

Temporal topics in online news articles: Migration crisis in Venezuela

*Paulette Vázquez, †*Juan Carlos García, ‡María José Luna, and *Carmen Vaca

**Facultad de Ingeniería en Electricidad y Computación, FIEC*

Escuela Superior Politécnica del Litoral, ESPOL

Campus Gustavo Galindo Km 30.5 Vía Perimetral, Guayaquil, Ecuador

†*Facultad de Ingeniería Industrial*

Universidad de Guayaquil, UG

Cdla. Salvador Allende, Av. Delta y Av. Kennedy, Guayaquil, Ecuador

‡*Maestría de investigación en derecho*

Universidad de Los Hemisferios

Paseo de La Universidad N° 300 y Juan Díaz - Urbanización Iñaquito Alto, Quito, Ecuador

{jpvazque, jugarcia, cvaca}@fiec.espol.edu.ec, juan.garciapl@ug.edu.ec, mjllunlara@hotmail.com

Abstract—The migration process of citizens of a country in crisis can be extended by months or years depending on the social, economic, political, institutional situation of the country of origin. There are several articles that approach Venezuelan migration based on information provided by international organizations such as ONU and ACNUR. However, so far there is not a study that analyzes the information media offers about this issue, so we have temporarily analyzed online news to be able to obtain the topics that emerges from this social phenomenon. First, we extract 10K news articles published online in different newspapers across Latin America since 2015 until May 2019. Second, we build a binary classifier to discriminate whether the article is related to migration or not. Finally we apply topic modeling and word embeddings techniques to extract the most important issues discussed each year. The automatic text analysis, in time, reveals how the country moves from an intense migratory flow to an exodus of people of all ages and conditions who are welcomed by the host countries of the region starting with the closest neighbors. Our temporal analysis shows evidence that the migration process continues to increase and it is spreading throughout the continent mainly due to unemployment, insecurity and the lack of medicines and food.

Index Terms—migration crisis, venezuelan migration, text classification, online news, word embedding, topic modeling.

I. INTRODUCTION

In July 2019, the International Organization for Migration (OIM), and the UN Refugee Agency (UNHCR) stated that 3,929,560 people had left Venezuela until the first semester of 2019¹. According to the Organization of American States (OAS), the Venezuelan migration process could be the largest in history if their magnitude continues to grow as it has happened until now. Such crisis requires to analyze the causes and the consequences for people leaving Venezuela and for the receiving countries.

According to the latest ONU's figures, from 2015 to date (2019) 4,486,860 Venezuelans have left their country and

¹<https://www.iom.int/news/refugees-and-migrants-venezuela-top-four-million-iom-and-unhcr>

have arrived in Colombia, Peru, Chile, Ecuador, Brazil and Argentina, South American countries with internal socio economic issues to solve [1]. Regional migration flows have only increased these years with 81.5% of the migrants heading to different countries in Latin America and the Caribbean [2]. Such flows are imposing burdens over the neighbor-countries' economies [3]. Consequently, news related to the crisis has been constantly in the media these years. We aim to see the crisis through the lens of digital texts published in the last four years and let algorithms to uncover the temporal evolution of the discussed topics.

Migrations crisis evolution over time has been analyzed, starting from published news, in countries such as Italy [4] since that country has been facing many challenges because of the migration flows. Similar work has been conducted for other countries in Europe including France and UK [5]. However, to the best of our knowledge, previous work has not addressed the media coverage of the Venezuelan migration crisis and here we aim to present such analysis.

We start by collecting news from different newspapers with online presence in South America. Although it is possible to collect large volumes of information automatically, it is very easy to end up with noisy data that needs to be properly filtered. Therefore, we apply machine learning algorithms to keep only relevant information.

The information we have collected from the media shows the recurring subject in the public sphere, in other words, the published opinion, which becomes public opinion. Not only are they issues discussed on the street, with the family, in the workplace, but also they are the ones that are treated as public issues and as such become part of the government agenda. As a result of our analysis we make the following contributions:

- A collection method for news across years.
- A classifier to discriminate news media articles (written in Spanish) whether related to migration or not.
- An analysis of automatically discovered topics over four years of published news

The rest of the paper is organized as follows. Section II includes previous work. Section III explains the collection process of our online-news dataset. Section IV describes the experimental framework. Section V presents the results and Section VI discusses our findings.

II. RELATED WORK

Venezuelan migration is an issue that has been addressed in research in recent years [6]–[8]. Researchers have discussed causes and consequences of the Venezuelan migration under a broad range of topics such as the health crisis [9], the governability risks for countries in the region [10], the human trafficking [11] among others. Most of these studies are conducted on user generated data published on social networks or data provided by public or private entities [12].

