Reconhecimento de Gestos em Linguagem de Sinais

Filipi Xavier D. Braggio[1]

[1]PESC - IA
COPPE - UFRJ
filipixavier@cos.ufrj.br

Introdução

O reconhecimento automático de gestos humanos se mostra hoje como uma importante e desafiadora tarefa. Um exemplo de possível aplicação está na interpretação de linguagens de sinais. Outras aplicações incluem as novas possibilidades de interação de humanos com computadores e robôs.

Tal problema vem sendo muito atacado ultimamente, porém ainda apresenta grandes desafios. Uma das maiores dificuldades está no fato de que pessoas diferentes executam um mesmo gesto de maneiras muito distintas, além de as observações estarem sujeitas a variações nas características do ambiente, como iluminação por exemplo. Outro fato importante a se considerar é a similaridade que pode existir entre os gestos a se reconhecer, por exemplo ao distinguir um zig-zag e uma ondulação.

Neste estudo foi atacado especificamente o problema de reconhecimento de 15 diferentes gestos em LIBRAS (LInguagem BRAsileira de Sinais). A seção seguinte apresenta informações acerca da base de dados utilizada.

Base de dados

A base de dados utilizada foi o LIBRAS Movement Dataset disponibilizado por pesquisadores da Universidade de São Paulo. A base já foi utilizada anteriormente em outros estudos de reconhecimento de gestos (2). Possui 360 instâncias divididas em 15 classes correspondentes as 15 diferentes gestos. Cada classe possui 24 instâncias. Na base, cada instância representa uma única execução de um determinado gesto, e possui 91 atributos. Os atributos representam 45 pares de coordenadas horizontais e verticais, além do atributo de classe.

Cada instância foi obtida através de uma filmagem de 7 segundos da execução de um gesto. A partir de cada filmagem foram amostrados 45 quadros, para os quais foram calculadas as posições centrais horizontal e vertical da mão que executa o gesto. As coordenadas são representadas como valores reais entre 0.0 e 1.0, o atributo de classe é representado por um inteiro.

A seguir constam exemplos para cada uma das classes pertencentes à base. Para cada uma delas foi gerado um gráfico com os 45 pontos amostrados.

Pré-processamento

Ao longo do trabalho, foram experimentadas diversos formatos de entrada para alimentar os algoritmos. tais formatos são discutidos nas seções seguintes.

Formato original

Para fins de estudo, foi utilizado também o formato original dos dados, que consiste em um vetor de 90 números reais entre 0 e 1. As posições do vetor correspondem às coordenadas horizontal e vertical intercaladas.

Representação de transições

A primeira variação feita no formato de entrada corresponde à representação das 44 transições entre os 45 pontos amostrados. O formato de entrada passou a ser um vetor de 88 posições, sendo cada posição um valor real entre -2 e 2 correspondente ao deslocamento exercido nos eixos horizontal e vertical.

Representação de sentido de transições

A variação seguinte foi a substituição dos valores reais dos deslocamentos pelos valores -1, 0 e 1. Os novos valores passaram a representar somente o sentido da transição ou a não ocorrência dela. Para definir a ocorrência ou não de transição, foi utilizado um parâmetro chamado de limiar. Apenas as transições com módulo maior que este limiar fora consideradas de fato, e foram representadas por -1 ou 1 de acordo com o seu sentido. Transições com módulo abaixo do limiar foram representadas com 0.

O parâmetro limiar também foi variado na busca por melhores resultados. Foram utilizados diversos valores calculados com combinações dos valores da

média dos deslocamentos e do desvio padrão dos deslocamentos dentro de uma mesma instância.

Representação de transição em segmentos

Uma estratégia utilizada na representação dos dados foi a segmentação das instâncias em partes de tamanho igual. Essa estratégia visou reduzir a dimensão dos vetores de entrada. Foram utilizados para cada instância 9 segmentos correspondentes a 5 posições adjacentes cada um. Neste formato, as entradas passaram a contar com 18 valores reais, sendo para cada segmento um valor relativo ao deslocamento horizontal e outro relativo ao deslocamento vertical. Os deslocamentos fora calculados como a diferença entre as posições no primeiro e no último quadro de cada segmento.

