

Goal detection in football by using Support Vector Machines for classification¹

N. Ancona, G. Cicirelli, A. Branca, A. Distante

Istituto Elaborazione Segnali ed Immagini - C.N.R.
Via Amendola 166/5 - 70126 Bari - Italy
e-mail: ancona@iesi.ba.cnr.it

Abstract

In this paper we present a technique for detecting goals during a football match by using images acquired by a single camera placed externally to the field. The method does not require the modification neither of the ball nor of the goalmouth. Due to the attitude of the camera with respect to football ground, the system can be thought of as an electronic linesman which helps the referee in establishing the occurrence of a goal during a football match. The occurrence of the event is established detecting the ball and comparing its position with respect to the location of the goalpost in image. The ball detection technique relies on a supervised learning scheme called Support Vector Machines for classification. The examples used for training are appropriately filtered version of views of the object to be detected, previously stored in form of image patterns. We have extensively tested the technique on real images in which the ball is both fully visible and partially occluded. The performance of the proposed detection scheme are measured in terms of detection rate, false positive rate and precision in the ball localization in image.

1 Introduction

In this paper we focus on the problem of detecting the occurrence of a goal during a football match, by using methods and devices which does not require the modification neither of the ball nor of the goalmouth. Automatic goal detection in football is an open problem which is getting particular attention from referee associations, sport press and supporters. In fact, there

are not rare situations in which a goal occurs¹, but the referee and his collaboratores (linesmen) are not able to detecting the goal and, more important, do not award any point to neither teams correctly. Such situations occur for example when, after a shooting, the ball touches the internal side of the crossbar, bounces off the field having crossed completely the goal line and goes back, without touching the net. One of the most significative evidences of this event, named *ghost goal*, occurred during the World Cup final match on 1966 between England and West Germany. In that case, an english player struck a shot towards the German goal and the ball cannoned down from the crossbar, hit the ground and bounced back out into the play. English players claimed a goal, that the ball had passed completely over the line, but the referee, after consultation with his linesman, did not awarded the point to the England.

Optical sensors, like standard TV cameras, seem to be appropriate for approaching the problem at hand for several reasons. First of all they satisfy the main constraint of the problem, because their exploitation does not require modifications of neither the ball nor the goalmouth. They can be placed externally and also very far from the field and, if equipped with appropriate zoom lens, they provide images of the goalmouth area usefull for solving the goal detection problem. Moreover, they permit to have a direct evidence of the occurrence of a goal because the perceived images can be recorded on an analog or digital support for a successive analysis, for example by an external referee.

Reid and Zisserman [1] proposed an uncalibrated binocular vision system for solving the problem of goal detection in football. Their method exploits two images of the field acquired simultaneously from two different viewpoints. In both images both the goal area and the

¹ **Acknowledgements:** this paper describes research done at the Istituto Elaborazioni Segnali ed Immagini, C.N.R. in Bari. Partial support was also provided by the italian football association Federazione Italiana Gioco Calcio FIGC.

¹ A goal occurs in football when the ball completely crosses the goal line.



Figure 1: View of the goalmouth perceived by our electronic linesman placed closed to the corner flag.

goalmouth are visible. The computation of the vertical vanishing point in both images and of the homography between the two images induced by the field are used for measuring the projection of the ball onto the ground plane. The distance of this point with respect to the goal line is used for establishing if a goal occurs. A strong limitation is that their method is not able to compute the three-dimensional coordinates of the ball, and so it cannot be used as an automatic goal detector. In fact, the algorithm computes only the projection of the ball on the ground plane, and it is not able to distinguish if the ball crosses the goal line below or above the crossbar. Finally, an open problem in their paper is the detection and localization of the ball in both images for triangulating.