In the past, some authors have analyzed migrations crises in other latitudes from the media point of view, i.e. analyzing news published in national or international media channels. Allen and Blinder [13] used a “corpus linguistic” approach to analyse the British press coverage of migration founding patterns through the analysis of collocations, for example: “immigrants” and “migrants”, “refugees” and “asylum seekers”. Fotopoulos and Kaimaklioti [14] applied the same methodology on media discourse of Greek, German and British press, concluding that the media seems to have taken a uniform approach for covering this crisis. Besides, they mention that all the newspapers examined paid a great deal of attention to problems related to refugee children.

Ylä-Anttila et al. [15] analyzed the countermedia style and its adoption by parliamentarians in the politization of migration, in the case of Finland. Three datasets were integrated: countermedia (WTF), mainstream media (HS), and parliamentarians’ discourses in plenary sessions of the Finnish parliament. They compared the bigram frequencies of WTF and HS, and found that WTF media discusses migration mostly in conjunction with crime illegality and the police activity in a populist style.

Chavez et al. [16] made a content analysis of the major U.S. Newspapers’ coverage of Mexican immigration. A total of nine topics were used to categorize each document of the corpus: “border issues”, “crime”, “economics”, “education”, “family”, “health”, “legislation”, “prejudice” and “statistical analysis”. Some of the most important findings of this work is that most of the studied news were mainly used to influence public opinion. The majority used the information presented to influence readers’ emotions, opinions, or actions and a smaller number gave readers strictly informational content.

We propose to analyze online news articles regarding the Venezuelan migration and extract the most relevant topics in a chronological fashion applying topic modeling techniques. Topic modeling tools have become very popular for these tasks, being the following the most common models: Latent semantic analysis (LSA), Probabilistic latent semantic analysis (PLSA), Latent Dirichlet allocation (LDA), and Correlated topic model (CTM). [17]

TABLE I: Documents per year description

Year	Number of documents	Number of words
2015	104	83830
2016	114	58349
2017	126	79915
2018	530	511488
2019	326	175517

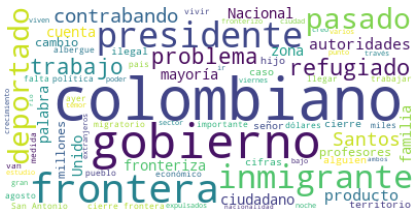
Online news platforms such as Google News have been used as a source of data for refugee [18] and migration [19] studies in Europe. Through keyword searches, we can obtain news related to a specific topic. However, the information obtained is diverse and may include non-relevant content. In this sense, automatic text classification methods are needed to determine whether a news item is related to a specific topic such as migration. Nowadays, deep learning techniques have proven to be more successful than traditional machine learning techniques. The convolutional neural network (CNN) have achieved a remarkable performance in text classification tasks [20]. CNN also have been applied in sentiment analysis [21] prediction over Twitter data, identifying text political tweets [22], predicting political ideology [23], [24]. Commonly, CNN architectures work together with semantic distribution models that represent words in numerical vectors called word embeddings. Recent word embedding approaches such as Glove [25] have proven to be an effective approach for classifying topic-relevant text.

III. DATASET

The data used in this present research comes from news articles published online in newspapers across Latin America regarding Venezuelan migration. We first selected the most popular newspapers in each of the nine Latin American countries included in this study. Next, we gather articles published in the corresponding online sites since January 2015 until May 2019. In that way, we build a ten thousand article corpus, storing three fields for each news: headline, content and publication date.

We search newspaper sites using a list of keywords: “migración”, “éxodo”, “venezuela”, “venezolanos”, “venezolana”, “venezolano”. Although we conduct a keyword based search, the resulting corpus contains a fraction of articles not related to the topic of interest. Therefore, we implemented a binary classifier to decide whether articles are “migration related” or not. We describe such classifier in the Methods section.

After automatically classifying each article, we are able to filter those dealing with entertainment, sports, religion and everything that has nothing to do with the migration process in the Andean country. As a result we end up in a corpus containing 1,200 records with a temporal distribution shown in Table I.



(a) Wordcloud 2015



(b) Wordcloud 2016



(c) Wordcloud 2017

Fig. 1: It is evident that topics appearing in the news are changing across time.

IV. METHODS

A. Pre-processing

After the exploratory analysis of the data, a first filtering was done. The content extracted from the online news not only consisted in the body of the news, but also included headers of information such as advertising, so all of these irrelevant sections for the analysis were removed.

In the same way, we also removed records in which their content had less than 20 words, was empty or duplicated. Finally, the date column was separated into two features: month and year, since we are interested in doing a temporary analysis of the documents based on the year in which they were published.

B. Identifying migration related news

Every day thousands of news articles are accessible on the Internet. They include a wide variety of topics, even when they are filtered by keywords. For example, doing a search for news online with the keywords “Venezuela” and “Migration” results in documents related to sports, music, politics, and the exodus itself. In this investigation, a document is defined as a newspaper article related to the crisis of Venezuelan refugees who migrated at a large scale between 2015 and 2019. Given the large number and variety of documents available, it is essential to develop an automatic classification method.

