generalization (Allwein et al. 2000) allows arbitrary combinations. Clearly, one against all and one against one are special cases of ECOC.

The predictions of the binary classifiers must then be combined into an overall prediction. Commonly used techniques include voting and finding the nearest neighbor in the ECOC decoding matrix (Allwein et al. 2000).

Cross-References

▶ Preference Learning
▶ Rule Learning

Recommended Reading

Allwein EL, Schapire RE, Singer Y (2000) Reducing multiclass to binary: a unifying approach for margin classifiers. J Mach Learn Res 1: 113–141
Dietterich TG, Bakiri G (1995) Solving multiclass learning problems via error-correcting output codes. J Artif Intell Res 2:263–286
Fürnkranz J (2002) Round robin classification. J Mach Learn Res 2:721–747. http://www.ai.mit.edu/projects/jmlr/papers/volume2/fuernkranz02a/html/.

Class Imbalance Problem

Charles X. Ling and Victor S. Sheng
The University of Western Ontario, London, ON, Canada

Definition

Data are said to suffer the *Class Imbalance Problem* when the class distributions are highly imbalanced. In this context, many ▶ classification learning algorithms have low predictive accuracy for the infrequent class. ▶ Cost-sensitive learning is a common approach to solve this problem.

Motivation and Background

Class imbalanced datasets occur in many real-world applications where the class distributions of data are highly imbalanced. For the two-class case, without loss of generality, one assumes that the minority or rare class is the positive class, and the majority class is the negative class. Often the minority class is very infrequent, such as 1 % of the dataset. If one applies most traditional (cost-insensitive) classifiers on the dataset, they are likely to predict everything as negative (the majority class). This was often regarded as a problem in learning from highly imbalanced datasets.

However, Provost (2000) describes two fundamental assumptions that are often made by traditional cost-insensitive classifiers. The first is that the goal of the classifiers is to maximize the accuracy (or minimize the error rate); the second is that the class distribution of the training and test datasets is the same. Under these two assumptions, predicting everything as negative for a highly imbalanced dataset *is often the right thing to do*. Drummond and Holte (2005) show that it is usually very difficult to outperform this simple classifier in this situation.

Thus, the imbalanced class problem becomes meaningful only if one or both of the two assumptions above are not true; that is, if the cost of different types of error (false positive and false negative in the binary classification) is not the same, or if the class distribution in the test data is different from that of the training data. The first case can be dealt with effectively using methods in cost-sensitive meta-learning (see ▶ Cost-sensitive learning).

In the case when the misclassification cost is not equal, it is usually more expensive to misclassify a minority (positive) example into the majority (negative) class, than a majority example into the minority class (otherwise it is more plausible to predict everything as negative). That is, $FNcost > FPcost$. Thus, given the values of *FNcost* and *FPcost*, a variety of cost-sensitive meta-learning methods can be, and have been, used to solve the class imbalance problem (Japkowicz and Stephen 2002; Ling and Li 1998). If the values of *FNcost* and *FPcost* are not unknown explicitly, *FNcost* and *FPcost* can be assigned to be proportional to the number of positive and negative training cases (Japkowicz and Stephen 2002).

In case the class distributions of training and test datasets are different (e.g., if the training data is highly imbalanced but the test data is more balanced), an obvious approach is to sample the training data such that its class distribution is the same as the test data. This can be achieved by oversampling (creating multiple copies of examples of) the minority class and/or undersampling (selecting a subset of) the majority class (Provost 2000).

Note that sometimes the number of examples of the minority class is too small for classifiers to learn adequately. This is the problem of insufficient (small) training data and different from that of imbalanced datasets.

Recommended Reading

Drummond C, Holte R (2000) Exploiting the cost (in)sensitivity of decision tree splitting criteria. In: Proceedings of the seventeenth international conference on machine learning, Stanford, pp 239–246

Drummond C, Holte R (2005) Severe class imbalance: why better algorithms aren't the answer. In: Proceedings of the sixteenth European conference of machine learning, Porto, vol 3720. LNAI, pp 539–546

Japkowicz N, Stephen S (2002) The class imbalance problem: a systematic study. Intell Data Anal 6(5):429–450

Ling CX, Li C (1998) Data mining for direct marketing – specific problems and solutions. In: Proceedings of fourth international conference on knowledge discovery and data mining (KDD-98), New York City, pp 73–79

Provost F (2000) Machine learning from imbalanced data sets 101. In: Proceedings of the AAAI'2000 workshop on imbalanced data

Classification

Chris Drummond
National Research Council of Canada, Ottawa, ON, Canada

Synonyms

Categorization; Generalization; Identification; Induction; Recognition

Definition

In common usage, the word classification means to put things into categories, group them together in some useful way. If we are screening for a disease, we would group people into those with the disease and those without. We, as humans, usually do this because things in a group, called a ▸ class in machine learning, share common characteristics. If we know the class of something, we know a lot about it. In machine learning, the term classification is most commonly associated with a particular type of learning where examples of one or more ▸ classes, labeled with the name of the class, are given to the learning algorithm. The algorithm produces a classifier which maps the properties of these examples, normally expressed as ▸ attribute-value pairs, to the class labels. A new example whose class is unknown is classified when it is given a class label by the classifier based on its properties. In machine learning, we use the word classification because we call the grouping of things a class. We should note, however, that other fields use different terms. In philosophy and statistics, the term categorization is more commonly used. In many areas, in fact, classification often refers to what is called clustering in machines learning.

Motivation and Background

Classification is a common, and important, human activity. Knowing something's class allows us to predict many of its properties and so act appropriately. Telling other people its class allows them to do the same, making for efficient communication. This emphasizes two commonly held views of the objectives of learning. First, it is a means of ▸ generalization, to predict accurately the values for previously unseen examples. Second, it is a means of compression, to make transmission or communication more efficient. Classification is certainly not a new idea and has been studied for some considerable time. From the days of the early Greek philosophers such as Socrates, we had the idea of categorization. There are essential properties of things that make them