In this paper we describe a method that, as the one proposed by Reid and Zisserman in [1], does not provide a general and fully automatic solution to the problem of goal detection in football. Our attempt was to design a method well suited to the detection of a particular event potentially occurring during a football match, namely a ghost goal. A deeper analysis of the problem at hand, and in particular of the ghost goal detection problem, shows that the best attitude for a linesman for detecting the complete overcoming of the ball of the goal line during a ghost goal is close to the corner flag. In fact, for detecting the goal from this view point, the linesman has to simply establish if the ball is to the left (right) with respect to the goalpost (see figure 1). Then, in general, a monocular observer having its optical axes lying on the goalmouth plane with a viewing direction oriented towards the goal line can detect the occurrence of a goast goal simply evaluating the relative position between the ball and the goalpost

in the perceived image. So the main problem that an electronic linesman has to solve for establishing the occurrence of a ghost goal is ball detection in images. The plan of the paper is as follow. In the next section we present a general framework for object detection in images and possible approaches for its solution. In section 3 we briefly discuss the main properties of Support Vector Machines (SVM) for classification, the supervised learning scheme used in this paper for detecting balls in images. In section 4 we discuss the steps performed for training an SVM and show performances of the ball detector applied on real images in terms of detection rate, false positive rate and precision in the localization of the ball in images. Conclusions follow.

2 Object detection in image

Under this perspective the problem of detecting goal can be reduced to the problem of detecting the ball in images of the goalmouth taken from a suitable viewpoint. This is a particular instance of a more general problem that is the one of detecting three dimensional objects by using the image projected by the object on the sensing plane of a standard camera. This problem, widely recognized as challenging by the computer vision and, more recently, by the neural network communities, has been previously approached by using standard vision algorithms. Region growing, edge detection, snakes, texture analysis are the most common used techniques for facing with the problem of detecting the most relevant visual information from the image at the aim of establishing whether a given object is present in the perceived image. A huge amount of possible pattern variations in the image, that are difficult to parametrize analytically, makes the problem of automatic object detection untractable if we base the detection mechanism on the aforementioned techniques. As an example, think to the problem of detecting human faces from an image [2]. The attitude of the face, the color of the skin, the expression, the presence or absence of common structural features like moustache or glasses, the shadows caused by particular light source distributions are common sources of pattern variations which make clear the difficulty of the problem at hand.

Recently, the problem of detecting objects has been addressed by a new and appealing perspective [3] in which the criteria for establishing whether or not a given image pattern is the instance of an object is based on *views* of the object, previously stored in form of image patterns. Under this new perspective, the problem of object detection can be regarded as a *learning from examples* problem in which the examples are particu-

lar views of the object we are interested to detect, and then many of the supervised learning schemes can be usefully applied for solving the problem at hand. More specifically, object detection can be seen as a *classification* problem, because our ultimate goal is to determine a separating surface, optimal under certain conditions, which is able to separate object views from image patterns that are not instances of the object. It is well known that the general problem of learning from examples, and in particular classification, can be interpreted as the problem of approximating a multivariate function from sparse data [4], where the data are in the form of (input, output) pairs, obtained by random sampling the unknown function in the presence of noise. This problem is clearly ill-posed, since it has an infinite number of solutions and, in order to choose one particular solution, we need to have some a priori knowledge of the function that has to be reconstructed (see [5] and the references therein). In [4, 5] the authors approach the problem of multivariate function approximation by using regularization theory and the a priori knowledge of the function takes the form of a smoothness functional. More recently, Vapnik [6] has introduced a new learning scheme, well founded in the framework of the statistical learning theory, called Support Vector Machines (SVM) for approaching classification and regression problems. The basic idea of the Vapnik's theory is closely related to regularization [7]: for a finite set of training examples, the search for the best model or approximating function has to be constrained by an appropriately small hypothesis space, that is the set of functions the machine implements. If the space is too large, functions can be found which fit exactly the data, but they will have a poor generalization capabilities on new data. Vapnik's theory formalizes these concepts and shows that the solution is found minimizing both the error on the training set (empirical risk) and the complexity of the hypothesis space, expressed in terms of VC-dimension. In this sense, as we show in the next section, the function found by SVM is a tradeoff between closeness to the data and complexity of the solution.