Automatic classification of documents is a fundamental task to discard those articles news that produce noise. The news extraction process generated a corpus of 10,000 documents. However, not all documents are relevant to migration. This is because the search by keywords is not contextual to the topic analyzed. The table II shows the training data formed by two columns: the text of the news and the binary class that determine whether the news is related to migration. For example, the news about the Venezuelan singer ‘El Puma’ who receives recognition for his career. In his speech he regrets the situation of migration in his country. However, this news is not directly related to migration, therefore its class is defined as 0. On the other hand, the news with reference to the charity of good people alleviates migrant needs in Colombia is marked as 1. Our classifier must Learn to discriminate directly related news the transit of the migratory flow from Venezuela to other countries.

Deep neural network has proven to be an effective technique for categorizing text. For instance, we use an architecture of a

TABLE II: Training corpus with a binary class

Headline	Class
‘El Puma’ recibe Grammy a la excelencia lamentando migración venezolana y destrucción del país	0
Miraflores acoge primera expo-feria emprendedora peruana-venezolana	0
Integrantes de Los Cadillacs y Aran de las Casas heridos en accidente aéreo registrado en Caracas	0
Organizan en Bruselas conferencia de solidaridad por migración venezolana	0
Solo la caridad de buenas personas alivia necesidades de migrantes en la ruta Cúcuta- Bucaramanga	1
Ante poca posibilidad de cambio migración continuará y urge movilizar a la región dicen expertos	1
En Ipiales según los registros se quedaron a vivir 376 venezolanos	1
Se estima que a diario ingresan a Colombia cerca de 40.600 migrantes del vecino país	1

convolutional neural network (CNN) [20] to accomplish such tasks in our data. The CNN architecture contains several layers such as embedding, convolutional, max pooling, dense units, and the fully connected output. Our text classification pipeline includes several tasks: parameter selection, model training, model evaluation and results improvements.

1) *Parameters selection*: Selecting the model parameters properly has a significant impact on the accuracy of deep models. They are also known as hyper-parameters, which determine the structure of the neural network (e.g. number of convolutional layers, number of filters, filter size) and how it is trained (e.g. learning rate, maximum number of epochs, dropout rate). The Table III shows the tuning parameters for our CNN architecture.

TABLE III: Parameters selected for the CNN model

Parameter	Value
Embedding Dimension	300
Word Embeddings	Pre-trained Glove
Epoch	12
Batch Size	128
Filters	64
Kernel Size	3
Pooling	Max
Dense units	10

The **embedding layer** is the first in the neural network architecture. It consists of transforming each word into a numerical vector. Word embeddings represent the vocabulary of words in a continuous vector space. They capture the

meaning similarities of the words, which allow to extract the semantic features. For this purpose, we use a pre-trained word vectors [26] with Glove [25] embedding technique using different sources on the web such as Spanish Wikipedia.

The **convolutional layer** allows you to learn the important words in a news document. The output of the embedding layer is the input for the 1-dimensional (1-D) convolution operation which was configured with 64 filters and a kernel size of 3. Therefore, word convolutions are performed in equivalent combinations to kernel size. This allows the neural network to learn the context when words are related to other words.

The **max-pooling layer** converts each 3-word convolution into a single output by choosing the largest value in each kernel. This operation is used to reduce over-fitting allowing to generalize by extracting high level features.

Two **dense layers** were used after max-pooling. The first had 10 units with a ReLu activation function. The other dense layer had 1 unit with the sigmoid activation function.

The batch size in training is 128. The Adam optimizer was used with adaptive learning to minimize binary cross-entropy loss in 12 epochs.

2) *Model training*: The training phase was carried out with 200 news documents labeled manually. The binary tag defines 1 as migration and 0 as non-migration. The training corpus is balanced with 50 % of documents of each class. This is to avoid the problem of the unbalanced class.

3) *Model validation*: We apply 10-fold cross-validation as a classifier evaluation procedure. Then, we calculated the macro F1 score. The precision obtained was 82.4%.

Next, we apply the classifier to detect migration documents over a 1,000 news articles. After this execution the classifier detected document news relevant to migration, however there were still false positives.

4) *Model evaluation*: The next step was to increase the size of the training news with additional 200 documents, including those that in the first instance were detected as false positives. Likewise, the process was evaluated with 10-fold cross-validation. The precision obtained was 88.9%.

Finally, the model was applied to detect articles of emigration on the corpus with 10,000 news obtaining 1,200 successful results.

C. Uncovering emerging topics

Our main interest is to find out how topics in the news have temporarily changed. Although this process can be done manually, when the amount of documents is considerably large, it would be time consuming.

