Violence Recognition in Spanish Words using Data Mining

Adolfo Flores Moreno, Silvia B. González-Brambila and Juan G. Vargas-Rubio
Dívisión de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Azcapotzalco, México, D.F., México

Keywords: Violence, Classification, Data Mining.

Abstract: Violent behavior in our society has been studied from many points of view, yet many cause-effect relations remain unexplained. Security personnel are normally trained to be alert and recognize potential violent behavior, but they cannot be 100% effective in recognizing it due to the monotonous nature of their job. This paper presents the first results of a work in progress detecting violence from the analysis of words in conversations. We used a set of videos with two person conversations in Spanish and classified them as violent and non violent. The audio of the conversations was extracted and converted to text. We used "Ward", "K-means" and "PAM" (clValid, 2014) to group words, performing a clValid analysis we found that the hierarchical technique was the best. The percentages of frequency were computed for each term and the SVM (Meyer, 2014) technique was applied, from which we found that there were unclassifiable terms. In three of the tests the prediction was erroneous and in another three we obtained good predictions with respect to the test set.

1 INTRODUCTION

One of the reasons for surveillance is to keep order and safety in public and private places. The use of video cameras is necessary, but the personnel in charge of watching the images may get bored and miss some events.

Violence is a deliberate behavior with the intention of causing physical or psychological injury to other person, and it may also lead to property destruction. Violence is generally associated with aggressiveness, but the second is not always destructive (Villanueva et al, 2007).

Violence is hard to define due to its ambiguity and subjectivity (Derbas & Quénot, 2014). In technical definitions violence is defined by visual-auditory indicators, like high-speed movements (Gong et al, 2008) and words.

There are many factors that lead to violence, bad mood, frustration, substance abuse, prescriptions, also social and environmental factors like family situations, job instability, friend circle, etc. that can contribute to it (RStudio, 2014). Violence can appear in any situation and consequences can be physical and psychological, generating behaviors and unwanted situations.

This work shows how standard video surveillance can be improved by automatically triggering some alarms when a conversation that is monitored has a high probability of being violent. When this happens some actions can be taken to avoid violence.

Most of the work related to violence is dedicated to detect violent scenes in videos. Violent scene detection (VSD) is an important research problem and promises several applications like movie/film inspection, video on demand, semantic video indexing and retrieval. Recent works are using low-level and mid-level features to represent violence concept (Lam et al, 2013) or visual and spatio-temporal features (Derbas & Quénot, 2014). However, at present, classification and filtering is done manually (Fujii & Yoshimura, 2011).

The rest of this paper is organized as follows. We describe the work in Section 2. Section 3 includes detailed information of the experiments, results and our analysis. Finally, conclusions and future work are presented in Section 4.

2 DESCRIPTION OF THE WORK

For this work, we collected different video segments where a two person conversation was taking place, and they were classified as violent or nonviolent. The audio track was extracted and converted to text.

Flores Moreno A., González-Brambila S. and Vargas-Rubio J..
Violence Recognition in Spanish Words using Data Mining.
DOI: 10.5220/0005072502100216
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2014), pages 210-216
ISBN: 978-989-758-048-2

The text files were preprocessed to build a term matrix. The frequency of each word was computed and presented in a graph and in a word cloud. The grouping results were validated using "Ward", "K-means" and "PAM" with *clValid*, and we used SVM for the classification.

2.1 Collecting Videos

We collected 100 video segments from different sources in the internet in the formats mp4 y wmv. The segments contain conversations of two persons in Spanish that were classified manually as violent or non violent (Tables 1 and 2).

Table 1: Collected video segments.

Type	Qty.
Violent	53
Non violent	47
Total	100

Table 2: Duration of the videos.

	Minutes
Minimum	1:12
Maximum	8:39

2.2 Audio Extraction and Processing

Once the video segments were classified, the audio was extracted in mp3 format using "Adobe Premiere Pro CC" (Adobe, 2014). Using the same program, the audio was converted to text with the Speech Analysis Model for Spanish. The results were saved in a .txt file.

There were some problems with the conversion. The conversion of each audio segment lasted from 3 to 10 minutes, but it was not always successful. Some of the conversations required manual transcription because the model could not identify some words correctly. The number of files successfully converted by Adobe was 83 while the other 17 had to be converted manually.

2.3 Pre-Processing

The file "stopwords.txt" that contains all 617 meaningless words like prepositions, articles and conjunctions was created to filter the texts. Also the gender was removed manually from some words to obtain better results.

