Feature (machine learning)

In machine learning and pattern recognition, a **feature** is an individual measurable property or characteristic of a phenomenon being observed.[1] Choosing informative, discriminating and independent features is a crucial step for effective algorithms in pattern recognition, classification and regression. Features are usually numeric, but structural features such as strings and graphs are used in syntactic pattern recognition. The concept of "feature" is related to that of explanatory variable used in statistical techniques such as linear regression.

Contents

Classification

A set of numeric features can be conveniently described by a feature vector. An example of reaching a two-way classification from a feature vector (related to the perceptron) consists of calculating the scalar product between the feature vector and a vector of weights, comparing the result with a threshold, and deciding the class based on the comparison.

Algorithms for classification from a feature vector include nearest neighbor classification, neural networks, and statistical techniques such as Bayesian approaches.

Examples

In character recognition, features may include histograms counting the number of black pixels along horizontal and vertical directions, number of internal holes, stroke detection and many others.

In speech recognition, features for recognizing phonemes can include noise ratios, length of sounds, relative power, filter matches and many others.

In spam detection algorithms, features may include the presence or absence of certain email headers, the email structure, the language, the frequency of specific terms, the grammatical correctness of the text.

In computer vision, there are a large number of possible features, such as edges and objects.

Extensions

In pattern recognition and machine learning, a **feature vector** is an n-dimensional vector of numerical features that represent some object. Many algorithms in machine learning require a numerical representation of objects, since such representations facilitate processing and statistical analysis. When representing images, the feature values might correspond to the pixels of an image, while when representing texts the features might be the frequencies of occurrence of textual terms. Feature vectors are equivalent to the vectors of explanatory variables used in statistical procedures such as linear regression. Feature vectors are often combined with weights using a dot product in order to construct a linear predictor function that is used to determine a score for making a prediction.

The vector space associated with these vectors is often called the **feature space**. In order to reduce the dimensionality of the feature space, a number of dimensionality reduction techniques can be employed.

Higher-level features can be obtained from already available features and added to the feature vector; for example, for the study of diseases the feature 'Age' is useful and is defined as *Age = 'Year of death' minus 'Year of birth'* . This process is referred to as **feature construction**.[2][3] Feature construction is the application of a set of constructive operators to a set of existing features resulting in construction of new features. Examples of such constructive operators include checking for the equality conditions {=, ≠}, the arithmetic operators {+,−,×, /}, the array operators {max(S), min(S), average(S)} as well as other more sophisticated operators, for example count(S,C)[4] that counts the number of features in the feature vector S satisfying some condition C or, for example, distances to other recognition classes generalized by some accepting device. Feature construction has long been considered a powerful tool for increasing both accuracy and understanding of structure, particularly in high-dimensional problems.[5] Applications include studies of disease and emotion recognition from speech.[6]

Selection and extraction

The initial set of raw features can be redundant and too large to be managed. Therefore, a preliminary step in many applications of machine learning and pattern recognition consists of selecting a subset of features, or constructing a new and reduced set of features to facilitate learning, and to improve generalization and interpretability.

Extracting or selecting features is a combination of art and science; developing systems to do so is known as feature engineering. It requires the experimentation of multiple possibilities and the combination of automated techniques with the intuition and knowledge of the domain expert. Automating this process is feature learning, where a machine not only uses features for learning, but learns the features itself.

See also

- Covariate
- Dimensionality reduction
- Feature engineering
- Hashing trick
- Statistical classification
- Explainable Artificial Intelligence

References

1. Bishop, Christopher (2006). *Pattern recognition and machine learning*. Berlin: Springer. ISBN 0-387-31073-8.
2. Liu, H., Motoda H. (1998) *Feature Selection for Knowledge Discovery and Data Mining.*, Kluwer Academic Publishers. Norwell, MA, USA. 1998.
3. Piramuthu, S., Sikora R. T. Iterative feature construction for improving inductive learning algorithms. In Journal of Expert Systems with Applications. Vol. 36 , Iss. 2 (March 2009), pp. 3401-3406, 2009
4. Bloedorn, E., Michalski, R. Data-driven constructive induction: a methodology and its applications. IEEE Intelligent Systems, Special issue on Feature Transformation and Subset Selection, pp. 30-37, March/April, 1998
5. Breiman, L. Friedman, T., Olshen, R., Stone, C. (1984) *Classification and regression trees*, Wadsworth
6. Sidorova, J., Badia T. Syntactic learning for ESEDA.1, tool for enhanced speech emotion detection and analysis. Internet Technology and Secured Transactions Conference 2009 (ICITST-2009), London, November 9–12. IEEE