The Use of Supervised Learning Algorithms in Political Communication: Setting the future of Political Communication research?

Javier G. Marín – U. Granada (Spain)
Óscar G. Luengo – U. Granada (Spain)

Introduction

Since the 1970s, probably because of its progressive institutionalization, political communication has been articulated around certain methodological consensus, remaining close to positivist and quantitative approaches, which is particularly visible in the American academic context, and being the English the “lingua franca” in this area. All these elements have denoted a clear ethnocentrism. This stability has provided explanatory knowledge to (at present) so consolidated aspects as *agenda-setting*, *framing*, the formation of public opinion, the interaction between news coverage during political campaigns and voting behavior, or even the limits of information manipulation. Therefore, quantitative behavior studies have been the dominant paradigm in the field.

Traditionally, Political Science, Social Psychology and some other areas under the label Mass Communication studies, have given coverage to the analysis of the complex interaction between the world of communication and politics. However, the social and political dynamics in the past decade have meant a context without precedent regarding social relations, political processes, and communication interactions. Due to the sophistication of research and massive data collection techniques, other proposals have acquired special relevance as those from Neurosciences and Genetics (Luengo, 2016). Indeed, research in political communication has been affected by all this.

Some authors as Lance Bennet and Shanto Iyengar (2008) suggest that in recent years seems to exist a fracture in this building of Political Communication, indicating some skepticism in the ability of renovation of old concepts and regeneration of the methodological ways to approach the new scenarios, where the structures, processes and actors seem to act according to different parameters. A new horizon of the communicative processes has been defined, characterized by the "de-professionalization", with the irruption in the field of grass root citizen journalism as massive transmitters of information; fragmentation, by the multiplicity of media, new supports and new alignments in the map of the information companies; and the unpredictability, by the gradual complication in the modeling of these relationships and the consequent loss of the predictive capacity.

However, this environment of growing sophistication described coincides with a particular historical moment. On the one hand, greater technical resources have been developed in the field of automated/supervised analysis. On the other hand, have greater volume of available data accumulated for the analysis in the Social Sciences. In this new context, it seems that some of the classic tools employed in Political Communication, i.e. content analysis, have being displaced. Today, it is easy to verify hypothesis by analyzing the entire universe of study, without having to restrict the process to a sample. Also, armies of encoders are no longer required, and complex processes in order to test the reliability of the fieldwork is not necessary anymore. The use of algorithms, big data, neural networks, the adaptation of the techniques of machine learning or artificial intelligence to new fields of knowledge, is a good example of these new forms of research in Political Communication, without a doubt, they are involving going one step forward in the development of the study of, for example, the impact of the communication campaigns, the effects of the generation of frames in the news coverage, or the way in which citizens are informed.

In addition, the scope of comparative studies is acquiring a new dimension with these innovative research strategies, since some of the traditional barriers such as language skills or the sample size are significantly minimized. Therefore, the potentiallity is undeniable and makes possible the implementation of proposals based on comparative studies, such as the pointed out by Hallin and Mancini (2004) which became a constant reference of analytical framework.

This paper aims to explore the projection of the incipient use of machine learning (algorithms) in research in Political Communication. This field has developed in recent times some criticism due to the fog in which researchers have been tucked to give new answers. Needless to say, there is a need in this field of relocating objects of study in a context of increasingly fragmented audiences, extremely changing communicative and political institutions and actors readapting to them. The use of this type of research techniques may involve a decisive event: crossing a threshold in the improvement of traditional methodological strategies in Political Communication in general, as for example in agenda setting or framing research.

Machine learning

Machine learning is the subfield of computer science that, according to Arthur Samuel, gives "computers the ability to learn without being explicitly programmed." (Koza, Bennett, Andre & Keane, 1996). Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), learning to rank, computer vision, etc[1].

In Mathematics, Computer Science, and related disciplines, an algorithm is a well-defined, orderly and finite operations list that allows to find the solution to a problem. It can also be defined as the description of a pattern of behavior that is expressed using a finite repertoire of actions and basic information, identified, well understood and achievable. This repertoire is called lexicon (Scholl & Peyring, 1991). In basic terms, an algorithm is not more than a set of serialized operations. What is, then, the novelty? This is not in the existence of algorithms, known since ancient times, but in the ability to create very complex algorithms due to the possibilities offered by the exponential increase in computing capacity. i.e. no longer the lack of having a supercomputer is a limit to apply these tools. And, if in the 1950s the first computers allowed the use of new investigative techniques such as survey and its scientific treatment by statistics, at present machine learning may be a candidate to establish a new a set of tools that complement the existing ones.