Representação de sentidos de transição em segmentos

Foi utilizada uma variante do padrão anterior que substitui os valores reais das transições pelos valores -1, 0 e 1, representando apenas o sentido das transições ou a não ocorrência delas. De maneira análoga ao explicado na seção Representação de sentido de transições, foi utilizado um parâmetro limiar para definir a ocorrência ou não de uma transição.

Os efeitos do uso de cada um dos formatos de entrada são discutidos na seção Resultados.

Clusterização

Sobre a base de dados foi aplicada a ação de clusterizar, ou seja agrupar, instâncias semelhantes. O resultado desejado seria a separação em 15 grupos, cada um contendo as instâncias referentes a um mesmo gesto. Na busca pelos melhores resultados foram utilizadas diferentes técnicas e algoritmos conforme descritos nas seções seguintes. As implementações utilizadas foram aquelas disponibilizadas pelo framework java Encog (Enc).

K-Means

K-Means (4) é um método de agrupamento que visa organizar n entradas em k clusters de forma a minimizar a distância de cada um dos n elementos ao centro do cluster ao qual pertence. O algoritmo apresenta convergência garantida pela técnica de maximização da expectativa, em que para cada iteração o algoritmo reposiciona os centros dos clusters e redefine a pertinência dos elementos aos clusters. Neste problema, em nenhum cenário, foram necessárias mais que 50 iterações do algoritmo para atingir convergência.

Mapa auto-organizado

Os mapas auto-organizados são um modelo de redes neurais artificiais não supervisionadas. Tal modelo possui em sua camada de ativação neurônios que estão dispostos de forma que haja uma informação de vizinhança entre eles. Quando um padrão é apresentado como entrada, o neurônio com maior valor de ativação é dito o vencedor, e representará a saída da rede para este padrão.

No treinamento do mapa auto-organizado, a informação de qual neurônio é o vencedor é utilizada para a redefinição de pesos da rede. Esse método de treinamento chama-se aprendizado pro competição. A cada nova entrada apresentada, o mapa se atualiza de forma a melhor representar os padrões já apresentados.

Os mapas auto-organizados foram criados por Teuvo Kohonen em 1982 (3) por isso são comumente chamados mapas de Kohonen. São muito aplicados a problemas de clusterização e na redução de dimensionalidade de dados.

Neste trabalho foi utilizado um mapa com 15 neurônios na camada de saída, representando cada uma das 15 classes de gestos da base de dados.

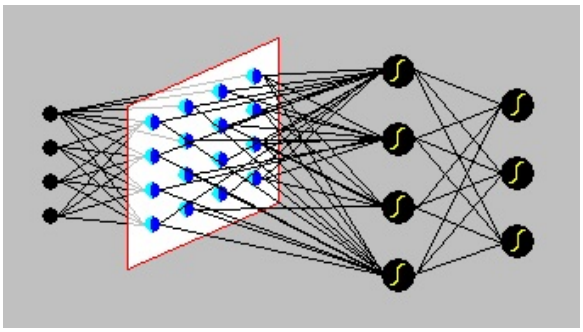

Figure 1: Rede de um mapa auto-organizado.

Treinamento

Todo o treinamento utilizado foi não supervisionado, sendo a quantidade de classes a única informação de classificação utilizada da base de dados para parametrizar os modelos. Ao longo do estudo foram experimentadas diferentes partições do banco de dados para treinamento. Em todas essas partições o número de instâncias por classe foi equilibrado.

Antes de cada treinamento, os dados originais foram pré-processados e ordenados de maneira a simular uma amostragem aleatória com probabilidades iguais para cada classe. Em seguida os dados pré-processados eram fornecidos como entrada aos dois algoritmos (K-Means e Mapa auto-organizado) com a mesma ordenação.

Teste

A avaliação de cada uma das técnicas de clusterização foi feita da seguinte forma: Foram fornecidas as entradas (sendo estas a base inteira ou uma partição desta) para treinamento do mapa e para a execução do algoritmo K-Means. Ao término dos processamentos, foram iteradas todas as instâncias existentes na base e foi registrada a classificação obtido em cada um dos métodos. Tendo as informações da classificação correta e da classificação obtida, foram montadas as tabelas cruzando tais informações.

