

Aplicação de classificadores na base NIST

Thiago Alexandre Nakao França

Universidade Tecnológica Federal do Paraná (UTFPR)
Campo Mourão – PR – Brasil

nakaosensei@gmail.com

Abstract. *The focus of this paper was to classify the 26 letters from the english alphabet using classifiers from artificial intelligence, to train these classifiers, it was used the images provided by NIST database, we used the classifiers such K-NN, SVM(Support Vector Machine) and Decision Tress. The train set was made by a total of 37440 instances, while the test set had just 11941 instances, we used the fusion from the 3 classifiers and we got the general accuracy of x%. Is is important to say that the characteristics extraction and normalization were made by the student Thiago Nakao, making use of gnu octave, the classifiers used came from the software Weka and Lib svm.*

Resumo. *O objetivo deste trabalho foi realizar a classificação das 26 letras do alfabeto inglês fazendo o uso de classificadores de inteligência artificial, para treinar esses classificadores, foi usada a base de imagens fornecida pelo NIST, foram usados os classificadores K-NN, SVM(Support Vector Machine) e árvore de decisão. O conjunto de treino era composto por 37440 instâncias, enquanto o de teste tinha apenas 11941 instâncias, usando a fusão dos 3 classificadores, a acurácia geral foi de x%. É importante ressaltar que a extração de características e normalização das mesmas da imagem foi realizada pelo acadêmico Thiago Nakao, fazendo uso do octave, quanto aos classificadores, foram usados os do software Weka unido ao lib svm.*

1. Introdução

Este foi um trabalho acadêmico do curso de Ciência da Computação da UTFPR de Campo Mourão, seu objetivo era fazer o uso de diferentes classificadores de IA para classificar um conjunto de imagens de letras do alfabeto inglês, as imagens foram obtidas pela base do NIST.

Os classificadores utilizados foram o KNN, SVM e Árvore de decisão, que são classificadores bastante utilizados em diversos trabalhos até os dias atuais.

A extração de características das imagens foi implementada em octave pelo aluno Thiago Alexandre Nakao França, basicamente houveram dois testes principais, um com um zoneamento das imagens em nove quadrantes, e outro teste com o zoneamento da imagem em 16 quadrantes, a heurística utilizada foi contagem de pixels pretos nas matrizes das imagens. Explicarei em maior profundidade esse processo nas seções seguintes.

A classificação ocorreu fazendo o uso de cada um dos classificadores separadamente, e também com o uso da fusão deles através do voto e máximo, foram usados os algoritmos do software Weka, com exceção do SVM que foi executado nele, mas na realidade de tratava do pacote lib-svm por trás.

Por fim, tivemos os resultados dos classificadores das extrações de contagem de pixel com 9 e 16 zoneamentos, descobrimos que com 16 zoneamentos a acurácia foi superior. No entanto, os resultados estarão presentes nas próximas seções.

2. Extração de Características

Para o treinamento dos classificadores, era preciso extrair as características das imagens do conjunto de treino e teste, no conjunto de treino haviam 37440 imagens, enquanto no de teste existiam 11941, todas essas imagens tiveram de passar pelo processo de extração de características.

A característica utilizada nesse experimento foi simplesmente a contagem de pixels pretos aliada ao zoneamento da matriz da imagem, foram realizados experimentos com zoneamento de 9 e 16 seções em todos os classificadores para efeito de comparação.

Basicamente, foram implementadas duas rotinas bastante semelhantes, essas rotinas recebiam como entrada caminho do arquivo indicando quais imagens pertencem ao conjunto em questão, e como saída gerava um arquivo contendo as características normalizadas daquela imagem, o que diferencia uma rotina da outra é que uma faz o zoneamento de 9 características, enquanto a outra faz o zoneamento para 16 características.

Foi evidente que nem sempre receberíamos imagens simétricas para realizar a extração, nesses casos, pode ocorrer de uma linha ou coluna de zoneamento seja um pouco maior que as demais, se tivéssemos imagens muito pequenas no conjunto, o problema poderia ser ainda maior.

