Fuzzy Classifiers for Imbalanced, Complex Classes of Varying Size

Sofia Visa
Department of ECECS,
University of Cincinnati,
Cincinnati, OH 45221-0030, USA
svisa@ececs.uc.edu

Anca Ralescu
Department of ECECS,
University of Cincinnati,
Cincinnati, OH 45221-0030, USA
aralescu@ececs.uc.edu

Abstract

In this paper we investigate the suitability of a fuzzy system as a classifier for imbalanced data problems. Primarily, the fuzzy model performance is evaluated on artificial data sets, generated with various levels of size, complexity and imbalance. It is investigated what combination of the three problematic issues makes the learning problem harder [4]. A theoretic analysis shows that for a fuzzy classifier the "imbalance problem" is no longer a problem. By considering a relative frequency *to the class size* the imbalance factor is eliminated.

Keywords: Classification, Fuzzy Systems, Imbalanced Data.

1 Introduction

Many real world problems deal with imbalanced learning. More exactly, there are very few examples available form a class in respect to the other class. If the function to be learned by the system it is complex too, besides the imbalance factor, the classifier has difficulties to model the data. In this paper it is investigated how much imbalance and complexity a classifier can handle. Results of research work in this area have important practical applications in fields such as:

- Medicine - diagnosis of rare disease;
- NASA - system failure of a shuttle in mission;
- Army - discriminating between friendly and hostile weapons;
- Industry - credit cards or phone calls fraud detection.

In all these cases a system capable to recognize the imbalanced class (small class) is desirable. As described in [5] in some cases, human operators of such systems learn to recognize conditions leading to failure and act before this actually occurs and, in many cases automatic systems can be trained to recognize failures too.

The issue of learning of imbalanced classes (obviously from imbalanced data sets) has recently received increased attention, as it can be seen from the scientific events dedicated to this topic [11]. A complete review of the various approaches adopted for this problem is not within the scope of this paper, although it is interesting to note that from the research to date it appears that the usual approaches to learning, such as decision trees, networks, etc, are not immediately applicable to this problem. A common, and justifiable approach is to re-balance the classes, either by increasing the number of data points for the small class *up-sampling* or by decreasing the number of points in the large class *down:sampling*. Another common approach is that of assigning different error penalties: an error on a data point belonging to the small class would have a larger penalty than one for the large class. However, it can be argued that the effect of assessing penalties is equivalent to changing the relative data distribution in the two classes, or, in other words, to re-balancing the data.

A fuzzy set based approach for this type of problem is new, having been introduced by the authors in [7], although the basic technique for deriving the fuzzy sets is not. It can be said that the study presented in [5] addressed the same problem although not explicitly and the emphasis there was more on the procedure of deriving the membership function for a class than on the type of classes to be learned. Different from [7] where the use of the fuzzy set approach was deemed necessary because of the overlapping nature of the data, in the current approach this is not the case. The results of [7] suggested that the overlap (and not the imbalance) affected the performance of the fuzzy classifier, fact that motivated the current study. Further motivation came from the desire of duplicating the experiments of [4] were a neural network classifier was derived for modeling imbalanced classes. It is expected that, because a fuzzy set reflects the size of the training data used to derive it, such an approach does not need preprocessing of the data, such a re-balancing or penalty assessment.

Table 1: Example of data sets: each table cell shows the number of data points for the positive(+), negative(-) classes, for the indicated combination of complexity (c), size(s) and imbalance levels(i).

2 INTERVALS (c=1)			**32 INTERVALS (c=5)**		
	s=1	s=5		s=1	s=5
i=1	157+ 20-	2500+ 156-	i=1	160+ 16-	2496+ 160-
i=5	157+ 157-	2500+ 2500-	i=5	160+ 160-	2496+ 2496-

2 Data Sets Generation

For this study the binary learning problem (positive and negative class) is considered. The data sets (total number of points is 5000) are generated randomly from the interval $0 - 1000$ using a uniform distribution. There are $5^5 = 125$ data sets each of which consists of one dimensional data points labeled 1 or 0 for the positive or negative class respectively. In generating the data set five levels of complexity $(c = 1, \ldots, 5)$, size $(s = 1, \ldots, 5)$ and imbalance $(i = 1, \ldots, 5)$ are considered such that each of the 125 sets is characterized by a different combination of these three aspects: complexity/size/imbalance. The domain is divided into 2^c equal intervals such that contiguous intervals have opposite class labels. The number of alternating intervals gives the complexity level [4]. The number of points sampled from each class (and hence interval) depends on the size of the domains as well as on the imbalance level.