3 Support Vector Machines for classification

In this section we review the basic concepts of SVM for two classes classification problems [6] for the general case of not linearly separable classes with linear and not linear surfaces. We are given a training set $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{\ell}$ of size ℓ where $\mathbf{x}_i \in R^n$ and $y_i \in \{-1, 1\}$, for $i = 1, 2, \dots, \ell$. In other words we assume that the examples in S belong to either of two classes. In the general hypothesis of not linearly

separable classes, the optimal separating hyperplane $\mathbf{w}^* \cdot \mathbf{x} + b^* = 0$ found by SVM is solution of the following quadratic programming (QP) problem with linear constraints:

Problem 1

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{\ell} \xi_i \\ \text{subject to} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) + \xi_i \geq 1 \quad i = 1, 2, \dots, \ell \\ & \xi_i \geq 0 \end{aligned}$$

where C is a positive number and the non negative slack variable ξ_i (one for each point in S) measures the amount of misclassification of the point \mathbf{x}_i with respect to the optimal separating hyperplane. In fact $\xi_i = 0$ if the point \mathbf{x}_i is correctly classified. Some considerations are in order. As we have shown, the slack variables assume positive values for misclassified points, while they vanish for correctly classified ones. So, the term $\sum_{i=1}^{\ell} \xi_i$ in the objective function of the problem (1) is a quantity proportional to the number of misclassified points of the training set. Notice that other quantities can be used for measuring the amount of misclassified points, such as $\sum_{i=1}^{\ell} \xi_i^2$. However, the optimal separating hyperplane obtained by using $\sum_{i=1}^{\ell} \xi_i$ as a misclassification measure is more robust from a statistical point of view, because the solution is less sensitive to the presence of outliers in the training set.

The objective function of the problem (1) expresses two properties of the solution in the general case of linearly non separable classes. In fact, minimizing the first term is equivalent to maximizing the distance between the optimal separating hyperplane and the closest points in S . Moreover, minimizing the second term is equivalent to minimizing the number of misclassified points. The constant C , which can be regarded as a *regularization parameter*, controls these two terms during the training process. In fact, for small values of C , the optimal separating hyperplane tends to maximize the distance of the closest point of S . For large values of C , the optimal separating hyperplane tends to minimize the non correctly points of S . For intermediate values of C , the solution of the problem (1) is a tradeoff between maximum margin and minimum number of misclassified points.

The QP problem with linear constraints (1) can be solved by using the standard technique of Lagrange multipliers. At this aim, we introduce ℓ non negative slack variables λ_i relative to the constraints $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) + \xi_i \geq 1$, and ℓ non negative slack variables μ_i

relative to the constraints $\xi_i \geq 0$. If we denote with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_\ell)$ and with $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_\ell)$ the 2ℓ Lagrange multipliers relative to the constraints of the problem (1), then solving (1) is equivalent to determining the saddle point of the Lagrangian function:

$$L = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \lambda_i [y_i (\mathbf{w} \cdot \mathbf{x}_i + b) + \xi_i - 1] - \sum_{i=1}^N \mu_i \xi_i \quad (1)$$

where $L = L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$. So the optimal \mathbf{w}^* is:

$$\mathbf{w}^* = \sum_{i=1}^{\ell} \lambda_i^* y_i \mathbf{x}_i \quad (2)$$

where the optimum $\boldsymbol{\lambda}^*$ is solution of the dual problem of the problem (1):