Segue um exemplo da lógica empregada no zoneamento, no caso, para 16 zonas:

- 1) Armazene o número de linhas de uma imagem em uma variável e o número de colunas em outra

2) Divida esses numeros por 4 e salve ambos, o resultado da divisão das linhas será o acrescimo de cada zona no eixo y, enquanto o resultado da divisao das colunas sera o acrescimo de cada zona no eixo x.

3) Salve a informação de “limite” de cada nivel para o eixo x e o y, por exemplo, em uma imagem 30x40, dividimos 30/4 e temos o valor 7 para as linhas, isso significa que os elementos entre as linhas 1 e 7 estarao na primeira zona das linhas, da 2 a 14 na segunda, da 15 a 21 na terceira e da 21 a 30 na quarta, fazemos o mesmo processo para as colunas, assim, temos quatro variaveis que informam o limite dos quadrantes das linhas, e quatro que informam o limite das colunas.

4) Agora basta percorrer as linhas de cada quadrante separadamente, com base nos indices, e contar os pixels.

Uma vez que as características foram extraidas, ainda se mantendo no exemplo de 16 zonas, teremos como resultado um array de 17 posições, vindo primeiro as 16 características seguida da classe, o nome da classe é obtido através do nome da pasta que a imagem se encontra.

3. Normalização

Os dados provenientes da extração de características não poderiam ser passados para os classificadores de forma “crua”, pois os tamanhos das imagens era muito variado, por isso, após a contagem de pixel, foi realizada a normalização sobre cada uma das 16(ou 9) características, utilizamos minmax, onde cada valor recebe o proprio subtraido do menor valor do conjunto, dividido pelo maximo subtraido do minimo.

4. Experimentos e Classificação

Após a extração/normalização dos dados, finalmente fomos para a etapa de classificação, lembrando que os resultados foram obtidos fazendo o uso do software Weka, iniciarei expondo os resultados dos experimentos fazendo o uso do zoneamento de 9 quadrantes.

4.1 Resultados classificação com 9 Zonas.

Seguem as matrizes de confusão e as taxas:

```
a b c d e f g h i j k l m n o p q r s t u v w x y z <- classified as
261 66 0 0 7 6 6 7 0 0 0 28 0 61 0 0 1 2 4 0 0 0 0 5 3 2 0 | a=a
3 290 0 7 5 1 27 0 0 0 12 0 17 0 0 0 3 37 0 0 0 0 1 17 5 1 1 | b=b
0 0 402 1 22 0 10 0 0 0 2 56 0 0 0 10 1 0 0 0 0 0 0 0 0 14 | c=c
0 62 0 261 0 0 22 0 0 0 0 0 1 0 2 4 33 0 3 0 2 0 3 0 1 2 | d=d
0 95 3 0 190 1 11 1 0 0 30 3 0 0 0 4 2 1 1 0 0 0 19 2 1 1 | e=e
1 0 2 1 10 89 0 1 0 0 2 4 6 0 0 0 293 4 0 0 6 0 1 1 2 3 1 | f=f
0 57 5 17 9 1 147 2 0 0 2 4 6 0 0 0 2 9 1 0 0 10 0 117 0 0 | g=g
16 34 0 0 0 1 1 254 0 0 7 1 39 5 0 0 1 1 0 0 2 3 36 0 1 0 | h=h
3 4 10 7 86 9 0 0 192 29 3 2 1 2 0 0 0 6 114 5 1 5 0 76 78 182 | i=i
0 2 0 80 7 0 0 0 0 5 237 1 0 2 0 1 2 9 0 15 4 0 0 0 0 0 61 | j=j
1 12 0 38 0 3 2 0 0 0 269 6 9 0 0 0 0 5 0 0 0 3 21 6 0 0 | k=k
0 0 0 0 1 0 0 0 0 0 1 481 0 0 0 0 0 0 0 0 7 0 3 0 0 3 | l=l
49 52 0 0 0 0 6 49 0 0 5 0 189 0 1 24 7 12 0 0 1 3 58 0 4 0 | m=m
9 118 0 0 0 1 14 15 0 0 6 0 26 23 0 5 13 1 0 0 28 5 172 1 2 0 | n=n
0 23 2 229 1 0 19 0 0 1 0 0 0 1 57 1 128 0 0 0 4 1 0 0 1 0 | o=o
1 11 0 12 0 9 2 0 0 0 0 0 0 4 0 0 406 3 0 0 0 1 1 0 1 16 0 | p=p
0 129 0 28 5 1 60 0 0 0 0 0 8 2 2 1 186 3 2 0 5 1 7 1 1 2 | q=q
7 85 3 2 51 6 8 1 0 0 190 1 22 1 0 4 4 44 2 0 0 0 2 9 0 4 | r=r
1 123 2 85 7 0 10 0 5 13 6 1 10 0 1 0 30 7 78 0 0 0 2 5 3 56 | s=s
2 0 0 1 0 6 0 1 0 0 0 0 12 0 0 8 2 0 1 422 0 0 0 1 11 2 | t=t
0 12 0 6 1 0 6 2 0 0 7 8 2 1 1 1 10 0 0 343 10 47 1 0 0 | u=u
0 14 0 0 3 0 0 0 0 0 13 0 7 20 0 2 19 0 0 0 66 258 26 3 51 0 | v=v
2 41 0 1 0 0 1 4 0 0 7 2 3 8 0 2 3 0 0 0 39 2 359 1 0 0 | w=w
6 3 0 0 0 3 0 1 0 0 38 0 1 2 0 0 0 5 1 0 0 7 15 362 26 2 | x=x
3 5 0 1 1 5 0 5 0 0 8 0 24 1 0 34 0 0 0 4 0 26 1 13 322 0 | y=y
0 35 5 8 12 1 0 0 0 13 8 2 2 0 0 0 8 1 4 0 0 0 3 11 5 349 | z=z
```