More precisely, the complexity determines the number of alternating intervals in the domain $0, \ldots, 1000$ according to the equation (1):

$$complexity = 2^c, \quad c = 1, \ldots, 5 \qquad (1)$$

The different size levels are used to generate sets of different sizes according to the equation:

$$size = round\left[\left(\frac{5000}{32}\right)^s * 2^c\right], \quad s = 1, \ldots, 5 \qquad (2)$$

where c is a given complexity. For example the biggest data set has 5000 data points (for $s = 5$) and the smallest has 312 points (for $s = 1$).

The imbalance factor is applied only to the negative class, so that if from a total of M (here 5000) points N are sampled for the positive class, then $N * \frac{1}{32/2^i}$ data points are sampled for the negative class, for $i = 1, \ldots, 5$. The values $i = 1$ and $i = 5$ lead to highest and lowest (no imbalance) imbalance between the classes respectively. To further illustrate the data sets, table 1 shows eight examples of data sets (the extreme cases).

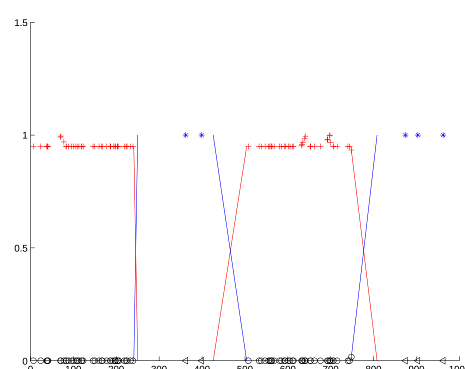

Figure 1: Test on file 211 - 0 errors.

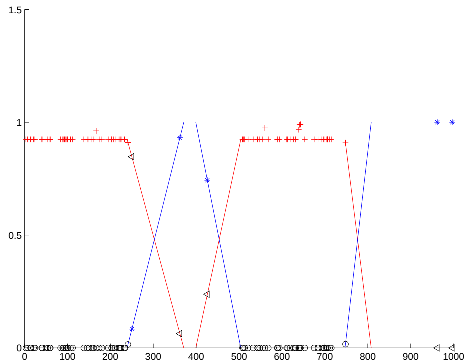

Figure 2: Test on file 211 - 1 error.

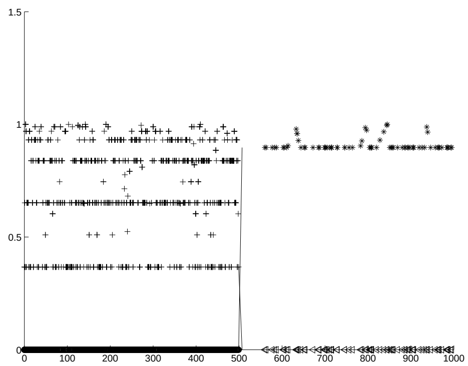

Figure 3: Test on file 151 - 0 errors.

3 Methods

As already mentioned in section 1 various exisitng approaches have been applied to the problem of learning of imbalanced classes. A brief critical review, which further justifies the current approach, is included below:

3.1 Existing Approaches

Traditional approaches to classifiers have been presented in several studies, including neural networks, decision trees, support vector machines, [3], [4], [10].

It appears that each of these approaches suffers when used for the problem of learning of imbalanced classes. For example, for decision trees, pruning of the decision tree may cause serious problems such as loosing entirely the small (and therefore the most interesting) class. Moreover, if the small class is assigned high penalty for errors (e.g. system failure in the space shuttle), then definitely a decision tree is not the right system to use for this classification problem.

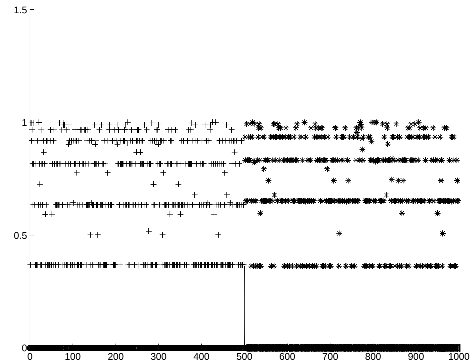

Figure 4: Test on file 151 - 1 error.