In recent years several tools that help us to perform this work automatically have emerged, of which one of the most used is Topic Modeling. It is an unsupervised learning approach to clustering documents, that helps us to discover topics according to their contents. When clustering documents we will have to process the individual words belonging to each document, causing the amount of data to increase. In order to do this, efficient dispersed data structures are needed. [27] We use two different approaches, which are detailed below.

1) *Latent Dirichlet Allocation (LDA)*: We use LDA to generate the distribution of five-year topic words. [28] Where rows and columns of each Document Term Matrix (DTM) are the news and the the total amount of words existing in all news respectively. The number of documents and the number of words used by the model for each year are described in the figure I. What we achieve with this process is to obtain five sets of words per year, where each set is made up of the most relevant words that describe it. The tables IV to VIII show the topics generated with LDA.

TABLE IV: LDA topics of 2015

Topic	Words
1	autoridades, colombianos, agosto, recibir, gobierno, ciudadanos, información, seguridad, deportados, frontera
2	frontera, seguridad, autoridades, unión, mayoría, recibir, información, mejor, zona, viven
3	información, unión, pasado, cambio, territorio, mejor, agosto, frontera, productos, zona, viven
4	mayoría, gobierno, territorio, trabajo, ciudadanos, viven, cambio, autoridades, presidente, pasado
5	colombianos, frontera, presidente, santos, pasado, contrabando, froteriza, cierre, productos, deportados

TABLE V: LDA topics of 2016

Topic	Words
1	gobierno, escasez, frontera, presidente, pasado, encuentra, comprar, alimentos, seguridad, prueba
2	información, frontera, recibir, diario, cosas, mejor, prueba, encuentra, comprar, alimentos, seguridad, prueba
3	frontera, recibir, pasado, presidente, gobierno, mejor, finanzas, económica, vive, conocimientos
4	alimentos, económica, recibir, pasado, gobierno, escasez, vive, frontera, ciudadanos
5	recibir, gobierno, pasado, mejor, finanzas, económica, conocimientos, vive, frontera, cosas

TABLE VI: LDA topics of 2017

Topic	Words
1	trabajo, ciudad, frontera, miles, migraciones, dinero, encuentran, trabajar, brasil, gobierno
2	niños, viaje, busca, vida, unión, pasadi, peruanos, dólares, nacional, incluso
3	autoridades, ciudad, miles, trabajar, viaje, brasil, director, unión, económica, peruanos
4	trabajar, viaje, brasil, director, unión, económica, autoridades, peruanos, diario, miles
5	permiso, permanencia, ciudadanos, temporal, migratoria, migraciones, nacional, unión, pasado, gobierno

2) *Nonnegative Matrix Factorization (NMF)*: We use NMF as another topic identification technique, this also uses DTM described in the previous subsection and model factors the high-dimension vectors into a lower representation. Further, it has an inherent property of clustering, so it generates two matrices from DTM, where one of them represents the clusters of topics identified in the articles and the other represents the matrix of weights for each topic. This is how we get five topics for each year described in the tables IX to XIII.

TABLE VII: LDA topics of 2018

Topic	Words
1	estrategia, conjunta, llega, frontera, huir, momento, latinoamérica, vuelta, encuentro, emigran
2	indican, pese, permitir, vuelta, varias, salido, secretario, latinoamérica, patria,ministro
3	quedan, decidieron, corte, desarrollo, colombianos, social, septiembre, migratorio, meses, patria
4	gobierno, pasaporte, millones, refugiados, frontera, ciudadanos, brasil, presidente, migratoria, onu
5	argentina, destinos, salen, futuro, emigrado, mejor, necesidad, quedó, precios, altos

TABLE VIII: LDA topics of 2019

Topic	Words
1	fronterizos, ayuda, nacional, salud, bajo, puente, incluso, hiperinflación
2	edad, apenas, hiperinflación, incluso, bajo, república, hizo, ayuda, froterizos, nacional
3	fronterizos, ayuda, salud, social, nacional, puente, bolívar, ayuda, apenas, bajo
4	millones, gobierno, refugiados, presidente, frontera, unidos, ayuda, internacional, humanitaria, ciudadanos
5	fronterizos, ayuda, nacional, salud, apenas, social, bajo, puente, incluso, hiperinflación

D. Semantically related issues

We use the vector-based language model to represent the semantic and syntactic relationships of words in the migration corpus. The corpus was divided into five parts corresponding to the years 2015-2019. A separate linguistic representation allows us to verify how the subject has evolved over time.

Word2Vec [29] is the word embedding technique that we applied to train each corpus to learn the semantic relationships of words. Then, four topics found by NMF were selected: food, violence, children, and work. Finally, we obtained the words that give a semantic context related to each subject analyzed.

The **embedding dimension** used to train the word vector was 300. A smaller embedding dimension may result in less precision. A larger value has an impact on computational efficiency. Usually 300-dimension [30] is a good balance between precision and computational performance.