Problem 2

$$\max_{\boldsymbol{\lambda}} \quad -\frac{1}{2} \boldsymbol{\lambda} \cdot \mathbf{D} \boldsymbol{\lambda} + \sum_{i=1}^{\ell} \lambda_i$$

$$\text{subject to} \quad \sum_{i=1}^{\ell} \lambda_i y_i = 0 \\ 0 \leq \lambda_i \leq C \quad i = 1, 2, \dots, \ell$$

where \mathbf{D} is a matrix of size $\ell \times \ell$, with $D_{ij} = y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$ per $i, j = 1, 2, \dots, \ell$. Moreover the optimal b^* can be computed by using the Kuhn-Tucker conditions:

$$(C - \lambda_i^*) \xi_i^* = 0 \quad i = 1, 2, \dots, \ell \quad (3)$$

$$\lambda_i^* [y_i (\mathbf{w}^* \cdot \mathbf{x}_i + b^*) + \xi_i^* - 1] = 0 \quad i = 1, 2, \dots, \ell \quad (4)$$

where ξ_i^* are the values of ξ_i at the saddle point. In fact, from the Kuhn-Tucker condition (4) we have that:

$$b^* = y_i - \mathbf{w}^* \cdot \mathbf{x}_i \quad \forall i \ni' 0 < \lambda_i^* < C$$

The points \mathbf{x}_i with $\lambda_i^* > 0$ are called *support vectors*. The classification of a new data \mathbf{x} involves the evaluation of the decision function:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^{\ell} \lambda_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^* \right) \quad (5)$$

where the solution is expressed evaluating the dot product between the data and some elements (support vectors) of the training set S .

3.1 Extension to non linear separating surfaces

The extension of the theory to the general case of non linear separating surfaces is done by mapping the input vectors \mathbf{x} in a higher dimensional space, called *feature*

space, and looking for the optimal separating hyperplane in this new space. Let $\phi(\mathbf{x})$ be the image of the point \mathbf{x} in the feature space, with:

$$\phi(\mathbf{x}) = (a_1 \phi_1(\mathbf{x}), a_2 \phi_2(\mathbf{x}), \dots, a_n \phi_n(\mathbf{x}), \dots)$$

where $\{a_n\}_{n=1}^{\infty}$ are real numbers and $\{\phi_n\}_{n=1}^{\infty}$ are real functions. In the feature space induced by the mapping ϕ , the optimal separating hyperplane found by SVM has the form:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^{\ell} \lambda_i^* y_i \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}) + b^* \right) \quad (6)$$

where the inner product of vectors in the feature space is:

$$\phi(\mathbf{x}) \cdot \phi(\mathbf{y}) = \sum_{n=1}^{\infty} a_n^2 \phi_n(\mathbf{x}) \phi_n(\mathbf{y})$$

Let K a function of two variables \mathbf{x} and \mathbf{y} of the input space which estimates the inner product of their corresponding images, $\phi(\mathbf{x})$ and $\phi(\mathbf{y})$, in the feature space, that is:

$$K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{y})$$

Then, the optimal separating hyperplane in the feature space (6) can be written as a non linear separating surface in the input space:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^{\ell} \lambda_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right) \quad (7)$$

represented as a linear combination of kernel functions centered on the support vectors only. The Mercer's theorem [6] establishes general conditions for a kernel function K to estimate inner products in Hilbert spaces. In fact, suppose K a continuous symmetric function, kernel of the positive definite integral operator:

$$(T_K f)(\mathbf{x}) = \int K(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}$$

Then K admits an expansion of the form:

$$K(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \lambda_n \phi_n(\mathbf{x}) \phi_n(\mathbf{y})$$

where ϕ_n are the mutually orthogonal eigen-functions and λ_n the corresponding eigen-values of the integral operator T_K , that is they are solution of the following integral equation:

$$\int K(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y} = \lambda \phi(\mathbf{x})$$

It is important to point out that the mutually orthogonal functions ϕ_n (features) span a Hilbert space in which the optimal classifiers lives. In other words, specifying the kernel function K used in SVM is equivalent to specify the set of all possible classifier that the machine implements, or the complexity of the function space in which the final classifier lives.