The following steps were used to pre-process the text files for both, violent and non violent conversations:

- Load the file.
- Build the corpus.
- Convert to lowercase.
- Eliminate spaces.
- Eliminate punctuation signs
- Load the file with meaningless words (stopwords.txt).
- Remove generic words (used in R).
- Remove meaningless words (from "stopwords.txt" and other meaningless words).
- Build the term matrix.

Table 3 shows how the number of terms is reduced due to the pre-processing. The 100 more frequent words for violent and non violent conversations are shown in Figures 1 to 4.

Table 3: Number of terms before and after pre-processing.

Type of conversations	Terms before Pre-processing	Terms after Pre-processing
Non violent	1971	1549
Violent	1388	1133
Total	3359	2682

From Figures 1 to 4 we observe that there are some words that appear frequently in both, violent and non violent conversations. The difference is that they are preceded by different words in each case. In the case of violent conversations, those preceding words are normally "bad words".

2.4 Clustering

To validate the best grouping using the clustering methods "Ward", "K-means" and "PAM" we used the *connectivity*, *silhouette and Dunn* index. Figure 5 shows the hierarchical grouping with 8 clusters is the best choice and it is validated with Dunn for the file with all the conversations.

Using the clValid package from R we determined that 8 groups is the best solution, and it is shown in the Dendrogram of Figure 6.

From the grouping generated by "k-means" we observe that in each cluster we have the words that appear frequently together in conversations. In clusters 1, 2, and 7 we have the words used in violent conversations, and in the others are the words used in non violent conversations (see Fig. 7).

casa	quiero	vas	quieres	miedo	crees	diciendo	digo	hombre	vida	dinero	favor	importa	marido	pienso	siento
25	23	22	17	14	11	11	11	11	10	9	9	9	8	8	8
entiendes	loc	persona	querias	querid	cansad	dijiste	dime	entiendo	forma	gusta	hij	hijs	mama	mira	pensar
7	7	7	7	7	6	6	6	6	6	6	6	6	6	6	6
pense	puedes	vete	dando	dices	digas	dios	escucha	gente	hiciste	iba	maldit	pasa	paso	ven	amig
6	6	6	5	5	5	5	5	5	5	5	5	5	5	5	4
cara	caso	chic	cobrar	diablos	entender	familia	fotos	hablando	hablar	hice	jugar	llevo	manana	morir	mujer
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
mundo	padre	pasar	problema	queda	queria	real	salir	senora	vale	vivir	aburrid	acabo	alguien	amor	apoyo
4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3
ayuda	bailar	bano	boca	buscando	cadaver	callate	carta	contrato	cosa	culpa	cumplir	das	debo	decirle	decirme
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
deja	dejado	dificil	diga												
3	3	3	3												

Figure 1: The most frequent 100 words in non violent conversations.

maldit	pendej	dinero	favor	quieres	hij	quiero	vas	casa	put	madre	cabron	mierda
25	22	20	18	14	13	13	13	12	12	11	10	10
perr	coge	dijiste	importa	hey	pasa	viej	amig	crees	diciendo	digo	dije	estupid
10	9	9	9	8	8	8	7	7	7	7	7	7
habias	hablar	puedes	adios	callate	dame	familia	hablando	loc	mama	oye	papa	pinche
7	7	7	6	6	6	6	6	6	6	6	6	6
vale	vete	vida	amigs	cono	gente	herman	hombre	juro	mira	necesito	oficial	ofrece
6	6	6	5	5	5	5	5	5	5	5	5	5
pregunta	problema	problemas	ves	ahorita	alguien	camino	carajo	chinga	comer	culpa	dices	digas
5	5	5	5	4	4	4	4	4	4	4	4	4
doy	escuchar	facil	hagas	huy	idiota	imaginas	imbecil	mano	mire	mujer	palabra	pasando
4	4	4	4	4	4	4	4	4	4	4	4	4
perdon	personas	pudrete	senor	siquiera	toques	trasero	trato	amor	baja	borrach	calmate	cara
4	4	4	4	4	4	4	4	3	3	3	3	3
carcel	caso	cochin	come	cosa	cura	dejame	deje	desgraciad				
3	3	3	3	3	3	3	3	3				

Figure 3: The most frequent 100 words in violent conversations.

Figure 2: A cloud of the 100 more frequent words for non violent conversations.