Resultados

Nesta seção são apresentados e discutidos os resultados obtidos com cada uma das técnicas utilizadas. Para todas elas, as informações expostas referem-se ao que foi obtido utilizando como entrada para os algoritmos toda a base de dados.

K-means

Para alimentar o algoritmo K-Means foram utilizadas as seguintes formas de representação de dados após pré-processamento: dados originais, dados de 44 transições (diferenças entre pares de posições adjacentes), sentidos (-1, 0, e 1) dos movimentos nos eixos horizontal e vertical nas 44 transições, transições (diferenças entre posições) para 9 segmentos de 5 pontos cada, sentidos (-1, 0, e 1) dos movimentos em cada um dos 9 segmentos.

A representação que apresentou melhores resultados foi a representação original dos dados. Para este formato de entrada foram experimentados diferentes números de iterações do algoritmo K-Means. Foi notado que a partir de 50 iteração a convergência já havia ocorrido.

Foi tentado sem sucesso reduzir o número de clusters fornecido ao algoritmo. A intenção seria errar menos entre classes parecidas agrupando-as, para em um processamento futuro distinguir entre tais classes. Conforme foi diminuido o numero de classes, o agrupamento de classes parecidas não funcionou como esperado, e novas classes começavam a se misturar no novo modelo.

A tabela 1 a seguir apresenta o melhor resultado obtido através do K-Means, as colunas correspondem às classes originais e as linhas correspondem às classes encontradas como resposta. Toda a base foi utilizada, a representação dos dados é a original, foram solicitados 15 clusters diferentes.

Mapa auto-organizado

Assim como para o algoritmo K-Means, foram utilizadas as seguintes formas de representação de dados após pré-processamento: dados originais, dados de 44 transições (diferenças entre pares de posições adjacentes), sentidos (-1, 0, e 1) dos movimentos nos eixos horizontal e vertical nas 44 transições, transições (diferenças entre posições) para 9 segmentos de 5 pontos cada, sentidos (-1, 0, e 1) dos movimentos em cada um dos 9 segmentos.

A representação que obteve melhores resultados foi a representação dos sentidos de transição dos 9 segmentos. Para este formato foi variado o parâmetro limiar que define se houve (-1 ou 1) ou não houve (0) transição nas direções horizontal e vertical para um dado segmento. A primeira utilização do parâmetro constava de um valor único utilizado para as duas direções de todas as instâncias. Em seguida foi implementada uma forma que considerava um valor para cada instância, calculado como uma combinação da média e do desvio padrão do módulo das transições de uma mesma instância. Adiante foi implementado o cálculo de um limiar para a direção horizontal e outro para a direção vertical, considerando média e desvio padrão das transições nessas direções. Por fim, a forma utilizada, e que obteve os melhores resultados, considera para ambas as direções o maior entre os limiares horizontal e vertical, sendo cada limiar $0,5$ vezes a média dos módulos das transições em cada uma das direções.

Adicionalmente no estudo do mapa auto-organizado foi feita a variação do parâmetro chamado taxa de aprendizado, que consiste em um valor entre 0 e 1 que regula a intensidade dos ajustes feitos na rede a cada nova entrada inserida. O comportamento da clusterização se mostrou bastante estável para uma grande faixa de valores da taxa de aprendizado. Para valores entre $0,2$ e $0,7$ o comportamento foi virtualmente o mesmo, estando qualquer diferença dentro das flutuações causadas pela própria variação da ordenação das entradas para o algoritmo. Para valores fora dessa faixa, o mapa acabou colapsando mais de 85% das entradas em um mesmo cluster.

Assim como no caso do K-means, foi tentado sem sucesso reduzir o número de clusters. Finalmente, os resultados apresentados na tabela 2 dizem respeito à seguinte configuração: representação de direção de transições em 9 segmentos, limiar calculado separadamente por instância e por direção como $0,5$ vezes a média dos módulos das transições (prevalecendo o maior entre os limiares horizontal e vertical), taxa de aprendizado de $0,5$, 15 clusters. As colunas correspondem às classes originais e as linhas correspondem às classes encontradas como resposta. Toda a base de dados foi utilizada no mapa.