Imagem 01 – Matriz SVM

```
=== Confusion Matrix ===
a b c d e f g h i j k l m n o p q r s t u v w x y z <- classified as
359 37 0 4 2 3 9 12 1 0 9 0 5 0 0 1 3 4 0 0 0 0 3 6 0 1 | a=a
16 320 1 10 4 0 11 1 0 1 3 0 10 1 0 1 28 20 1 0 1 0 5 0 0 1 | b=b
0 0 317 7 30 5 6 0 0 0 14 88 0 0 1 10 1 3 0 1 0 0 0 0 35 | c=c
3 31 2 331 1 0 5 0 0 0 0 0 1 0 0 0 20 0 0 0 1 0 0 0 1 | d=d
0 80 2 3 225 5 2 0 0 0 5 4 1 0 0 2 1 20 1 1 0 0 8 1 1 3 | e=e
0 1 0 2 11 143 1 1 0 0 2 1 3 0 0 241 1 1 0 8 0 1 0 0 2 0 | f=f
8 79 5 33 14 0 165 2 0 0 2 3 4 3 2 0 24 0 0 0 4 0 41 0 0 0 | g=g
19 21 0 1 1 0 2 313 0 0 4 1 4 9 0 1 3 1 1 0 2 1 17 0 1 0 | h=h
6 14 6 10 102 18 0 0 319 51 2 5 1 0 1 12 0 17 67 7 0 4 1 16 13 143 | i=i
0 4 0 42 8 0 1 0 4 304 0 0 0 0 0 1 10 0 12 6 0 0 0 0 7 27 | j=j
14 10 1 3 36 3 1 0 0 0 254 7 3 1 0 0 0 17 0 0 1 3 17 6 0 0 | k=k
0 0 1 0 2 0 2 0 0 0 1 483 0 0 0 0 0 0 0 0 0 1 1 0 0 5 | l=l
51 57 1 2 0 0 3 117 0 0 1 0 117 19 1 17 6 44 0 0 2 0 21 1 0 0 | m=m
14 74 0 4 5 1 16 49 0 0 4 0 14 162 0 2 9 7 0 0 15 5 58 0 0 0 | n=n
7 30 9 314 0 2 8 0 0 0 0 0 1 0 33 0 51 0 0 0 4 0 0 0 0 0 | o=o
1 10 0 18 1 8 0 1 0 0 3 0 11 2 0 403 4 1 0 0 0 1 1 0 2 0 | p=p
0 106 1 27 4 0 45 1 0 0 0 0 5 1 2 3 223 8 4 0 12 1 7 1 0 | q=q
13 102 1 3 56 4 4 1 0 0 82 2 16 1 0 7 5 139 0 0 0 0 2 6 0 2 | r=r
3 98 4 31 5 0 30 0 2 17 4 2 4 1 1 0 22 7 181 0 0 0 0 6 3 24 | s=s
3 1 0 0 1 3 0 0 0 13 0 0 1 0 0 8 0 1 0 429 0 0 0 0 9 0 | t=t
0 7 0 5 2 0 4 3 0 0 5 1 3 9 0 0 3 0 1 0 360 12 43 0 0 | u=u
0 0 0 0 1 1 0 0 0 1 6 0 0 5 2 0 0 0 0 76 327 25 3 35 0 | v=v
0 22 0 2 0 0 4 6 0 0 6 0 6 20 0 2 3 3 1 0 22 3 375 0 0 | w=w
6 8 0 1 1 3 1 3 1 1 44 1 3 0 0 0 3 1 5 0 0 4 8 360 16 2 | x=x
3 4 0 0 1 8 0 3 0 0 2 0 6 2 0 30 2 0 0 4 1 45 2 12 328 0 | y=y
0 47 3 17 12 1 3 0 1 0 1 2 0 0 0 0 5 2 2 0 0 0 8 4 0 359 | z=z
```

