Combining Dialectical Optimization and Gradient Descent Methods for Improving the Accuracy of Straight Line Segment Classifiers

XXIV SIBGRAPI - Conference on Graphics, Patterns and Images

Rosario A. Medina Rodriguez and Ronaldo Fumio Hashimoto

Institute of Mathematics and Statistics
IME – USP

August 28th to 31st, 2011

SIBGRAPI 2011

INTRODUCTION

- Pattern Recognition Problems (OCR, handwritten and face recognition, etc.).
- Recent classifier which uses Straight Line Segments on its definition, called *SLS classifier*[1].
- One important step to get a good results for classification is to find the optimal positions of the straight line segments given a training data set.

[1] "A NEW MACHINE LEARNING TECHNIQUE BASED ON STRAIGHT LINE SEGMENTS" IN ICMLA 2006

OBJECTIVE

- Combine the traditional gradient descent method (GD) with a novel evolutionary algorithm called Dialectical Optimization Method (DOM)[2] at the training phase to obtain the capability of escaping from local optimum.

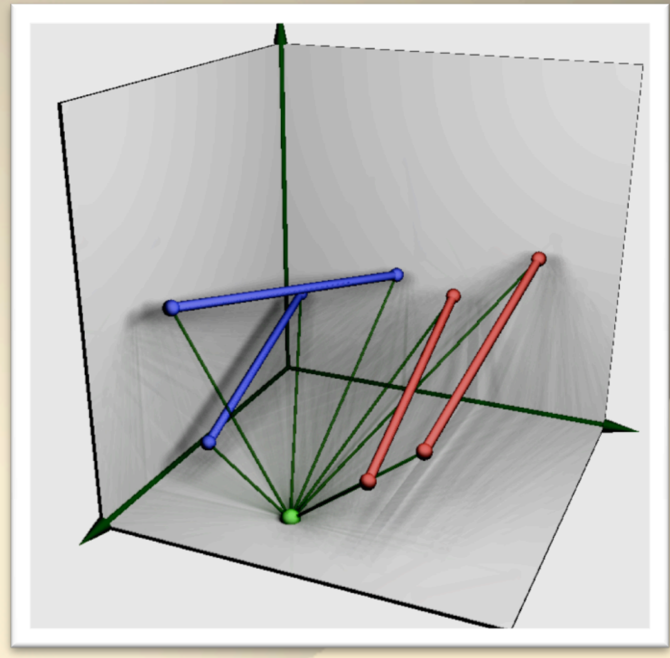
[2] "OPTIMIZATION BASED ON DIALECTICS" IN IJCNN, 2009

Outline

- Straight Line Segments Classifier.
- Dialectical Optimization Method.
- Hybrid of Dialectical Optimization and Gradient Descent Methods.
  - Adaptations to DOM concepts.
- Experimental Results.
- Conclusions

Classifier based on Straight Line Segments

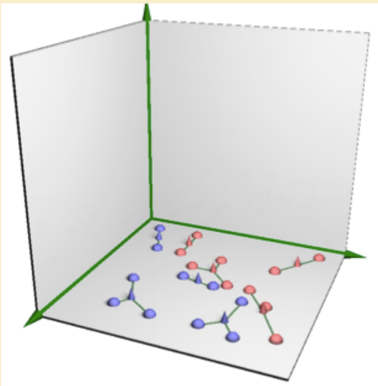
- Find two collections of SLSs such that the classification function minimizes a certain risk function.
- Distances between a set of points and two sets of straight line segments (SLSs)[3].

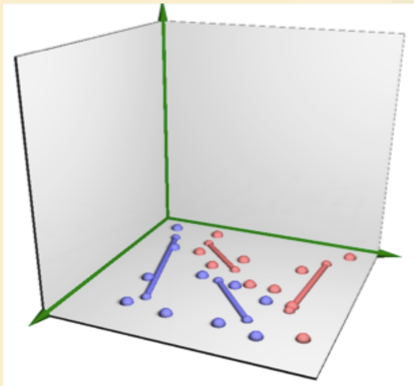


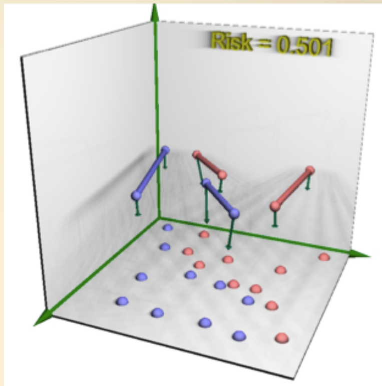
[3] A new training algorithm for pattern recognition technique based on straight line segments," in Computer Graphics and Image Processing, 2008.