Classes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	2	0	0	0	0	12	0	0	11	0	17	0	3	0
1	0	0	0	0	0	0	3	0	0	0	0	1	0	0	5
2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	3	0	0	0	6	4
4	10	7	14	7	0	7	6	2	13	6	0	6	0	0	1
5	0	0	0	0	0	0	0	12	0	0	17	0	16	0	0
6	0	0	0	15	0	9	0	0	0	0	0	0	0	0	0
7	0	0	0	1	16	0	0	0	0	0	0	0	0	0	0
8	6	6	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	3	0	0	6	0	7	0	1
10	0	0	0	0	0	0	0	0	0	4	0	0	0	8	7
11	0	0	0	0	1	8	0	0	0	0	0	0	0	6	0
12	0	1	9	0	2	0	3	3	10	0	1	0	1	1	0
13	8	8	0	1	5	0	0	0	0	0	0	0	0	0	0
14	0	0	1	0	0	0	0	3	0	0	0	0	0	0	6

Table 1: Clusterização K-means

Classes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	6	7	0	0	0	2	0	0	3	0	2	0	4	0	0
1	0	0	0	17	0	10	0	0	1	2	0	0	0	0	0
2	0	0	0	0	17	0	0	0	0	0	0	0	0	1	0
3	0	0	12	0	0	1	0	0	1	2	0	7	0	0	0
4	0	0	0	3	0	0	20	0	6	6	0	6	0	0	4
5	0	0	0	0	0	0	0	0	1	0	9	0	7	0	0
6	16	14	0	0	0	0	0	0	2	2	3	0	3	0	0
7	0	0	0	0	0	0	0	16	3	0	5	0	8	0	0
8	0	0	0	0	7	0	0	0	0	0	0	0	0	0	12
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
10	2	3	1	0	0	0	0	0	1	0	5	0	2	0	0
11	0	0	0	0	0	0	4	8	4	4	0	0	0	3	7
12	0	0	0	2	0	3	0	0	0	7	0	0	0	6	0
13	0	0	11	1	0	0	0	0	1	1	0	11	0	0	0
14	0	0	0	1	0	8	0	0	0	0	0	0	0	14	1

Table 2: Clusterização Mapa auto-organizado

Conclusões

Comparando as duas tabelas, é possíve ver que o mapa auto-organizado obteve uma performance ligeiramente superior à do K-Means. Nota-se uma maior distribuição das instâncias entre as classes e um isolamento um pouco maior entre classes, sugerindo uma menor distorção intra-classe no mapa auto-organizado. No K-Means foi obtido o maior cluster (4) com quase 80 instâncias. No mapa auto-organizado foi obtida a classe (9) de menor população, com 1 instância somente. Em ambas as tabelas, fica explícita a proximidade entre algumas classes. São elas: balanço curvado(1), balanço horizontal(2), balanço vertical(3); movimento anti-horário(4), círculo(6); zig-zag horizontal(10), onda horizontal(12); zig-zag vertical(11), onda vertical(13).

Notou-se durante o estudo que o mapa auto-organizado se beneficiou da simplificação de representação dos dados (reduzindo dimensão e discretizando valores), e o mesmo não ocorreu para o K-Means, que teve seu melhor desempenho em cima da representação original dos dados. Ficou evidente que para alguns modelos de redes neurais não supervisionadas o pré-processamento dos dados é fundamental para o sucesso da classificação, assim qualquer conhecimento sobre o problema deve ser aplicado no sentido de simplificar a representação dos dados.

O estudo mostrou as dificuldades relacionadas ao reconhecimento automático de gestos. As instâncias podem variar dramaticamente de acordo com o sujeito que executa o gesto. Além disso existem uma grande proximidade entre algumas classes, como o caso dos zig-zagues e das ondas, comumente fazendo que as instâncias dessas classes se agrupem.