Imagem 02 – Matriz KNN

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	<- classified as
302	30	0	3	1	4	13	12	0	3	7	1	29	3	0	0	4	31	0	0	3	0	5	5	0	3	a=a
47	163	0	10	9	0	47	10	0	2	7	0	28	1	0	2	38	14	3	0	1	1	46	1	0	5	b=b
0	0	332	4	13	5	9	0	0	0	0	49	65	0	0	1	7	4	6	4	7	0	0	0	1	0	11 c=c
6	21	12	262	3	1	13	0	0	1	0	0	0	1	11	7	37	0	0	0	17	0	1	0	1	2	d=d
4	62	6	0	134	10	26	1	0	2	47	2	7	0	0	3	9	29	1	0	8	0	6	0	1	7	e=e
1	3	1	0	8	134	5	0	0	3	1	0	3	3	0	210	11	8	0	9	0	1	6	3	9	0	f=f
12	55	4	39	10	1	158	1	0	4	3	2	5	1	2	1	37	5	1	0	11	3	31	1	0	2	g=g
20	19	0	0	0	0	3	252	0	2	6	1	33	22	0	2	2	6	2	0	5	1	24	1	1	0	h=h
13	15	9	13	59	33	3	2	209	33	17	26	3	5	2	10	0	25	66	12	12	15	0	39	60	134 i=i	
2	6	1	19	8	5	2	0	11	241	1	0	0	0	2	2	4	0	22	7	0	0	0	13	49	31	j=j
9	6	9	1	31	1	9	7	0	0	216	2	3	5	0	0	2	30	0	0	13	2	10	11	1	9	k=k
3	0	9	1	2	0	3	0	1	0	18	444	1	0	1	0	0	1	0	0	2	0	2	4	0	4	l=l
41	33	2	4	1	1	1	94	0	0	13	0	137	30	0	4	9	27	0	0	2	1	60	0	0	0	m=m
18	42	0	1	2	1	14	52	0	0	3	0	29	109	0	2	9	7	0	0	25	1	124	0	0	0	n=n
10	11	33	280	1	0	5	1	0	1	0	0	0	0	51	0	48	0	0	0	17	0	1	0	0	0	o=o
5	9	1	10	2	14	4	3	0	0	2	0	4	4	2	376	15	0	0	1	3	5	4	0	3	0	p=p
21	80	5	23	8	0	57	1	0	0	6	0	7	2	4	7	191	6	4	0	12	1	16	0	1	0	q=q
16	62	4	1	23	4	17	10	0	1	63	3	25	4	0	4	19	106	2	0	15	0	10	23	0	4	r=r
7	80	2	53	17	8	26	4	6	32	5	4	12	2	1	6	18	11	63	0	3	0	10	10	4	41	s=s
0	1	0	1	3	9	0	1	2	34	3	3	1	0	0	8	0	3	1	375	0	0	0	0	21	3	t=t
1	6	0	2	0	0	1	8	0	0	3	4	1	17	0	0	6	1	1	0	354	11	41	0	1	0	u=u
0	5	0	0	4	2	0	10	0	0	10	0	1	19	0	1	1	0	2	0	64	239	47	7	70	0	v=v
5	14	0	1	0	0	2	10	0	0	9	1	15	67	0	1	2	0	3	0	31	5	308	0	1	0	w=w
9	9	1	0	3	5	3	2	1	7	57	11	5	4	0	1	0	15	2	0	3	8	11	292	10	13	x=x
1	16	0	1	4	3	0	1	0	5	4	4	1	3	0	18	0	2	1	13	2	31	5	13	325	0	y=y
3	46	6	10	12	1	7	2	2	8	4	12	1	1	0	6	4	14	9	3	3	0	5	45	1	262	z=z