Classifier based on Straight Line Segments – Training Phase

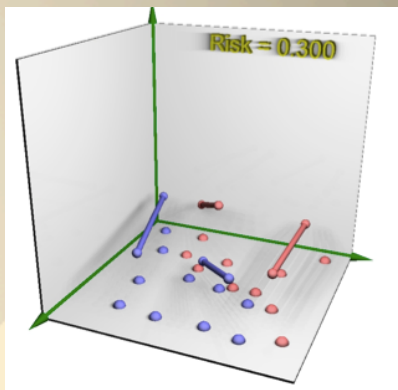
- **Placing (I,II):** Find the initial positions of the straight line segments using K-means clustering algorithm.
- **Tuning (III,IV):** Minimize the mean square error function, using gradient descent method.


(I)


(II)



(III)



(IV)

DIALECTICAL OPTIMIZATION METHOD

- Evolutionary method based on the materialist dialectics for solving search and optimization problems based on the dynamics of contradictions between their integrating dialectical poles.
- It has a lot of iterations and recombination process.

Dialectical Optimization Method

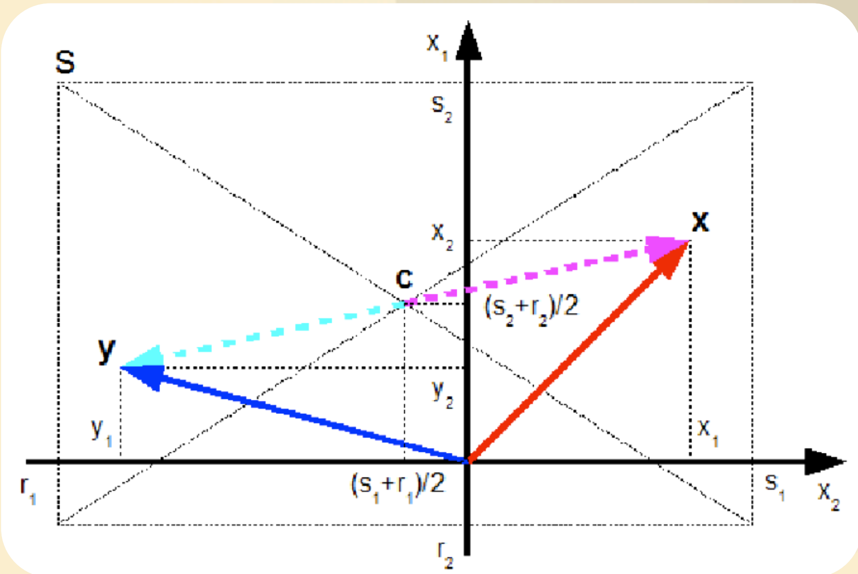
- **Pole:** corresponds a candidate solution to the problem.

- **Set of poles:** $\Omega = \{w_1, w_2, \ldots\ldots, w_m\}$, where each pole is defined as: $w_i = (w_{i,1}, w_{i,2}, \ldots., w_{i,n})^T$.

- **Social Force:** is associated to the objective function of the optimization problem, denoted by $f(w_i)$.

DIALECTICAL OPTIMIZATION METHOD

- **Contradiction:** Given two poles w_p and w_q is defined as: $\delta_{p,q} = dist(w_p, w_q)$
- **Antithesis:** $w\breve{}i = b - w_i + a$, where $a \le w_i \le b$, and $a, b \in R$.



Hybrid of DOM and GD Method

- The main goal of DOM is to assist the gradient descent method by providing to it a new set of initial positions (the output of the dialectical optimization method).

1. Generate initial poles.
2. For each phase of DOM, apply the gradient descent to each pole in the population in order to obtain one optimum local for each pole.
3. Proceed with the next steps of DOM.
4. If the number of phases is reached, stop.

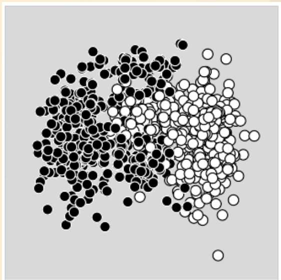
   Otherwise, return to step 2.