Trabalhos futuros

Após os métodos de pré-processamento, foi possível identificar um certo efeito que traz impactos negativos e que deveria ser contornado. A representação de algumas instâncias de uma mesma classe era praticamente idêntica (na representação de sentidos de transição), entretanto podia observar-se que os dados pareciam estar transladados em algumas dessas instâncias. Tal fenômeno ocorre devido a variações no tempo de início dos movimentos nas capturas dos gestos. Algumas instâncias começavam o movimento mais adiantadas ou atrasadas em relação a outras. O impacto dessa diferença nos algoritmos é grande, pois fragiliza a comparação de componentes que eventualmente ocorre em cada técnica.

Uma possibilidade para contornar esse problema seria considerar a transformada de Fourier das entradas, dado o caráter de invariância dessa transformada com relação ao tempo. Outra possibilidade seria aplicar uma técnica de processamento de sinais onde cada entrada fosse comparada a um padrão correspondente a cada classe, e fosse utilizada sua forma transladada que apresentasse maior verossimilhança com algum dos padrões considerados.

References

Encog machine learning framework. http://www.heatonresearch.com/encog.

Daniel B. Dias, Renata C. B. Madeo, T. R. H. H. B. S. M. P. (2009). Hand Movement Recognition for Brazilian Sign Language: A Study Using Distance-Based Neural Networks. [In international joint conference on neural networks, atlanta, ga.].

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. [In biological cybernetics].

MacQueen, J. (1967). Some Methods for classification and Analysis of Multivariate Observations. [In proceedings of 5th berkeley symposium on mathematical statistics and probability].

Anexos

Gráficos das Instâncias

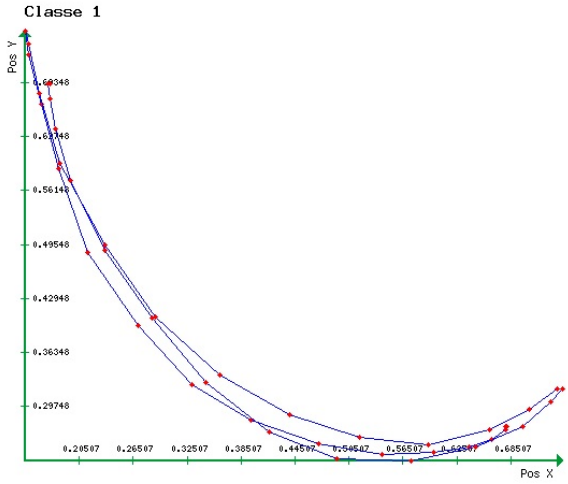

Figure 2: Balanço curvo.

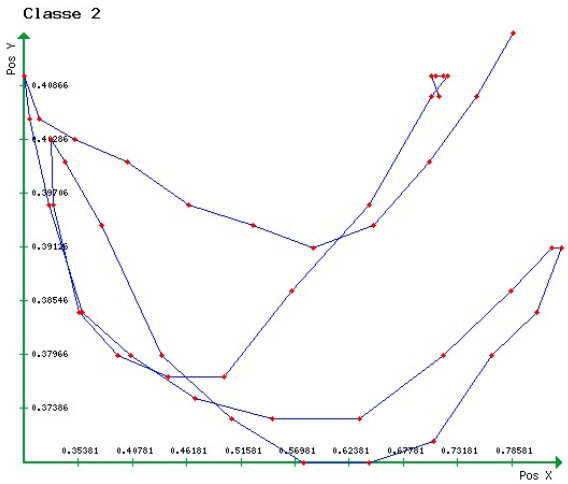

Figure 3: Balanço horizontal.

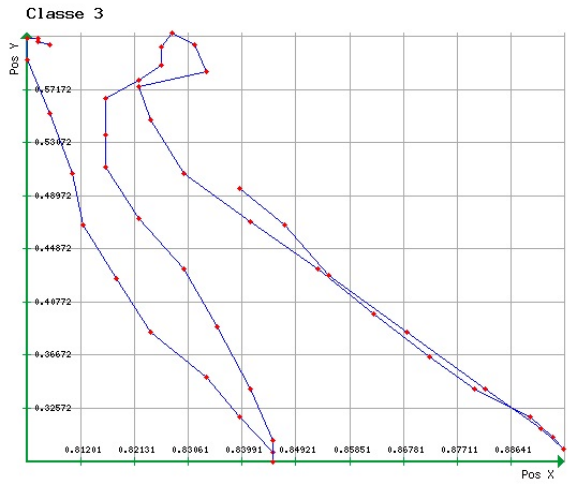

Figure 4: Balanço vertical.