The context learned is determined by the **windows size**. A large context window can perform statistically significant improvements in different tasks such as classification [23]. In this research, we used a windows size of 10.

The appropriate language model for our research is Skip-Gram. For a given context we need to predict the words. In this case, the context is given by the topics previously found by the NMF method. The table XIV shows the parameters used in the training. Through this method we were able to find the semantic words to enrich the topics. These topics are shown in the results section.

V. RESULTS

Applying the automatic migration news classifier, we successfully created a corpus with 1,200 news documents. Then,

TABLE IX: NMF topics of 2015

Topic	Words
1	fronteriza, gobierno, colombiano, contrabando, deportados, cierre, frontera, presidente, santos, colombianos
2	ahora, últimos, situación, cifras, trabajo, mayoría, mayor, gobierno, habitantes, presidente
3	recibir, pasado, mejor, deportados, xona, unión, ciudad, seguridad, frontera, autoridades
4	cierre, presidente, trabajo, fronteriza, deportados, territorio, frontera, unión, cambio, pasado
5	colombianos, viven, colombiano, pasado, trabajo, recibir, cambio, territorio, frontera, información

TABLE X: NMF topics of 2016

Topic	Words
1	vive, seguridad, comprar, trabajo, presidente, pasado, gobierno, alimentos, económica, escasez
2	finanzas, conocimiento, llegada, encuentra, frontera, prueba, cosas, diario, recibir, información
3	escasez, comprar, alimentos, últimos, frontera, económica, capital, finanzas, trabajo, semana
4	vive, económica, trabajo, seguridad, recibir, alimentos, frontera, gobierno, encuentra, fronteriza
5	mejor, encuentra, conocimientos, finanzas, prueba, llegada, recibir, frontera, información, llegada

using an unsupervised method, our approach was able to discover the main issues in the migratory exodus from Venezuela. Using the word embedding technique, we analyze the semantic evolution of four topics. The table XV shows the semantic relations. We analyze the evolution of these topics in the discussion section.

VI. DISCUSSION

From the data extracted, the evolution of the Venezuelan migration process can be noted. Beginning in 2015, terms such as: “frontera”, “colombianos”, “fronteriza” begin to emerge, and this is how the mobilization of Venezuelans to their closest border, Colombia, is increasing. In 2017, within the topics we speak of Peru, in 2018 of Brazil, of “éxodo”; in 2019, the perspective is changed, by resorting to terms that have a strong international connotation, such as: “refugiados” “visa”, “humanitaria”, “ingreso”, “pasaporte”, “ONU”, “proceso”, “ACNUR”.

As of 2018, the emergence of topics related to large humanitarian organizations, especially ACNUR and the OAS, shows the need to mitigate the problems that this unusual migratory transit entails. In fact, it is not only the largest population displacement in the region, but also is considered a worldwide migratory emergency.

A mobilization of these characteristics and magnitude entails serious social, legal and economic problems both for those who take refuge and for the recipient countries. They arrive with an eventual immigration status, which complicates the ability to find housing, work, given their instability; that is, a state of subsistence that faces adverse situations. The consequences for the receiving nations are complicated, being an underdeveloped region, in which each nation has its

TABLE XI: NMF topics of 2017

Topic	Words
1	presidente, ingreso, nacional, diario, pasado, vida, frontera, gobierno, semana, situación
2	económica, número, vida, trabajar, diario, dinero, busca, peruanos, miles, trabajo
3	cifras, peruanos, nacional, temporal, permanencia, ciudadanos, permiso, migratoria, migraciones
4	encuentran, internacional, autoridades, viaje, nacional, unión, niños, encuentran, niños, ciudad
5	dinero, miles, mientras, frontera, ciudadanos, trabajo, dólares, viaje, permanencia, temporal

TABLE XII: NMF topics of 2018

Topic	Words
1	onu, miles, gobierno, unidos, éxodo, internacional, población, brasil, últimos, millones
2	cédula, documento, exigencia, identidad, medida, territorio, ingreso, ecuatoriano, ingresar, pasaporte
3	comida, madre, conseguir, familia, compatriotas, dinero, trabajo, terminal, reunión, representante
4	organización, comisionado, agencia, regional, naciones, unidas, embajada, onu, acnur, refugiados
5	migratorio, educación, pasaportes, humanitaria, Chávez, andina, documentos, ciudadanos, movilidad, Iationamérica

own economic, political and social problems; in addition to the aggravating fact that criminal organizations and common criminals have served from this diaspora.