ADAPTATIONS TO DOM CONCEPTS

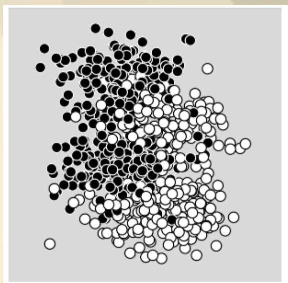
- **Pole:** Vector consisting of the extremities of the SLSs belonging to class 0 and class 1, such as $[L_0|L_1]$.
- **Antithesis:** Given a pole $[L_0|L_1]$, *the antithesis* is redefined as $[L_1|L_0]$.
- **Set of poles:** 50% randomly generated and the other 50% is generated with antithesis poles.
- **Social Force:** $$R_{E_n}(\mathcal{L}_0, \mathcal{L}_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - F_{\mathcal{L}_0,\mathcal{L}_1}(x_i))^2$$
- **Contradiction:** $| R_{En}(w_a) - R_{En}(w_b) |$

EXPERIMENTAL RESULTS

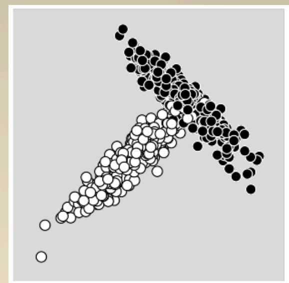
- **Artificial Data Sets**



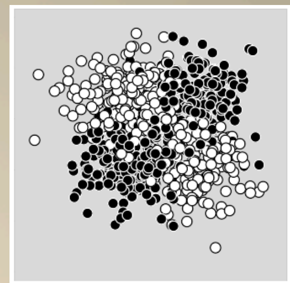
F-Dist



S-Dist



Simple-Dist



X-Dist

- It makes possible to apply the Bayes classifier.
- The probability density function is known, it is possible to use numerical integration to calculate the classification rate.

Experimental Results

- Number of Examples: 100 , 200, 400 and 800.
- Number of SLSs per class: 1, 2, 3 and 4
- Methods used at Training Phase:
  - Gradient Descent (GD)
  - Gradient with Genetic Algorithms (GD-AG)
  - Gradient Descent with Dialectical Optimization (GD-DOM)

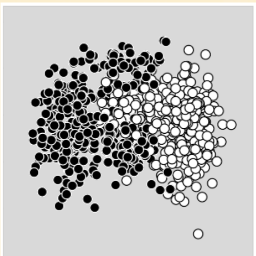
<u>Parameters – GD Method</u>

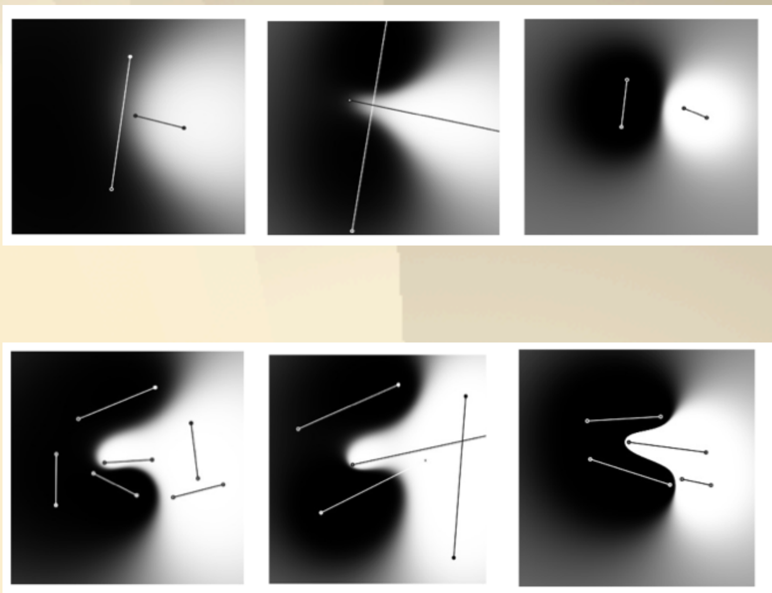
- Number of Iterations = 1000
- Initial Value = 0.1
- Displacement Increment = 0.1
- Displacement Decrement = 0.5
- Minimum Value = 10^{-5}

<u>Parameters – DO Method</u>

- Number of poles = 30,
- Number of phases = 20,
- Number of iterations = 15,
- Minimum Value=10^{-3} ,
- Learn Rate = 0.99,
- Crisis Effect value = 0.2.