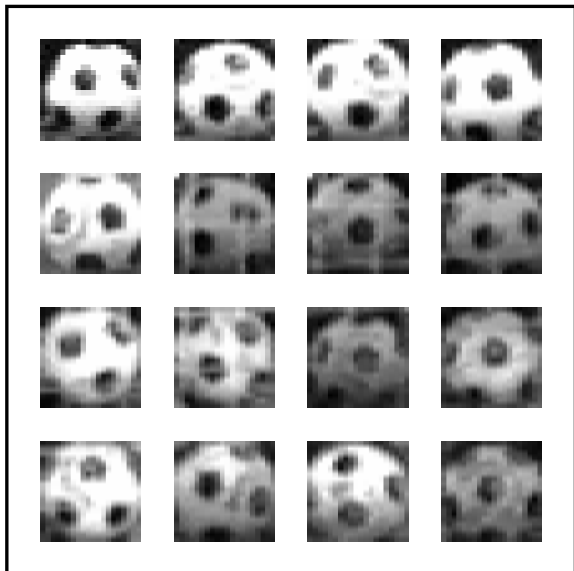


Figure 2: Image patterns of football.

4 Experimental results

In this section we describe the steps performed for collecting data, training an SVM and estimating the performances of the obtained classifier. We used a standard TV camera with a zoom lens having a focal length of $f = 75mm$. At the aim of reducing motion blurring effects, a shutting time of $1/10000sec$ was used. The camera was placed externally to the football ground, the height of its optical center was $1.5m$ roughly and its optical axis was manually aligned with the goal line. The distance of the camera with respect to the center of the two goalposts was $48m$. The figure (1) shows a typical image acquired by the camera in this attitude.

One of the main advantages associated with the chosen camera attitude is that the football size projected on the camera plane is almost constant moving the football inside the area being monitored for detecting goal. In fact, for 3D points belonging to this area, the image formation process can be described, from a geometric point of view, by using orthographic projection, instead of perspective projection. This is mainly due to the distance between the camera and the goalmouth, to the adopted camera focal length and to the fact that the area we are interested to monitoring is close to the optical axis of the camera. It is well known that, under orthographic projection, the size of the objects perceived by a camera is invariant with respect to the distance between the sensing plane and the object. In our context, this means that the football size projected on the camera plane does not change moving the football

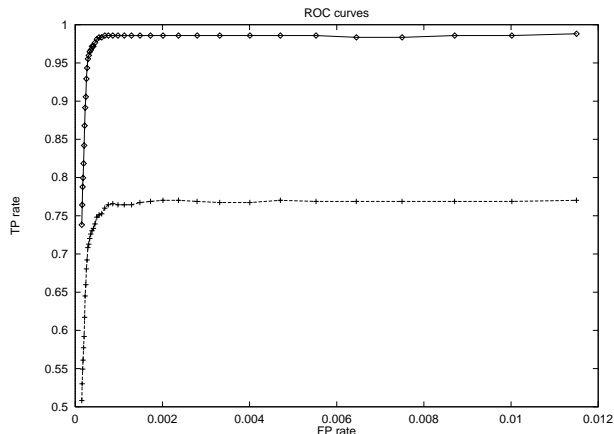


Figure 3: ROC curves on the whole test set. The upper curve is relative to fully visible balls. The lower curve was computed considering also occlusion cases.

inside the area being monitored for goal detection. In fact, we have experimentally verified that the perceived football size varies of $1pxl$ roughly moving the football from one goalpost to the other and close to the goal line. This implies that, due to the chosen camera attitude and the adopted focal length, any algorithm for detecting football in the area close to the goalmouth does not need to manage scale variations of the object in the image.

For collecting positive examples, we acquired 2004 images containing one textured, standard football each. In the present context, positive examples are views of the football, in form of images (see figure (2)). We acquired the football in different attitudes and illumination conditions, and in different positions inside the area being monitored, for example internally and externally to the goalmouth. Each example, with a size of $20 \times 20pxl$, was manually extracted. In all the positive examples the football was totally visible. This means that, in the current implementation, the problem of detecting partially occluded footballs was not taken into account.