Imagem 03 – Matriz Dtree

Total de instâncias conjunto de teste: 11941

Acurácia:

SVM(Kernel Linear): 54.1914 %, acertou 6471 instâncias.

Knn(K=5):61.3768 %, acertou 7329 instâncias.

Arvore de decisão: 50.7076 %, acertou 6055 instâncias.

Voto: 59.2831 %, acertou 7079 instâncias.

4.2 Resultados classificação com 16 Zonas.

Seguem as matrizes de confusão e as taxas:

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	<- classified as	
353	6	0	3	0	1	4	8	0	0	21	0	36	0	0	0	8	4	0	0	0	0	1	13	1	0	a=a	
5	306	0	6	0	0	6	0	0	2	0	1	10	0	0	1	11	1	0	0	1	0	1	1	2	1	b=b	
0	1	456	1	29	1	11	0	1	0	0	11	0	0	1	1	1	1	0	0	0	0	0	0	0	3	c=c	
0	51	0	311	0	0	0	0	0	2	0	0	2	0	0	7	6	11	0	0	0	3	0	2	0	0	1	d=d
0	60	8	1	222	7	20	0	2	0	3	3	1	0	0	1	4	3	0	0	1	0	0	1	0	28	e=e	
0	1	0	0	2	268	1	0	0	1	1	2	0	0	94	2	0	0	43	0	0	0	0	0	1	3	f=f	
3	56	2	12	7	2	225	0	0	0	3	2	17	0	7	0	14	0	0	0	18	0	21	0	0	0	g=g	
50	6	0	0	0	1	5	204	0	0	3	0	110	1	0	0	2	1	0	0	1	2	10	5	1	0	h=h	
4	7	4	6	49	17	0	0	335	58	0	6	1	0	1	1	0	6	2	9	0	2	0	77	82	148	i=i	
0	16	0	27	2	0	1	5	6	335	0	0	0	0	0	0	3	0	8	0	0	0	0	0	0	20	j=j	
13	2	2	0	20	15	5	3	0	0	225	6	7	0	0	0	1	45	0	0	3	3	6	21	0	0	k=k	
1	0	10	0	5	0	4	2	0	1	0	457	0	0	0	0	0	1	0	0	5	0	6	0	0	4	l=l	
47	10	0	0	0	1	2	80	0	0	8	0	267	0	0	7	0	2	0	0	1	0	32	1	2	0	m=m	
23	17	0	4	0	2	1	109	0	0	7	0	29	110	0	0	4	4	0	0	2	5	119	2	1	0	n=n	
1	27	3	194	0	1	2	0	0	6	0	0	2	0	140	0	81	0	0	0	2	0	0	0	0	0	o=o	
4	2	0	13	0	14	0	0	0	0	0	0	1	0	1	411	14	1	0	0	0	1	0	1	4	0	p=p	
2	28	0	20	4	3	32	0	0	1	2	0	2	1	2	0	328	10	0	0	12	0	0	0	1	0	4	q=q
17	5	3	0	18	26	3	0	0	29	3	10	0	0	4	24	281	0	0	1	0	1	0	8	1	12	r=r	
0	52	0	25	3	2	19	0	1	27	0	1	5	0	0	0	5	1	292	1	0	0	0	0	1	10	s=s	
2	0	0	0	0	3	0	0	1	0	0	6	0	0	4	0	0	0	442	0	0	0	0	0	0	9	2	t=t
6	16	0	21	1	0	9	9	0	1	1	1	2	0	9	0	7	0	0	0	334	21	19	1	0	0	u=u	
1	2	0	0	1	4	0	10	1	1	6	0	1	6	0	0	0	1	0	0	49	297	41	20	41	0	v=v	
1	10	0	2	0	0	1	47	0	1	8	1	7	13	0	0	8	0	0	19	3	351	3	0	0	0	w=w	
14	0	0	0	2	8	0	3	3	0	23	0	0	0	0	0	0	1	8	0	0	0	2	3	377	26	2	x=x
1	4	0	0	0	10	0	7	0	0	3	0	4	1	0	4	0	0	0	10	0	25	3	14	367	0	y=y	
1	29	0	6	10	0	0	0	0	16	1	0	0	0	0	0	0	2	0	1	0	0	0	1	6	394	z=z	