Within this context, the recipient countries have found themselves in need of a series of legal reforms, as well as rethinking public policies that allow them to face this process and guarantee the rights of this emerging population, as well as that of their own citizens, analysis that will be the subject of further investigation. The general dynamics of migration, and the Venezuelan process specifically, is a real, continuous and permanent event; characteristics that make it impossible to reach conclusions with the scientific academic collection that can be applied to other types of studies. But the data addresses the issue sufficiently to be able to determine, for example, with words from Arango that: “economic disparities are, without a doubt, a necessary condition for most migratory flows, but in no way sufficient condition for these occur”. [2]

With this reflection, it can be corroborated that in 2015 the labor factor is the topic that is most repeated, and this is related to the official information that in Venezuela the unemployment rate has increased by 7.6% in 2013 to 26.4% in 2019.² The analysis of the issue about this embedding is marking the path to search for opportunities, difficulties, informality and even trying to stabilize and achieve new possibilities and above all acceptance. Christina Boswell explains that: “Economic opportunities in destination countries are also crucially important in influencing decisions to migrate and the direction of flows. Thus in addition to macro push factors and migration systems, there are a number of significant “pull” factors in receiving countries”. [31]

²www.indexmundi.com/g/g.aspx?c=ve&v=74&l=es

TABLE XIII: NMF topics of 2019

Topic	Words
1	ciudadanos, gobierno, frontera, nacional, presidente, ingresar, pasaporte, medida, migratoria, visa
2	régimen, asesinatos, políticos, refugiados, fuerza, guerra, petróleo, mundial, territorio, vida
3	comunidad, recibir, humanitaria, interno, ayuda, millones, presidente, unidos, Iván, duque
4	hiperinflación, niños, ayudar, dinero, llegaron, hijos, refugio, familia, ayuda, edad
5	económica, vida, cifras, lugar, argentina, ciudad, segundo, migratorio, extranjeros, cantidad

TABLE XIV: Word2Vec parameter selection

Parameter	Value
Embedding Dimension	300
Word Embeddings	Word2Vec
Language model	Skip-gram
Windows size	10
Epochs	12
Sampling	non-negative

And on the reverse of the currency, seen from the countries that receive the Venezuelan population, the same effect is evidently produced, the greater the migration, the lower the employment rate in general and an increase in underemployment, begging and crime, both for nationals and foreigners in addition to the violation of the human rights of migrants, who are willing to work for less money, more time than allowed by the law and without social security, etc.

In fact, only in Pichincha, the province of Ecuador whose capital is Quito, has the highest number of Venezuelan migrants, and today, it has reached the highest average in extreme poverty rate in the country, with 2, 4%.³

Another factor that accelerates the flow of migration is insecurity and violence. The Venezuelan Violence Observatory (OVV) is a non-governmental organization that investigates criminal indices in collaboration with the academy, due to the lack of official information, which on the contrary deals with this problem. According to the figures 20 years ago 4,550 homicides were registered per year and in 2016, 28,000 victims, in 2018, 23,000 violent deaths. The homicide rate increased from 19 per 100,000 inhabitants in 1998 to 81.4 in 2018, when violence was even higher than in countries in armed conflict.

The aforementioned is aggravated by the prevailing impunity, since the greater number of homicides are classified as resistance to the authority, and applies the penalty without trial under the police and the military called: “Operation of Liberation of the People.”⁴ As well as unregistered deaths, with a high rate of missing people.

The issue of violence can be seen more clearly in the embeddings topics, as the word “grave” is changed for: “muertos”, “estalló”, “violación”, “violaciones” and other violence that is the “xenofobia”, which appears in 2019. This last term is

³Instituto Nacional de Estadística y Censo (INEC).

⁴<https://observatoriodeviolencia.org.ve/>

TABLE XV: Semantic Topics for food, children, violence and work

2015	2016	2017	2018	2019
Alimentos (Food)				
escasez	medicinas	escasez	medicinas	medicinas
básicos	saqueos	hormigas	escasez	escasez
venden	necesidad	medicinas	productos	nube
inflación	precio	padeció	recesión	agua
Niños (Children)				
sacar	cruzaban	cruzaban	niñas	embarazadas
contó	mujeres	mujeres	adolescentes	padres
montañas	embarazadas	edades	ancianos	menores
preocupa	agradecidos	registros	adultos	vulnerabilidad
Violencia (Violence)				
grave	solicitudes	chile	exiliarse	xenofobia
descuidamos	pobres	diarios	persecución	violación
familiar	aumentado	escucha	desigualdad	generalizada
pobres	publicó	aumentado	salvar	explotación
Trabajo (Work)				
esposa	buscar	pido	trabajadores	grupo
oportunidades	venir	duro	pilotos	aceptan
jubilada	retornó	encontrar	barco	posibilidades
trabajando	vendió	falta	viaja	urbes

related to the effects that the migratory wave has caused in the receiving countries, whose citizens have been displaced in labor matters because they constitute the Venezuelan cheap labor force, as explained above ; objectively, unemployment and underemployment grew in the host regions; and the crime rate increased, evidently due to the increase in people residing illegally.