Problems may arise also when using a neural network approach for imbalanced classification: neural network might consider the small class as noise and end up not learning it at all or the computed weights will be biased towards the class containing many examples. To avoid this type of mistraining, the neural network is trained using a re-balanced data set.

The problems mentioned above are not present in a fuzzy system as this will neither loose the small class nor consider it as noise. The two classes are modeled independently, each by a fuzzy set which are used in the subsequent testing step. The fuzzy model both **learns the class representation and discriminates between classes** (it combines the discrimination and recognition methods).

Moreover, training of the fuzzy classifier is not order dependent, whereas training of a neural network depends on the order in which the training data is fed, sometimes to such extent that for some particular order of data the network does not converge (although for a shuffle of the same data it will). For decision trees, different trees result for different selection of features.

Finally, from a computational point of view, fitting the weights of a neural network or finding the smallest decision tree are both NP-complete problems [2]. For this reason heuristic algorithms such as gradient descent (for neural networks) and greedy search (for decision trees) are necessary and have been applied with great success. By contrast, from this point of view too, the fuzzy classifier, presents an additional advantage, namely that it is **computationally deterministic**.

3.2 The Current Approach : The Fuzzy Classifier for Imbalanced Classes

As already mentioned the current work extends the ideas first proposed in [7]. For the sake of completeness this section includes some the ideas presented there.

3.2.1 Generating a fuzzy set from data

Given a collection of data points, summarized by a normalized histogram (or empirical discrete probability distribution),

$$H = \{f_{(i)}; 0 \leq f_{(i)} \leq 1; \sum_{i=1}^{n} f_{(i)} = 1\} \quad (3)$$

a fuzzy set representation for it can be defined as shown in equation (4).

$$\mu_{(k)} = kf_{(k)} + f_{(k+1)} + f_{(k+2)} + ... + f_{(n)} \quad (4)$$

where $f_{(k)}$ and $\mu_{(k)}$ denote the kth largest value of the frequency distribution and membership function respectively.

Equation (4) is derived as a particular case of a general procedure converting a relative frequency distribution into a fuzzy set [6]. The fuzzy set corresponding to this conversion is maximal in the sense that it can be shown to be the closest to the crisp (non-fuzzy) set with the same support. It is easy to see that when the data are sampled according to the uniform distribution such that $f_{(k)} = \frac{1}{n}$, for all k, the membership function obtained according to (4) reduces to the indicator function of a crisp set, that is, $\mu_{(k)} = 1$ for all k.

3.2.2 Training the fuzzy classifier

The two classes were separately modeled into fuzzy sets as described on previous paragraph. The fuzzy set from the positive class is extended with the endpoints of the subintervals from the negative class and their membership values to the positive class are set to 0. Similarly, the final fuzzy set was derived for the negative class. This augmentation step it is necessary for the subintervals margin areas: the examples used to train the classifier do not cover the whole subinterval and for the test points coming from the uncovered margins areas, the memberships values to the two fuzzy sets still has to be evaluated. In this way, the positive and negative fuzzy sets stretch on the whole domain, as can be observed from Figure 5. The alternating subintervals (belonging to different classes) are not overlapped but, due to the augmentation step, the constructed fuzzy set is. A geometrical illustration of the obtained fuzzy sets, after the augmentation step, is presented in Figure 5 for a level of complexity $c = 2$.

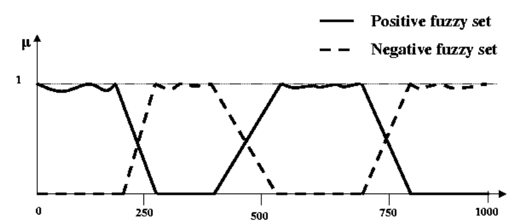

Figure 5: An intuitive example of positive and negative final fuzzy sets.

3.2.3 Testing the fuzzy classifier

For a testing point, two membership values to the two fuzzy sets obtained at the training step are computed and compared. The data point is consider to belong to the class to which the membership value it is greater. When a testing point is not present in the training set its membership value to the set is computed by interpolation of the nearest points that surround the given point (see Figures 1, 2, 3 and 4); in this figures the first digit on the file name represents the complexity, the second, the size and the third, the imbalance level. For the positive test data, the membership values to the positive fuzzy set are plotted as pluses and to the negative fuzzy set as circles; similarly, for the negative test data the membership to the positive fuzzy set is represented by triangles and to the negative fuzzy set by stars. The lines represent the interpolation line obtained from training at augmentation step in the endpoints regions of the subintervals, as described on previous paragraph.