Figure 4: A cloud of the 100 more frequent words for violent conversations.

```
Clustering Methods:
 hierarchical kmeans pam

Cluster sizes:
 2 3 4 5 6 7 8 9 10

Validation Measures:
                                 2        3        4        5        6        7        8        9       10

hierarchical Connectivity   2.9290   7.7869  10.7159  17.8361  18.5028  21.4317  24.3607  25.3607  26.3607
             Dunn           0.4749   0.4362   0.4862   0.5049   0.5049   0.5133   0.5586   0.5586   0.5586
             Silhouette     0.8218   0.8211   0.8156   0.8068   0.8033   0.7810   0.7720   0.7714   0.7713
kmeans       Connectivity  14.9071  16.6933  16.8361  21.6940  27.5520  36.9048  37.5714  34.6425  35.6425
             Dunn           0.4309   0.3369   0.3369   0.2139   0.2013   0.1833   0.1833   0.2357   0.2357
             Silhouette     0.8148   0.8029   0.7810   0.7631   0.7541   0.6432   0.6431   0.6432   0.6431
pam          Connectivity  84.7679 115.2321 175.2448 214.1349 237.7694 260.3389 288.7841 302.1587 316.8964
             Dunn           0.0693   0.0693   0.0690   0.0735   0.0786   0.0786   0.0786   0.0786   0.0786
             Silhouette     0.1555   0.0569   0.0784   0.0960   0.1141   0.1292   0.1437   0.1580   0.1670

Optimal Scores:

             Score  Method       Clusters
Connectivity 2.9290 hierarchical 2
Dunn         0.5586 hierarchical 8
Silhouette   0.8218 hierarchical 2
```

Figure 5: Grouping validation.

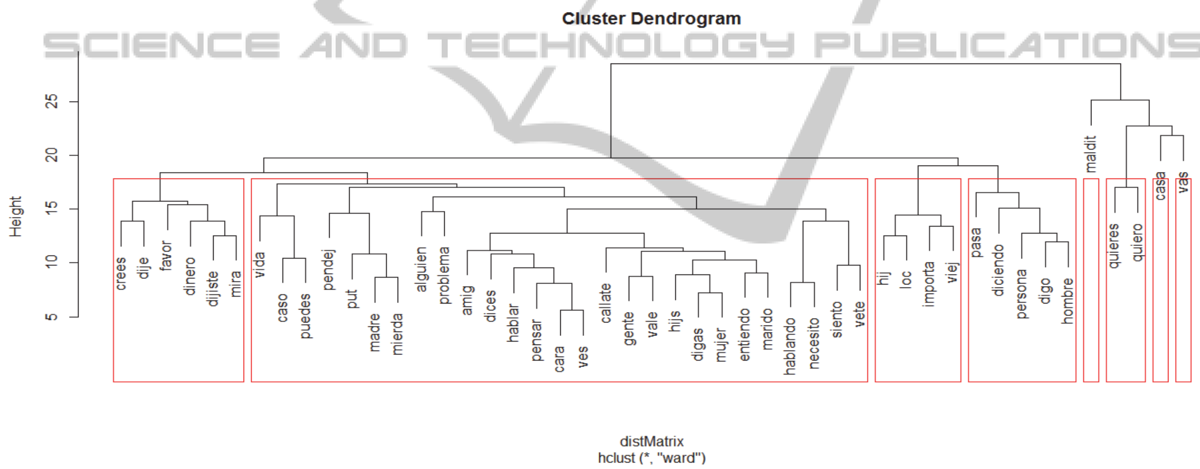

Figure 6: Dendrogram of 8 groups.

```
cluster 1: casa maldit quiero favor hij
cluster 2: coge pregunta put favor tema
cluster 3: dinero favor vas dame herencia
cluster 4: cobrar diciendo jugar escucha apoyo
cluster 5: quieres quiero miedo facil vida
cluster 6: insulto cuarenta desverguenza diputdo expresion
cluster 7: pendej estupid  adios cono mierda
cluster 8: vas crees mama dijiste miedo
```

Figure 7: The most important 5 words in each cluster using K-Means.

Using “K-medoids” (see Figure 8) the clusters 1 to 7 contain the words used in both, violent and non violent conversations, clusters 8 and 9 contain the words most probably used in violent conversations, and cluster 10 contains the words most used in non violent conversations.

In order to classify words with SVM the “e1071” package for R was used (SVM, 2014). The term matrix was saved along with the frequency of each term. If a term appeared more than 50% in non violent conversations (1 to 47), it was classified as non violent. On the other side, if a term appeared more than 50% in violent conversations (48 to 100), it was classified violent. If the percentage was the

```
cluster 1 : casa
cluster 2 :
cluster 3 : ido intrusos mantas sancionara
cluster 4 : usted
cluster 5 : favor
cluster 6 : quieres
cluster 7 : entonces
cluster 8 : apoco buen cabron ejemplo hablas haces hagas put sobrin trabajar
cluster 9 : cabana imbecil largate mal maldit  nacido oledor vete
cluster 10 : das dormir miedo puedo traeme vamos voy
```

Figure 8: Clusters obtained with the K- Medoids algorithm.

same (50%=V=NV=50%) it was marked not classified (SIN).