In 2016, the situation is exacerbated by the shortage, especially of food and medicine, which also have unaffordable prices. From a survey of 1,500 Venezuelan migrants living in Colombia, 72% left Venezuela because of shortage of food, 62% left in order to get money for their family, and 26% because of insecurity issues.⁵

Although since 2015 there is already talk of scarcity, inflation, corruption, in 2016 “saqueos”, “precio”, “medicina”, “necesidad”, “crónica”, “need” is added, which reveals the famine, inflation and general dissatisfaction. In 2017 of “padeecer”, in 2018: “medicamentos”, “elevados”, “recesión”, which in 2019 adds: “agua”, “bienes”, “cadenas”.

This information shows a causality, because of the high cost of the main inputs for the human being, such as water and food; as well as the lack of basic public services that must be provided by a rule of law, starting with the fundamental ones.

Regarding the profile of the migrant, it is evident that since 2016 not only men but also women and even pregnant women have already begun to cross the border 2017. By 2018, girls, adolescents, elderly, adults, pregnant women are counted, alone in pairs; In 2019 there is no distinction: pregnant, fathers, adults, minors, mothers, children, vulnerable, adolescents, babies and the elderly.

The above is an exodus, which according to the Dictionary of the Spanish Language, is the emigration of a town or a crowd of people; in the same way it can be called a diaspora,

⁵Proyecto Migración Venezuela

which is the dispersion of human groups that leave their place of origin.

“Thus, people, alone, accompanied, in family groups or in pairs, migrate, under factors of vulnerability and inequality such as age, gender, ethnic-racial belonging and socioeconomic status of origin generate different types and degrees of vulnerability and risk. Indeed, migrants have all ages. They move, leave their homes, neighborhoods, cities and countries motivated by the search for a job, new horizons, despair or family reunification. (...) When the decision to migrate is not voluntary, they are compelled to leave their place of residence and lose their assets and livelihoods”. [32]

The CEPAL inequality matrix in Latin America and the Caribbean states that, “the structuring axes of the social inequality matrix intersect, strengthen and chain throughout the life cycle, which result in a multiplicity of inequality or discrimination factors that interact simultaneously or accumulate over time. The confluence of multiple inequalities and discrimination characterizes the “hard cores” of poverty, vulnerability and social exclusion and makes them persist and reproduce ” [32]

VII. CONCLUSION

The diagnosis put into consideration, aims to serve as input for the generation, and implementation of public policies according to the needs posed both for migrants and for citizens of recipient countries, not in the abstract, but as a real search for the common good.

There are sociological and political studies, such as: “The analysis of the frames” in the work of William Gamson, among others, whose theoretical application is absolutely consistent in this academic work. Thus, for example, in a footnote of this work it is stated that: When can we say that we have achieved understanding about the events of reality? As a hypothesis, we can say that we have understood a succession of events when we can discern which the relevant objects or subjects are, that play a role within those events and when we have discerned the relevant existing relations between those objects and subjects. We understand a series of events when we distinguish between those elements that are central for those events to occur and those elements that are contingent or not necessary for those events to occur. [33]

What is described is a good synthesis of what is discussed in the paper, of how every public problem responds to a social reality that is reflected in the discourse, in the symbology, in a “cultural package” [33] Within the frameworks described by Gamson the most important is the “framework of injustice” about social movements, through which a value judgment can be made only under the certainty of the existence of elements and conditions that originates an insufferable situation for society. What has been fully supported in the case of Venezuelan migration.

REFERENCES

- [1] J. Dayton-Johnson, “Shaking up governance and inequality in south america: A political-economy account,” in *Handbook of South American Governance*. Routledge, 2017, pp. 320–333.