4 Results

The results are evaluated on average over 100 runs for randomly selected training and testing sets. For each run 50% of data were used in the train-

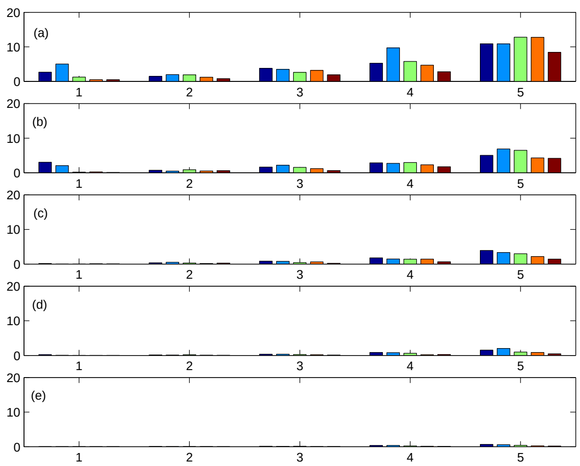

Figure 6: Cumulated error (positive and negative class). Figures $(a), \ldots, (e)$ correspond to $s = 1, \ldots, 5$; In each figure, the groups 1 to 5 correspond to complexities levels, $c = 1, \ldots, 5$ and, within each group, the bars correspond to the error for imbalance levels $i = 1, \ldots, 5$.

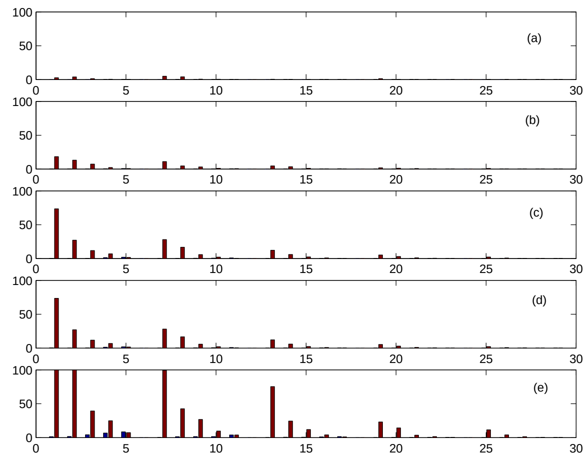

Figure 7: The error measured independently for the negative and positive class: the plots $(a), \ldots, (e)$ correspond to $c = 1, \ldots, 5$; the groups labeled 5, 10, 15, 20, 25 correspond to $s = 1, \ldots, 5$, and inside each group the imbalance level is $s = 1, \ldots, 5$, from left to right. For each pair of bars, the left bar represents the error for positive class and the right bar for the negative(imbalanced) class.

ing step to obtain the fuzzy sets and the other 50% of data were used for testing purpose.

Figure 6 shows the cumulated error over both classes for all 25 data sets, respectively. In fact the error for the positive class is much smaller in all the situations and this can be observed from the Figure 7 where the error is plotted individually for each class.

Figures 8, 9 and 10 show the error surface (on the imbalanced - negative - class, only) as a function of two aspects when the third is constant. For example, the top plot in figure 8 represents the error as a function of complexity and size for the imbalance level $i = 1$ (the highest imbalance level) while the bottom plot in the same figure represents the same for the imbalance level $i = 5$ (there is no imbalance).

From these figures it can be observed that the classification is poor only for a **critical combination** of *small size (s=1), high complexity (c=5) and high imbalance (i=1) (see Figures 8(top), 9(bottom) and 10(top)*. The critical combination is given by the extreme cases: very few data points to model the classes, high imbalance and very complex class to be learned by the system. When only two of these three critical factors are present, the fuzzy classifier performs well. For example, even for a very complex class with high imbalance when sufficient training points $(s = 5)$ are given, the fuzzy classifier can learn well the class as shown in figure 10(bottom). Similarly, the other two combinations are plotted in figure 8(top) for small size, high imbalance but low complexity $(c = 1)$ and in figure 9(bottom) for small size, high complexity and no imbalance $(i = 5)$.