Data was saved in a text file that contains: Terms, percentage in non violent conversations (NV), percentage in violent conversations (V), and the classification (Clas). Figure 9 shows the results of the classification.

```
Call:
svm(formula = Clas ~ ., data = dataset)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.0004363002

Number of Support Vectors:  1118

 ( 517 490 111 )


Number of Classes:  3

Levels:
 NV SIN V
```

Figure 9: Results of SVM classes and levels.

Among the 454 non violent terms in the test set, 398 were predicted as non violent and 56 as violent. Also there are 31 terms marked not classified from which 28 were predicted non violent and 3 as violent. On the other side there are 278 violent terms in the test set. From those, 60 were predicted as non violent and 218 as violent.

3 INTERPRETATION OF RESULTS

Table 4 shows the number of terms before and after pre-processing. Form the table we observe that many words are removed in order to obtain better results when classifying a conversation.

Table 4: Data from the term matrix, violent conversations vs. non violent conversations.

	Analyzed conversations	Terms before pre-processing	Terms after pre-processing
Non violent	47	1971	1549
Violent	53	1388	1133
Total	100	3359	2682

Table 5 shows the most frequent terms in violent and non violent conversations. We observe that there are words that appear in both like “Quieres” among others.

Table 5: Most frequent terms, violent conversations vs. non violent conversations.

Conversation type	Term	Repetitions
Non violent	Casa	25
	Quiero	23
	Vas	22
	Quieres	17
	Miedo	14
Violent	Maldit@	25
	Pendej@	22
	Dinero	20
	Favor	18
	Quieres	14

Table 6 shows the grouping validation between 2 and 10 clusters obtained with “clValid” where we can compare the results with “Ward”, “K-means” and “PAM”. Also, the validation measures *connectivity*, *silhouette* and *Dunn* are shown.

Observing the results in Table 6 it is possible to choose the most efficient technique given that the smaller values indicate better grouping.

From Table 7, the validation of results, it is observed that the hierarchical grouping is the best in each case, and the optimum is with the Dunn validation with 8 clusters.

4 CONCLUSIONS AND FUTURE WORK

After analyzing the results we consider pre-processing an important step to obtain the words

Table 6: Comparison of results using: “Ward”, “K-means” and “PAM”.

		2	3	4	5	6	7	8	9	10
hierarchical	Connectivity	2.929	7.7869	10.7159	17.8361	18.5028	21.4317	24.3607	25.3607	26.3607
	Dunn	0.4749	0.4362	0.4862	0.5049	0.5049	0.5133	0.5586	0.5586	0.5586
	Silhouette	0.8218	0.8211	0.8156	0.8068	0.8033	0.781	0.772	0.7714	0.7713
kmeans	Connectivity	14.9071	16.6933	16.8361	21.694	27.552	36.9048	37.5714	34.6425	35.6425
	Dunn	0.4309	0.3369	0.3369	0.2139	0.2013	0.1833	0.1833	0.2357	0.2357
	Silhouette	0.8148	0.8029	0.781	0.7631	0.7541	0.6432	0.6431	0.6432	0.6431
pam	Connectivity	84.7679	115.2321	175.2448	214.1349	237.7694	260.3389	288.7841	302.1587	316.8964
	Dunn	0.0693	0.0693	0.069	0.0735	0.0786	0.0786	0.0786	0.0786	0.0786
	Silhouette	0.1555	0.0569	0.0784	0.096	0.1141	0.1292	0.1437	0.158	0.167

Table 7: Optimum results for grouping.

	Score	Method	Clusters
Connectivity	2.9290	hierarchical	2
Dunn	0.5586	hierarchical	8
Silhouette	0.8218	hierarchical	2

most used in violent and non violent conversations. The word clouds and term graphs were very useful to visualize the most used words and their frequency in each type of conversationIt was observed that in both, violent and non violent conversations, there are terms that are used with similar frequencies, but in the violent conversations these words are accompanied by other words considered “bad words” or “rudeness”.

Once the word grouping was performed, eight groups of words that appear frequently together in conversations (violent or non violent) were obtained. Not all the grouping techniques used were adequate, but the hierarchic technique *ward* was the most efficient given that the closeness of their words was much better than *k-means and PAM.*

When classifying terms in violent or non violent using SVM it was observed that some terms cannot be classified due to the fact that they appear with similar frequency in both types of conversations. In the training sets of SVM it was observed that the performance depends on the size of the test set.