The procedure adopted for collecting negative examples, i.e. image patterns which are not instances of the exploited football, involved different steps. We acquired many images of the stadium in which the football was not present, framing peoples, advertising posters and places containing potential false positive patterns. Notice that each 20×20 sub-image of these images is a negative example and, in principle, it should be considered in the training process. Considering that each image not containing the football produces about 100,000 negative examples, some appropriate methods

has to be used at the aim of reducing the exponential growth of negative examples. In particular we need to use some technique which is able not only to keep low the number of negative examples used for training, but also, and more important, to select image patterns which are relevant for the problem at hand, i.e. for detecting footballs in the present context. At this aim (see [3, 8]), we collected 1230 negative examples, sampling on a regular grid a negative examples' image. A training set composed of 3234 examples was used for training an SVM for classification with a second degree polynomial kernel $K(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x} \cdot \mathbf{y})^2$ and a regularization parameter of $C = 200$. We tested the obtained classifier on a new image not containing instances of the football and on this image, we found 3647 false positive image patterns. The selected negative examples were added to the training set and the training process was repeated, by using a total of 6881 examples and the same kernel function and regularization parameter used before. This procedure of search of negative examples relevant for the problem at hand was iterated several times, each time using different negative example images. In particular, the 5 successive images produced 277 negative image patterns and the last 37 images produced 2817 negative image patterns. The final classifier was so obtained training an SVM on 9975 positive and negative examples by using the same kernel function and regularization parameter. Notice that the refinement process involved 43 false positive images for a total of over 4 millions of negative examples. The number of required support vectors for representing the optimal classifier was 2237, including 165 and 976 positive and negative errors respectively.

All the examples were appropriately preprocessed before training. First, pixels close to the boundary of each example window were removed in order to eliminate parts belonging to the background. Then a histogram equalization was applied to reduce variations in image brightness and contrast. The resulting pixels were used as input to the classifier.

For measuring the generalization capabilities of the learning machine, that is the ability of the machine to correctly classifying image patterns never seen before, we tested the classifier on 900 images acquired under different illumination conditions. Each test image was exhaustively scanned and all the sub-images with size $20 \times 20pxl$ were classified as instance of the football or not. The figure (1) shows a typical image used for testing. For better understanding the performances of the classifier, we analyzed all the test images checking for the visibility of the football, before of the classification process. We counted the images in which the football

was visible, occluded and partially occluded, with occlusion less than or greater than 50%. The ROC curves in figure (3) show the performances of the classifier on images with fully visible footballs (upper curve) and on images with occluded footballs (lower curve). In the first case we had a detection rate of 98.3% with a false positive rate of 0.2%; in the second case, where occluded footballs were considered too, we had a detection rate of 76.2% with a false positive rate of 2.6%.

5 Conclusions

In this paper we focus on the problem of detecting 3D objects by using the image they project on the sensing plane of a camera. A particular instance of this general problem is the detection of a ball in images at the aim of detecting goals during a football match. The system can be seen as an electronic linesman which helps the referee to establish the occurrence of a goal.

References

- [1] Ian Reid and Andrew Zisserman. Goal-directed video metrology. In *4th European Conference on Computer Vision '96*, Cambridge, April 1996.
- [2] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.
- [3] K. Sung and T. Poggio. Example-based learning for view-based human face detection. Technical Report A.I. Memo No. 1521, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1994.
- [4] T. Poggio and F. Girosi. A theory of networks for approximation and learning. Technical Report A.I. Memo No. 1140, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1989.
- [5] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7:219–269, 1995.
- [6] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.
- [7] T. Evgenious, M. Pontil, and T. Poggio. A unified framework for regularization networks and support vector machines. Technical Report A.I. Memo No. 1654, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1999.
- [8] Avrim Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97:245–271, 1997.