The imbalance factor affects the performance **only** for a combination of *high complexity* of the concept to be learned and very *small data size*. In this case the training data is not sufficient to cover well the complex domain so the classifier is unable to learn (parts of) the small class. For example,when $c = 5$, $s = 1$ and $i = 1$ many of the 16 subintervals of the imbalanced class do not have data points in the training step since there are only 8 points to model the class. In fact, in this case, the lowest recognition error that one might expect (using any classifier) is at least 50% (when each of the eight training points falls in a different interval). For increased size of the training set, even with a high complexity level the fuzzy classifier does well as the plot in figure 10(bottom) shows.

Comparing these results with a similar experi-

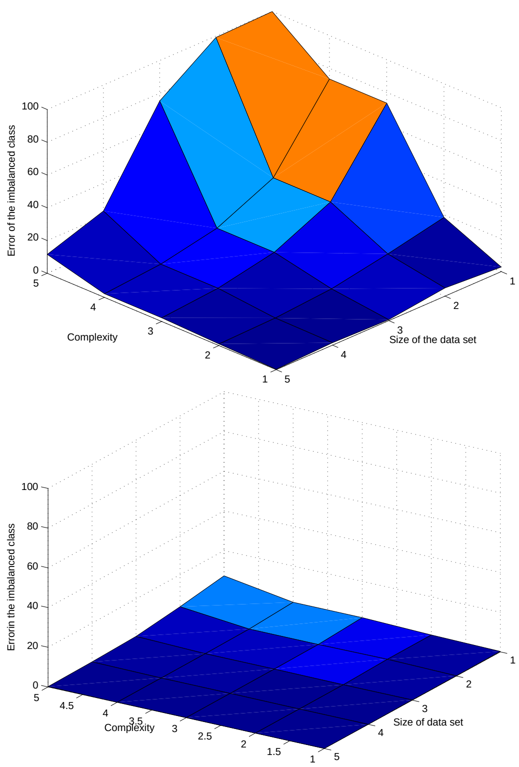

Figure 8: The error surface as a function of size and complexity for constant imbalance 1(top) and 5(bottom).

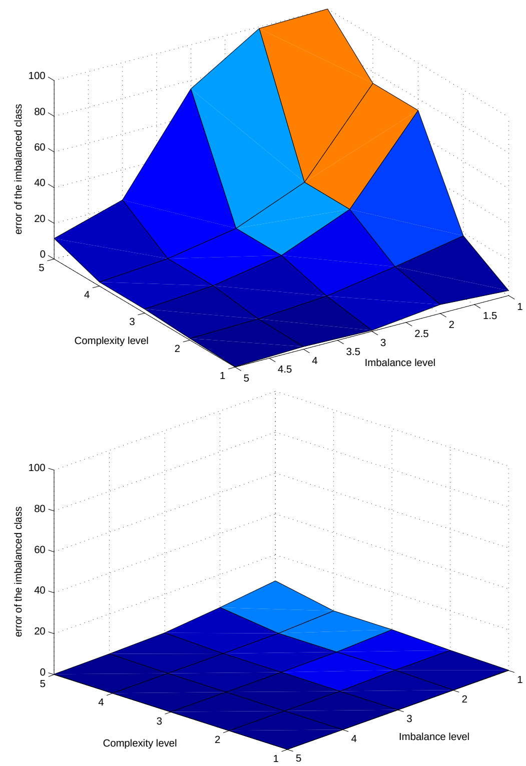

Figure 10: The error surface in function of complexity and imbalance for constant size level 1(up) and 5(down).

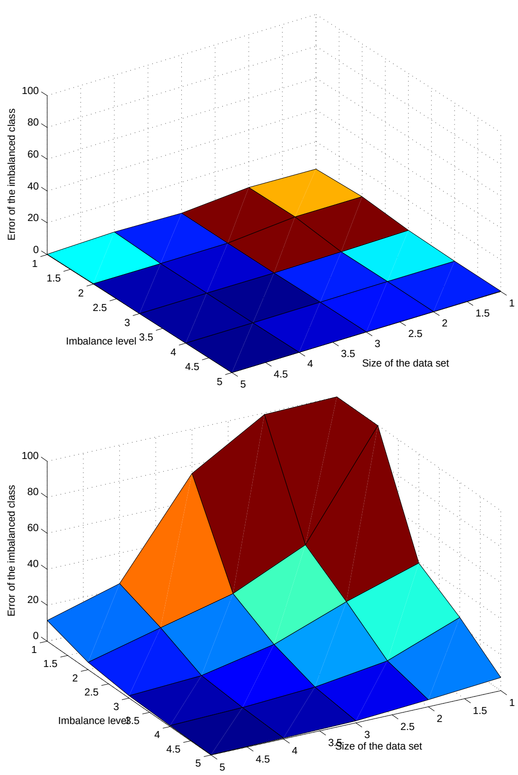

Figure 9: The error surface in function of size and imbalance for constant complexity 1(top) and 5(bottom).

ment [4] where a neural network was used as classifier, it can be concluded that the fuzzy model learns better both the balance and imbalanced classes and that its prediction power is much better (Figure 6).