Experimental Results



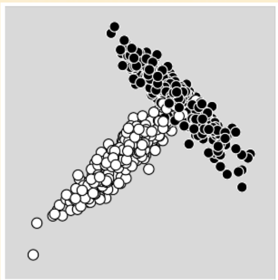


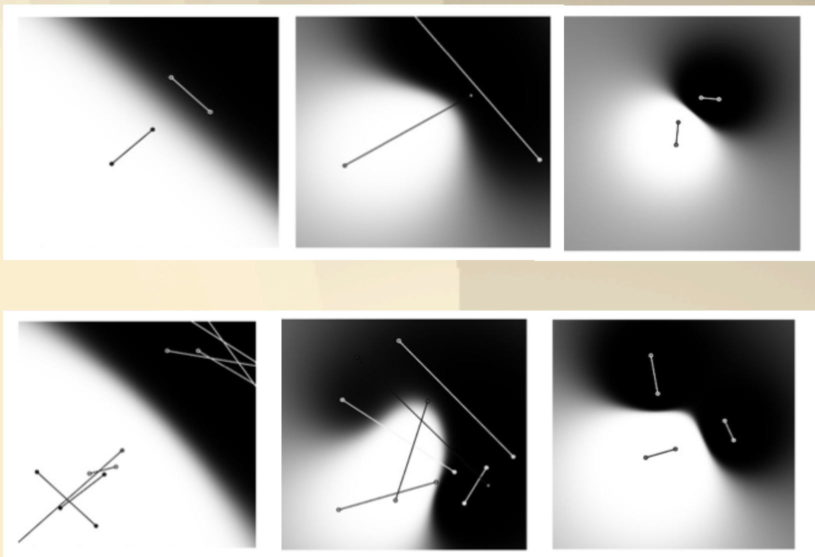
F- Distribution

	Examples	1 GD	1 GD-DOM	1 GD-AG	2 GD	2 GD-DOM	2 GD-AG	3 GD	3 GD-DOM	3 GD-AG	4 GD	4 GD-DOM	4 GD-AG
	100	82.89%	**83.32%**	59.94%	**85.99%**	84.84%	84.48%	**86.75%**	85.49%	73.63%	84.14%	**86.46%**	67.19%
F	200	75.02%	**86.24%**	73.29%	86.88%	**87.17%**	80.94%	**87.06%**	86.94%	81.76%	83.30%	**87.13%**	81.71%
(87.65%)	400	74.71%	**86.64%**	74.52%	**86.95%**	86.72%	85.02%	82.23%	**86.77%**	79.30%	83.95%	**86.51%**	75,04%
	800	76.86%	**84.69%**	76.77%	86.34%	**86.99%**	82.15%	86.71%	**87.14%**	75.70%	86.36%	**86.55%**	78.61%