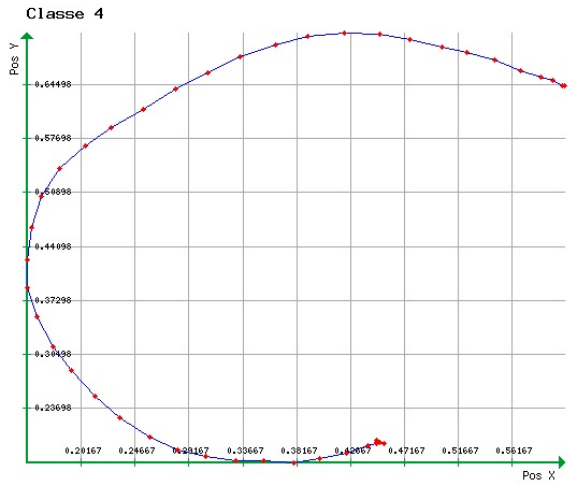

Figure 5: Movimento anti-horário.

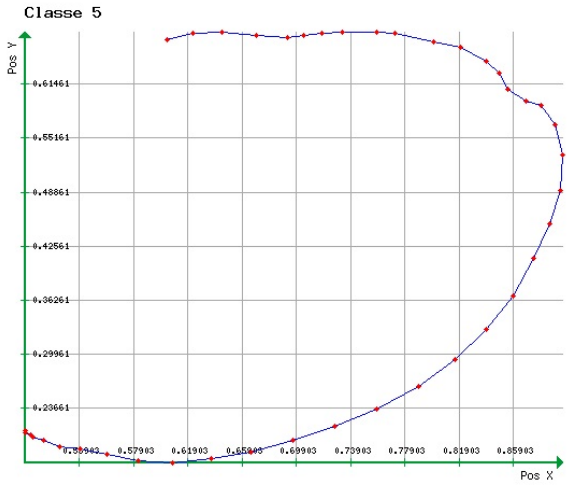

Figure 6: Movimento horário.

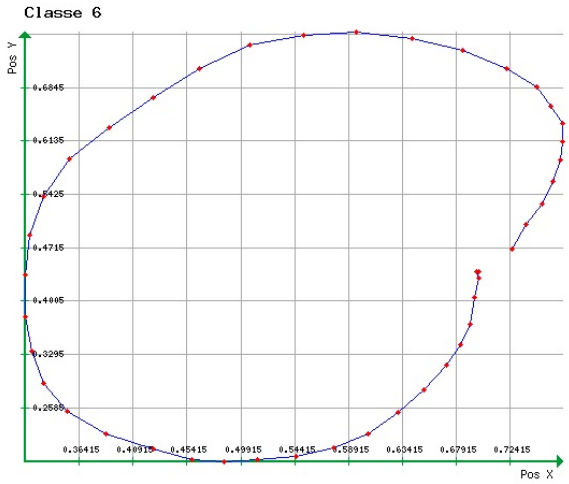

Figure 7: Círculo.

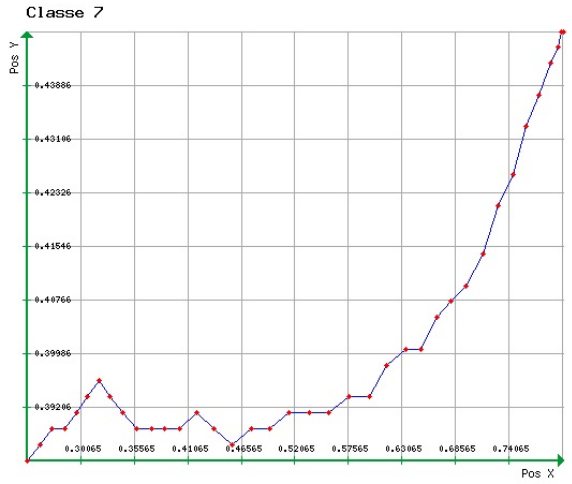

Figure 8: Linha horizontal.