Imagem 05 – Matriz SVM

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	<- classified as	
326	43	0	1	4	6	7	6	0	0	24	0	21	0	0	1	2	8	0	0	0	0	5	4	1	0	a=a	
6	392	0	10	5	1	24	1	0	0	10	0	15	0	0	2	32	12	0	0	0	0	10	3	1	1	b=b	
0	0	378	4	18	0	6	0	0	0	8	71	0	0	0	10	2	1	0	2	0	0	0	0	0	18	c=c	
0	33	0	306	0	0	14	0	0	0	0	0	0	0	0	5	32	0	0	0	0	2	0	1	0	1	d=d	
0	85	2	0	207	2	14	1	0	0	26	3	0	0	0	3	2	4	0	0	0	0	11	3	1	1	e=e	
1	0	2	1	6	119	0	1	0	0	1	0	1	0	0	271	5	1	0	6	0	0	1	1	1	1	f=f	
1	68	5	24	6	1	185	1	0	0	2	3	7	0	0	2	13	0	0	0	10	2	59	0	0	0	g=g	
20	17	0	0	0	0	3	291	0	0	5	1	23	5	0	0	2	3	0	0	3	3	26	0	0	0	h=h	
4	7	9	10	85	18	0	0	0	228	35	6	6	3	1	0	3	0	7	101	6	1	5	0	60	47	173 i=i	
0	2	0	47	8	0	0	0	0	1	295	1	0	0	0	1	0	11	0	13	7	0	0	0	0	7	33	j=j
6	2	6	1	35	1	4	2	0	0	280	5	5	0	0	0	0	6	0	0	0	2	16	6	0	0	k=k	
0	0	1	0	1	0	0	0	0	2	483	0	0	0	0	0	0	6	0	1	0	0	1	0	0	2	l=l	
48	47	0	2	0	0	3	79	0	0	6	0																