Number of Straight Line Segments per class: 1, 2, 3, 4

EXPERIMENTAL RESULTS

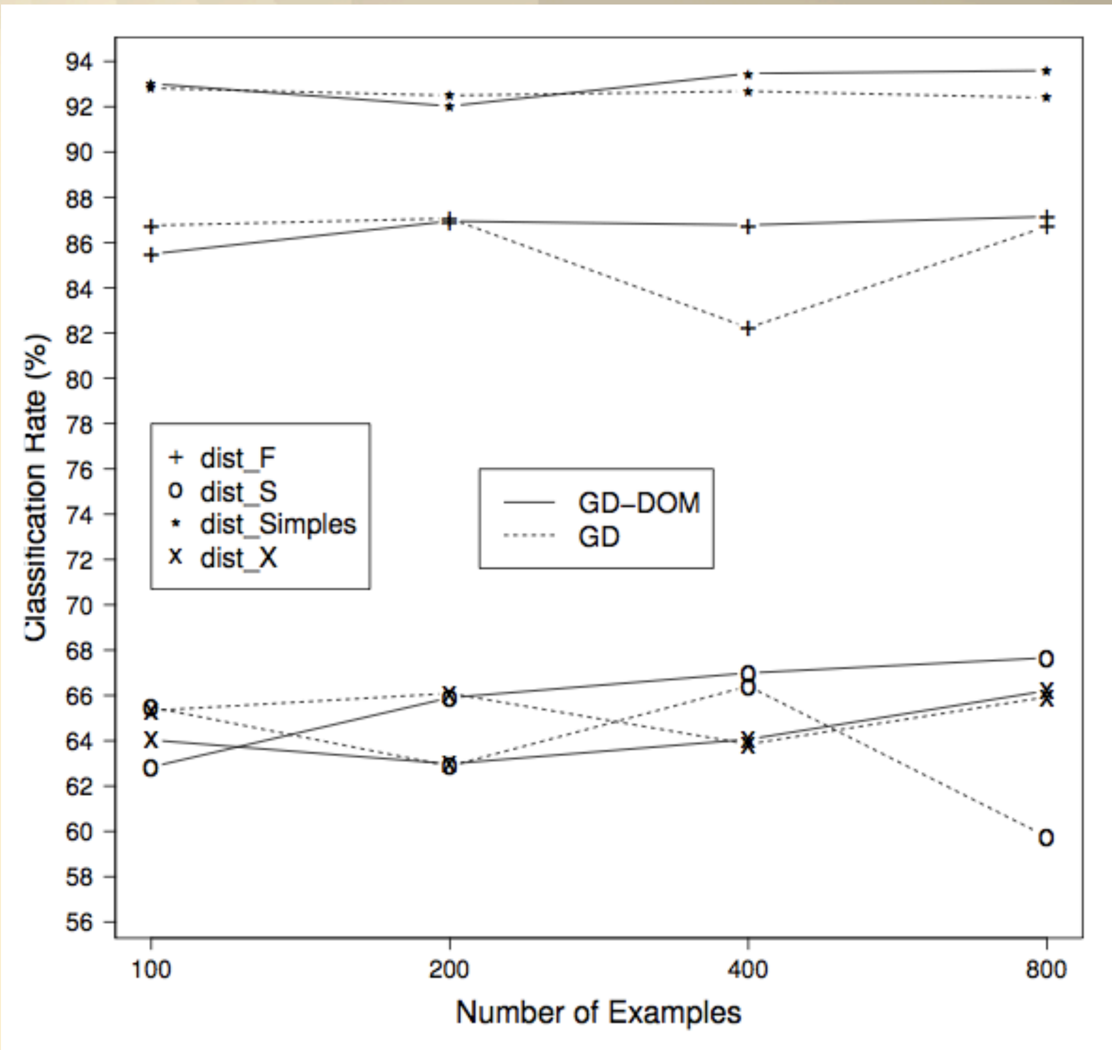




Simple - Distribution

		NUMBER OF STRAIGHT LINE SEGMENTS PER CLASS											
		1			2			3			4		
	EXAMPLES	GD	GD-DOM	GD-AG	GD	GD-DOM	GD-AG	GD	GD-DOM	GD-AG	GD	GD-DOM	GD-AG
SIMPLE (93.90%)	100	92.78%	**92.92%**	86.92%	92.78%	**92.99%**	91.98%	92.81%	**93.01%**	92.48%	92.78%	**92.96%**	88.92%
	200	92.51%	**92.66%**	89.30%	92.55%	**92.62%**	91.82%	**92.49%**	92.02%	92.22%	92.41%	**92.66%**	92.10%
	400	92.53%	**93.19%**	89.60%	92.61%	92.77%	**93.59%**	92.68%	**93.46%**	91.81%	92.68%	**93.45%**	93.08%
	800	91.75%	**93.50%**	90.02%	92.56%	**93.56%**	93.13%	92.39%	**93.59%**	93.04%	92.70%	**93.51%**	92.92%

Experimental Results



EXPERIMENTAL RESULTS

- **Public Data Set**
  - Breast Cancer Wisconsin (Diagnostic) Data Set.
  - 2 classes (B for Benign and M for Malign) and 10 attributes (features).

TABLE II

Classification Rate for Breast Cancer Data Set

Method	Number of SLSs per class			
	1	2	3	4
GradDesc	**96.78%**	96.92%	96.34%	**96.78%**
GradDesc- DOM	96.66%	**97.32%**	**96.99%**	96.03%

CONCLUSIONS

- Our main contribution is to improve the training phase (optimization Classification Rate (%) of SLSs positions).
- The classification rate was improved in an average of 2%.

Future Work

- While this method improves the classification rate, the computation time for the training algorithm increases. In addition it has been studied the use of threads on the implementation to reduce the training time.
- The presented results indicate that the SLS classifier using the proposed hybrid method can be potentially used in Computer Vision problems. We plan to do this analysis for future work and also extend the SLS binary classifier to a multiclass classifier.

Combining Dialectical Optimization and Gradient Descent Methods for Improving the Accuracy of Straight Line Segment Classifiers

XXIV Sibgrapi - Conference on Graphics, Patterns and Images

Rosario A. Medina Rodriguez and Ronaldo Fumio Hashimoto

Thank you!

Questions?

SIBGRAPI 2011

CNPq FAPESP