- [2] J. Arango, "La explicación teórica de las migraciones: luz y sombra," *Migración y desarrollo*, no. 1, 2003.
- [3] M. J. Camilleri and F. O. Hampson, "No strangers at the gate: Collective responsibility and a region's response to the venezuelan refugee and migration crisis," 2018.
- [4] L. Montali, P. Riva, A. Frigerio, and S. Mele, "The representation of migrants in the italian press: A study on the *corriere della sera* (1992–2009)," *Journal of Language and Politics*, vol. 12, no. 2, pp. 226–250, 2013.
- [5] A. Caviedes, "An emerging 'european' news portrayal of immigration?" *Journal of Ethnic and Migration Studies*, vol. 41, no. 6, pp. 897–917, 2015.
- [6] L. Freier and N. Parent, "The regional response to the venezuelan exodus," *Current History*, vol. 118, no. 805, pp. 56–61, 2019.
- [7] B. N. Gedan, "Venezuelan migration: Is the western hemisphere prepared for a refugee crisis?" *SAIS Review of International Affairs*, vol. 37, no. 2, pp. 57–64, 2017.
- [8] J. R. Torres and J. S. Castro, "Venezuela's migration crisis: a growing health threat to the region requiring immediate attention," *Journal of Travel Medicine*, vol. 26, no. 2, 2019.
- [9] A. J. Rodríguez-Morales, D. K. Bonilla-Aldana, M. Morales, J. A. Suárez, and E. Martínez-Buitrago, "Migration crisis in venezuela and its impact on hiv in other countries: the case of colombia," 2019.
- [10] V. M. Mijares and N. Rojas Silva, "Venezuelan migration crisis puts the region's democratic governability at risk," 2018.
- [11] M. John, "Venezuelan economic crisis: crossing latin american and caribbean borders," *Migration and Development*, vol. 8, no. 3, pp. 437–447, 2019.
- [12] A. R. Tuite, A. Thomas-Bachli, H. Acosta, D. Bhatia, C. Huber, K. Petrasek, A. Watts, J. H. Yong, I. I. Bogoch, and K. Khan, "Infectious disease implications of large-scale migration of venezuelan nationals," *Journal of travel medicine*, vol. 25, no. 1, p. tay077, 2018.
- [13] W. Allen and S. Blinder, "Migration in the news: Portrayals of immigrants, migrants, asylum seekers and refugees in national british newspapers, 2010-2012," 08 2013.
- [14] S. Fotopoulos and M. Kaimaklioti, "Media discourse on the refugee crisis: on what have the greek, german and british press focused?" *European View*, 11 2016.
- [15] T. Ylä-Anttila, G. Bauvois, and N. Pyrhönen, "Politicization of migration in the countermedia style: A computational and qualitative analysis of populist discourse," *Discourse, Context & Media*, vol. 32, pp. 1–8, 11 2019.
- [16] M. Chavez, S. Whiteford, and J. Hoewe, "Reporting on immigration: A content analysis of major u.s. newspapers' coverage of mexican immigration," *Norteamérica*, vol. 5, pp. 111–125, 12 2010.
- [17] R. Alghamdi and K. Alfalqi, "A survey of topic modeling in text mining," *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, vol. 6, no. 1, 2015.
- [18] M. Andretta and E. Pavan, *Mapping Protest on the Refugee Crisis: Insights from Online Protest Event Analysis*. Cham: Springer International Publishing, 2018, pp. 299–324. [Online]. Available: https://doi.org/10.1007/978-3-319-71752-4_11
- [19] L. Chouliaraki and R. Zaborowski, "Voice and community in the 2015 refugee crisis: A content analysis of news coverage in eight european countries," *International Communication Gazette*, vol. 79, no. 6-7, pp. 613–635, 2017.
- [20] Y. Kim, "Convolutional Neural Networks for Sentence Classification," pp. 1746–1751, 2014. [Online]. Available: <http://arxiv.org/abs/1408.5882>
- [21] P. M. Sosa, "Twitter Sentiment Analysis using combined LSTM-CNN Models," pp. 1–9, 2017.
- [22] L. S. Oliveira, P. O. S. V. de Melo, M. S. Amaral, and J. A. G. Pinho, "When Politicians Talk About Politics: Identifying Political Tweets of Brazilian Congressmen," no. 2011, 2018. [Online]. Available: <http://arxiv.org/abs/1805.01589>
- [23] X. Yang, C. Macdonald, and I. Ounis, "Using word embeddings in Twitter election classification," *Information Retrieval Journal*, vol. 21, no. 2-3, pp. 183–207, 2018.
- [24] S. Ruder, J. Glover, A. Mehrabani, and P. Ghaffari, "360 deg Stance Detection," pp. 31–35, 2018. [Online]. Available: <http://arxiv.org/abs/1804.00982>
- [25] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014. [Online]. Available: <http://aclweb.org/anthology/D14-1162>
- [26] C. Cardellino, "Spanish Billion Words Corpus and Embeddings," August 2019. [Online]. Available: <https://crscardellino.github.io/SBWCE/>
- [27] X. Cheng, X. Yan, Y. Lan, and J. Guo, "Btm: Topic modeling over short texts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.
- [28] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [29] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," *Proceedings of NAACL-HLT*, no. June, pp. 746–751, 2013. [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Linguistic+Regularities+in+Continuous+Space+Word+Representations#0%5Cnhttps://www.aclweb.org/anthology/N/N13/N13-1090.pdf>
- [30] P. Lison and A. Kutuzov, "Redefining Context Windows for Word Embedding Models: An Experimental Study," no. May, pp. 284–288, 2017. [Online]. Available: <http://arxiv.org/abs/1704.05781>
- [31] C. Boswell, *Addressing the causes of migratory and refugee movements: the role of the European Union*. UNHCR, 2003.
- [32] C. Maldonado Valera, J. Martínez Pizarro, and R. Martínez, "Protección social y migración: Una mirada desde las vulnerabilidades a lo largo del ciclo de la migración y de la vida de las personas," 2018.
- [33] A. Chihu Amparán and A. López Gallegos, "El" análisis de los marcos" en la obra de william gamson," 2004.