=== Confusion Matrix ===

```

a b c d e f g h i j k l m n o p q r s t u v w x y z <-- classified as
302 0 0 5 0 4 6 8 0 1 6 2 54 5 1 2 4 39 0 0 0 0 3 4 0 5 | a=a
8 292 6 10 11 3 18 9 1 0 2 0 14 0 1 0 43 0 1 0 1 0 8 0 1 6 | b=b
0 3 419 1 32 5 9 0 1 4 3 29 0 0 0 1 1 2 0 2 2 0 0 0 0 4 | c=c
1 35 4 267 1 2 8 1 0 1 0 0 2 1 24 2 41 2 0 0 0 1 3 0 0 0 | d=d
1 34 31 1 213 4 19 0 1 3 5 1 1 0 1 1 5 17 4 0 0 0 2 2 0 19 | e=e
0 0 1 1 4 259 5 0 1 21 0 1 7 0 0 66 9 2 1 7 0 18 6 2 6 2 | f=f
9 52 15 29 9 7 193 1 0 0 4 3 11 0 8 0 31 2 1 0 2 0 9 1 0 2 | g=g
12 6 1 1 0 0 3 269 0 1 12 0 49 13 0 0 3 1 0 0 5 4 29 2 0 0 | h=h
3 22 13 7 37 50 3 1 326 72 3 19 15 0 5 3 3 10 22 6 2 19 0 27 61 86 | i=i
0 5 2 5 2 2 3 1 5 349 1 0 0 0 3 1 1 0 26 5 0 2 0 2 7 4 | j=j
9 6 3 0 27 7 7 7 1 0 222 5 11 2 0 0 1 24 0 0 6 0 6 24 0 9 | k=k
1 1 23 0 3 0 2 0 0 2 1 440 0 1 0 0 0 1 1 0 6 0 1 0 0 13 | l=l
39 2 0 3 0 3 4 86 0 0 11 1 252 15 0 2 4 10 0 0 1 1 23 3 0 0 | m=m
19 4 1 1 1 1 0 92 0 2 2 0 35 181 0 2 6 1 0 0 4 2 84 0 0 1 | n=n
1 42 15 155 0 1 10 0 0 0 0 0 1 1 155 0 71 0 1 0 4 0 2 0 0 0 | o=o
6 4 1 10 1 39 8 0 0 1 0 0 15 0 1 350 14 0 0 10 0 3 3 0 1 0 | p=p
2 30 3 11 2 2 32 3 0 2 6 0 12 2 7 1 321 7 1 0 5 0 2 0 0 1 | q=q
15 13 5 1 35 6 3 6 0 6 22 11 20 0 0 1 38 229 1 0 3 0 4 7 0 20 | r=r
0 40 0 6 7 2 7 0 16 52 1 0 2 0 1 1 7 2 270 1 0 0 0 5 20 5 | s=s
0 0 0 0 1 9 0 0 0 16 1 0 0 0 0 1 13 1 0 0 415 0 0 0 1 10 1 | t=t
1 2 0 7 2 1 1 13 0 0 5 0 4 4 4 1 6 0 0 0 345 24 38 0 0 0 | u=u
1 0 0 0 0 2 4 1 1 3 1 0 5 4 2 1 0 3 0 0 23 320 64 9 38 0 | v=v
3 1 1 3 0 0 4 21 0 0 0 0 16 25 1 0 3 1 0 0 13 8 375 0 0 0 | w=w
9 3 0 0 4 5 3 4 2 7 48 7 10 0 1 1 0 10 3 2 3 7 3 323 5 12 | x=x
2 1 0 1 3 5 0 3 0 9 3 0 8 2 0 6 3 0 1 27 4 68 7 28 272 0 | y=y
3 19 1 5 15 2 4 0 5 8 6 2 1 0 0 1 0 2 3 1 1 0 2 1 0 385 | z=z