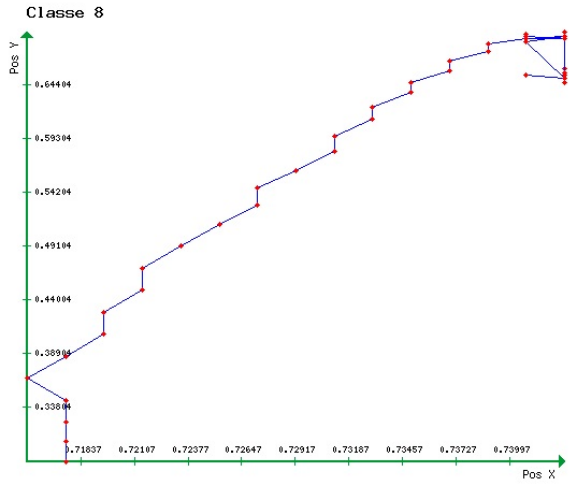

Figure 9: Linha vertical.

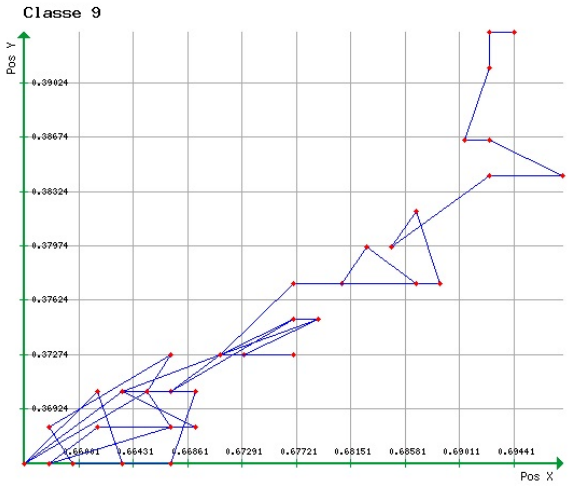

Figure 10: Tremer.

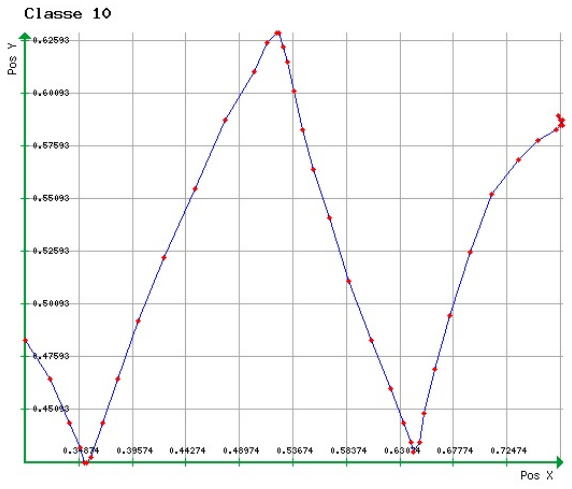

Figure 11: Zig-zag horizontal.

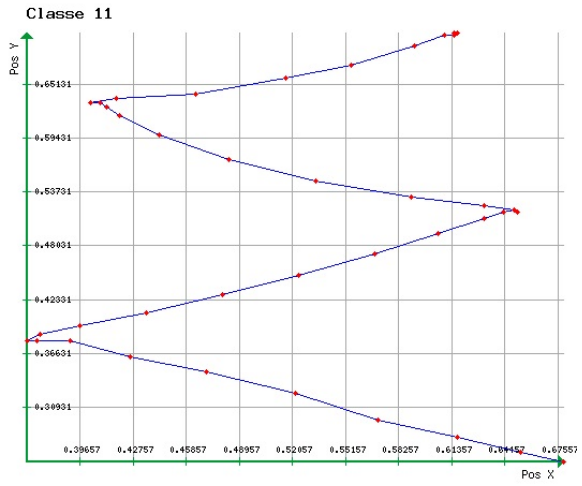

Figure 12: Zig-zag vertical.