Using the procedure presented in this work it is convenient to experiment with a larger number of video segments, and also use a better pre-processing that can include synonyms and removal of gender in most words. Also it is convenient to try more data mining techniques in order to make a thorough comparison and obtain better results due to the larger number of terms.

With the procedure presented in this work it is possible to design a system capable of classifying automatically conversations as violent and non violent. And this system can evolve to make this classification in real-time in order to trigger some alarm when a conversation turns violent in order to alert security personnel to take measures.

Another idea that can be explored is to pre-classify the video segments in categories like sports, political, family, commercial, etc. and also by region or social context in order to help the classification process.

REFERENCES

clValid: An R Package for Cluster Validation, [Online], Available at: http://www.jstatsoft.org/v25/i04/paper [Retrieved January 2014]

Meyer, D., “Support Vector Machines: The Interface to libsvm in package e1071”, September 2012, [Online], Available at: http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf [Retrieved January 2014]

Brun, R. E., Senso, J. A., Minería textual, [Online], Available at: http://www.elprofesionaldelainformacion.com/contenidos/2004/enero/2.pdf [Retrieved January 2014]

Montes-y-Gómez, M., Minería de texto: Un nuevo reto computacional, [Online], Available at: http://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf [Retrieved January 2014]

Villanueva, V. J., Escribano, M., Isorna, M., Pellicer, J., Alapont, L., Pellicer, P., Programa de apoyo al ámbito familiar: Agresividad y violencia, Editorial IES Pablo Serrano. Andorra (Teruel), España, 2007.

Adobe Premiere Pro CS6, [Online], Available at: http://www.adobe.com/mena_en/products/premiere.html [Retrieved January 2014]

Modelos de análisis de voz para Adobe Premiere Pro CS6, [Online], Available at: http://www.adobe.com/es/products/premiere/extend.displayTab3.html, [Retrieved January 2014]

RStudio v0.97.551, [Online], Available at: http://www.rstudio.com/ide/download/desktop [Retrieved January 2014]

Support Vector Machines in R, [Online], Available at: http://www.jstatsoft.org/v15/i09/paper [Retrieved February 2014]

An Introduction to R, [Online], Available at: http://cran.r-project.org/doc/manuals/R-intro.pdf [Retrieved January 2014]

R Data Import/Export, [Online], Available at: http://cran.r-project.org/doc/manuals/r-release/R-data.html [Retrieved January 2014]

Grün, B., Hornik, K., “Topicmodels: An R Package for Fitting Topic Models”, 2011, [Online], Available at: http://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf [Retrieved January 2014]

Feinerer, I., Hornik, K., “Package ‘tm’”, August 2013, [Online], Available at: http://cran.r-project.org/web/packages/tm/tm.pdf [Retrieved January 2014]

Wainschenker R., Doorn, J., Castro M., “Medición Cuantitativa de la Velocidad del Habla”, 2002, [Online], Available at: http://www.sepln.org/revistaSEPLN/revista/28/28-Pag99.pdf [Retrieved March 2014]

Zhao, Y., R and Data Mining: Examples and Case Studies, 2013, [Online], Available at: http://cran.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf [Retrieved March 2014]

Data Preprocessing Techniques for Data Mining, [Online], Available at: http://iasri.res.in/ebook/win_school_aa/notes/Data_Preprocessing.pdf [Retrieved March 2014]

Hastie T., Tibshirani R., Friedman J., The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Springer, 2001.

Bourel, M., Support Vector Machines, [Online], Available at: http://www.iesta.edu.uy/wiki/images/7/71/SVMSemEstadistica.pdf [Retrieved March 2014]

Derbas, N., Quénot, G., "Joint Audio-Visual Words for Violent Scenes Detection in Movies", Internation Conference on Multimedia Retrieval ICMR'14 Glasgow, United Kingdom, April 01-04, 2014

Gong, Y., Wang, W., Jiang, S., Huang, Q., Gao, W., "Detecting violent scenes in movies by auditory and visual cues. Advances in Multimedia Information Processing PCM 2008, 317-326, Spring Berlin Heidelberg, 2008

Lam, V., Phan, S., Ngo, T., Le, D., Duong, D., Satoh, S., "Violent Scene Detection Using Mid-level Feature", SolCt'13, Danang, Vietman. December 5-6, 2013.

Fujii, Y.,, Yoshimura, T., Ito, T., "Filtering Harmful Sentences based on Three-Word Co-occurrence", CEAS'11, Perth, Western Australia, Australia, September 1-2, 2011.