```

Imagem 03 – Matriz Dtree

=== Confusion Matrix ===

```

a b c d e f g h i j k l m n o p q r s t u v w x y z <-- classified as
386 6 0 1 0 2 2 5 0 0 9 0 28 0 0 0 3 6 0 0 0 0 1 9 0 1 | a=a
4 395 0 5 0 0 5 0 0 0 0 0 1 11 0 0 1 10 1 0 0 1 0 0 0 1 | b=b
0 1 461 1 18 4 9 0 1 0 0 20 0 0 0 1 1 1 0 0 0 0 0 0 0 | c=c
0 34 1 334 0 0 1 0 0 2 0 0 2 0 4 6 8 0 0 0 1 1 2 0 0 0 | d=d
0 47 8 1 260 5 10 0 2 0 3 2 0 0 0 1 3 4 1 0 1 0 0 1 0 16 | e=e
0 1 0 0 2 288 1 0 0 2 0 0 2 0 0 90 1 0 0 27 0 0 0 0 2 3 | f=f
3 53 1 14 7 3 247 0 0 0 2 2 14 0 8 0 15 0 0 0 5 0 15 0 0 0 | g=g
21 3 0 0 0 1 1 293 0 0 3 0 57 1 0 0 2 1 0 0 4 0 11 4 0 0 | h=h
4 10 7 6 36 14 1 0 447 69 0 9 7 0 1 0 2 4 4 7 0 1 0 28 66 92 | i=i
0 14 0 6 2 1 1 0 6 372 0 0 0 0 0 0 3 0 8 6 0 0 0 1 0 6 | j=j
10 1 2 0 17 10 6 3 0 0 279 3 6 0 0 0 0 18 0 0 0 2 4 15 0 1 | k=k
1 0 10 0 2 0 4 2 0 1 0 462 0 0 0 0 0 1 0 0 3 0 4 0 0 6 | l=l
36 8 0 0 0 0 2 95 0 0 5 0 288 2 0 3 0 1 0 0 1 0 19 0 0 0 | m=m
19 13 0 2 1 0 0 104 0 0 4 0 25 169 0 0 1 3 0 0 3 1 92 2 0 0 | n=n
0 25 6 194 0 2 3 0 0 3 0 0 2 0 151 0 70 0 0 0 3 0 0 0 0 0 | o=o
3 2 0 13 1 13 0 0 0 0 0 0 0 0 0 0 421 12 0 0 0 0 0 0 2 0 | p=p
1 23 1 11 2 3 21 0 0 0 0 0 0 0 0 3 0 369 7 0 0 9 0 0 0 2 | q=q
17 5 4 0 18 18 3 0 0 0 14 2 13 0 0 4 21 304 0 0 2 0 0 7 1 13 | r=r
0 44 0 14 2 1 13 0 1 20 1 0 2 0 0 0 3 1 334 1 0 0 0 1 5 2 | s=s
1 0 0 0 0 3 0 0 0 4 0 0 1 0 0 6 0 0 444 0 0 0 0 10 0 | t=t
2 6 0 7 1 0 3 11 0 0 1 1 1 0 4 0 4 0 0 379 24 14 0 0 0 | u=u
0 1 0 0 1 0 0 1 1 1 2 0 3 2 0 0 0 1 0 0 37 364 33 7 28 0 | v=v
1 1 0 2 0 0 2 26 0 0 3 0 9 11 0 0 3 0 0 0 13 3 401 0 0 0 | w=w
11 0 0 0 2 4 0 2 2 0 26 0 1 0 0 0 1 4 0 0 0 2 3 397 15 2 | x=x
1 3 0 0 1 3 0 2 0 0 3 0 4 0 0 4 0 0 12 0 37 5 10 368 0 | y=y
2 25 0 5 4 0 0 0 0 4 1 0 0 0 0 0 0 2 0 1 0 0 0 1 3 419 | z=z

```

Imagem 04 – Matriz Voto

Total de instâncias conjunto de teste: 11941

Acurácia:

SVM(Kernel Linear): 68.403 %, acertou 8168 instâncias.

Knn(K=5): 78.1593 %, acertou 9333 instâncias.

Arvore de decisão: 64.7768 %, acertou 7735 instâncias.

Voto: 75.6386 %, acertou 9032 instâncias.

Conclusão

Foi visível que no primeiro experimento, a combinação de classificadores apresentou resultados inferiores ao do KNN com k=5, isso se deve ao fato de outros dois classificadores com resultados piores estarem sendo levados em consideração.

Adicionalmente, foi possível notar que o uso de 16 zonas(características) na extração de imagens aumentou consideravelmente a porcentagem de acerto para a característica de contagem de pixels para todos os classificadores, no entanto, os resultados obtidos pela fusão de classificadores ainda foram inferiores ao KNN.