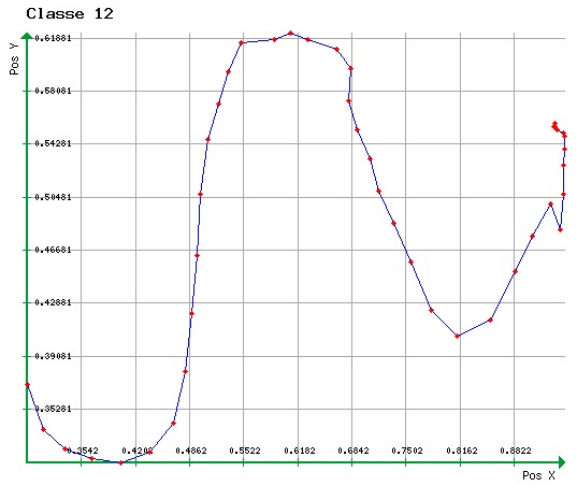

Figure 13: Ondulação horizontal.

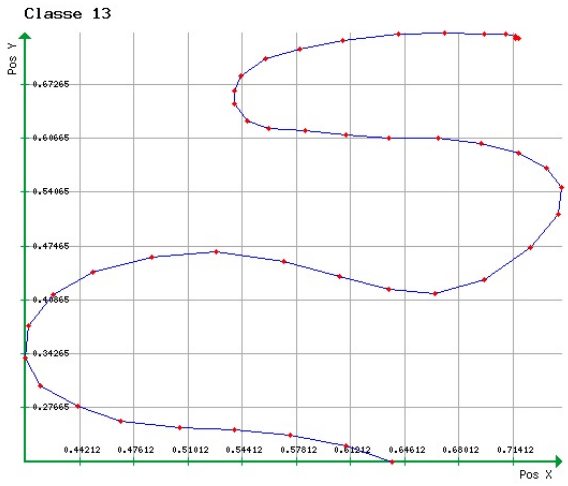

Figure 14: Ondulação vertical.

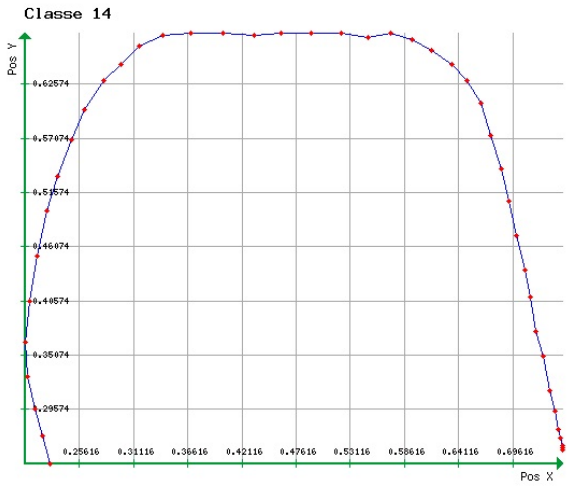

Figure 15: Parábola para cima.

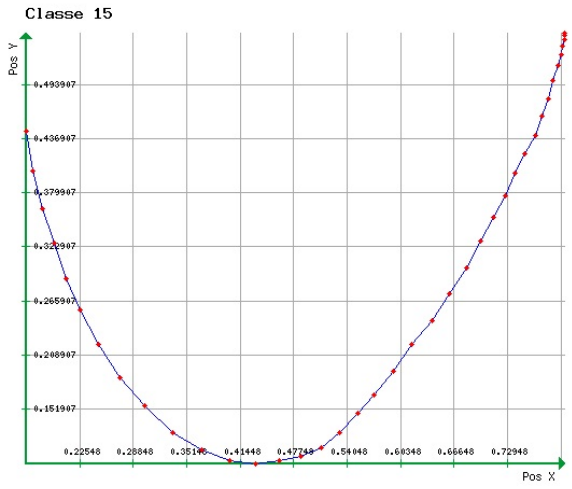

Figure 16: Parábola para baixo.