. & Elad, M. Dictionaries for sparse representation modeling Proc. IEEE, 2010, 98, 1045-1057

[13] T. Bourgeois, " Information Systems for Business and Beyond", Edition, Textbook Equity, Saylor Academy, 2014.

[14]Yin2008 Yin, H. Learning Nonlinear Principal Manifolds by Self-Organising Maps Lecture Notes in Computational Science and Engineering, 2008, 58, 68-95.

[15] Zhang2004 Zhang, Z. & Zha, H. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment SIAM J. Scientific Computing, 2004, 26, 313-338.