5 Conclusions

As already mentioned when applied to the problem of modeling imbalanced classes the fuzzy approach eliminates the effect of imbalance. This is due the fact that the fuzzy classifier models each class independently taking into account the size of the training set for each class rather than that of the whole training set.

The fuzzy set approach deals well with the imbalance aspect in all but one combination of the complexity and size levels: *high complexity, small overall size of data set*. For any other combination of imbalance / complexity / size, the model is robust and gives good classification rate. A fuzzy approach is a valid answer for the question *"What classifier is suitable for imbalanced classification* *problems?"*.

6 Future work

Taking into account the current research on fuzzy classifiers, and previous reported work on neural networks [4], decision trees [3], support vectors machines [10] for the problem of learning of imbalanced classes one can envision various directions for future work in this area. This should include probabilistic approaches and the suitability of the probabilistic classifiers for imbalanced data classification. It was assumed in the current study as well as in [4] that the two classes have the same underlying distribution (uniform). Studies where these distributions are different should also be considered. Another problem, most useful in the context of online learning, is the study of the imbalanced problem in the framework of the adaptive learning systems in which as data are obtained online the three factors of complexity, imbalance and size change.

Acknowledgements

Sofia Visa's work for this study was partially supported by a Graduate Fellowship from Ohio Board of Regents. Anca Ralescu's work for this study was partially supported by the Grant XXXXXXX of the Department of Navy.

References

[1] D. W. Aha. A study of instance-based algorithms for supervised learning tasks: Mathematical, empirical, and psychological evaluations. Department of Information & Computer Science, University of California, Irvine, 1990.

[2] A. Blum and R. Rivest. Training a 3-node neural net is NP-complete. In *Advances in Neural Information Processing Systems I*, pages 494-501, Morgan Kaufmann, 1989.

[3] C. Drummond and R. Holte. C4.5, Class Imbalance, and Cost Sensitivity: Why Undersampling beats Over-Sampling. In *Proceedings of the 20th ICML 2003 Workshop: Learning with Imbalanced Data Sets II*, pages 1-8, Washington, August, 2003.

[4] N. Japkowicz (2000). Learning from Imbalanced data sets: A Comparison of Various Strategies. In *Proceedings of Learning form Imbalanced data 2000*, pages 10-15.

[5] H. Narazaki and A. Ralescu (1994). Iterative Induction of a Category Membership Function. In *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 2.1, pages 91-100.

[6] A. Ralescu (1986). A note on rule representation in expert systems. In *Information Science*, 38.2,193-200.

[7] S. Visa and A. Ralescu. Learning Imbalanced and Overlapping Classes using Fuzzy Sets. In *Proceedings of the 20th ICML 2003 Workshop: Learning with Imbalanced Data Sets II*, pages 97-103, Washington, August, 2003.

[8] S. Visa, A. Ralescu, S. Yeung and A. Genaidy. A word is worth many numbers - Modeling the meaning of linguistic description of numerical data. In *Proceedings of the 13th Midwest Artificial Intelligence and Cognitive Science Conference (MAICS)*, April, 2002.

[9] S. Visa, A. Ralescu, S. Yeung and A. Genaidy. Linguistic Modeling of Physical Task Characteristics. In *Proceedings of the 9th Information Processing and Management of Uncertainty in Knowledge-based Systems Conference (IPMU)*, pages 1779-1786, July, 2002

[10] G. Wu and E. Chang. Class-Boundary Alignment for Imbalanced Dataset Learning. In *Proceedings of the 20th ICML 2003 Workshop: Learning with Imbalanced Data Sets II*, pages 49-56, Washington, August, 2003.

[11] *Proceedings of the 20th ICML 2003 Workshop: Learning with Imbalanced Data Sets II*, Washington, August, 2003.