Hence, in our case, we can say that an algorithm establishes a sequence to classify data into two or more groups based on conditions laid down in the own algorithm (case of non-supervised algorithms as the Latent Dirichlet Allocations, LDA) or by the user (case of supervised algorithms as SVM).

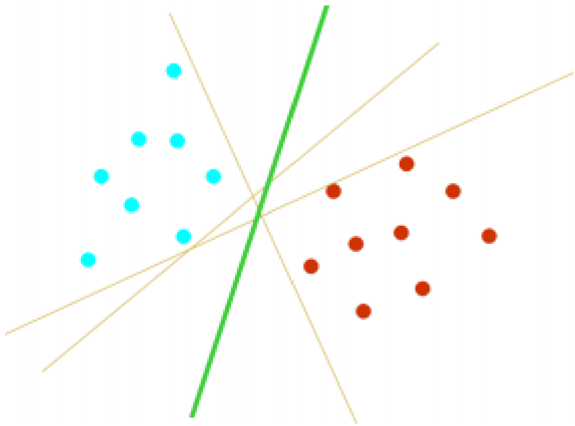
1 See, for example, the works in fields like Genetics (several authors, 2017), to detect plagiarism (VVAA, 2013), the cataloguing of images (Chum, Philbin, and Zisserman, 2008), etc.

Among the myriad of existing algorithms, we have chosen to work with Support Vector Machines. SVM is particularly fitted to work with texts due to being one of the best tested algorithms in industry and in the academic world (Joachim, 1998). There are thousands of works using it to measure sentiment in Twitter, for example, or in other social media (Facebook, Instagram, etc.). It is also better than neural networks for long texts (such as news stories) and uses much less computing power than other algorithms[2]. It is also a supervised algorithm, which means that we can train it to classify information using our variables.

SVM is based on the idea that any linear model is valid for classification if the classes are linearly separable, sufficing to find a hyperplane that discriminate both sets, i.e., any regression technique can be used for classification if we separate a sample into two groups: one first for training, where are assigned values of 1 for examples of one kind and 0 for the rest; and another group where the value of the regression is calculated and assigned to the corresponding class, comparing it to the value given by the investigators (measuring, thus, effectiveness). Of course, linear models can only obtain linear decision boundaries, there is the problem about how to use them for problems that are not linearly separable, such as those we find usually in the Social Sciences. In this case we can apply a non-linear model (increasing the complexity of the model), or also to build a linear classifier for a transformed space where a linear border may represent a nonlinear boundary in the original space. In other words: If we increase the number of space dimensions to sort, we can apply a linear model to separate classes while maintaining the simplicity of the model and, at the same time, serving for initially non-linear problems.

2 Especially less than other popular algorithms like Maximum Entropy, Random Forest and most Convoluted Neural Networks.

Chart 1 Optimal separating Hyperplane

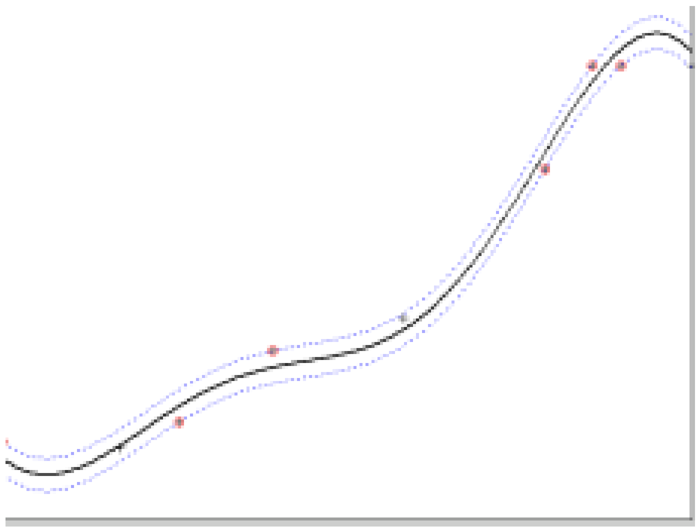


Source: Gunn, 1998: 11

Unfortunately, the universes to study are not usually presented in idyllic two-dimensional cases but rather an SVM algorithm should treat with a) more than two predictor variables, b) nonlinear curves of separation, c) cases where the data sets cannot be completely separated, d) classifications in more than two categories. But SVM is an algorithm designed to find linear classifiers in transformed spaces being highly efficient in nonlinear cases. What does it make? It creates a hyperplane of maximum margin, i.e., a *straight line* (if a two-dimensional space, which it is not) that separates two (or more) classes and is located at the maximum distance of both. How does it make it? If the data are not linearly separable we can use non-linear applications, which are called *kernel*, of which there are several (polynomial, gaussian, radial basis, perceptron and sigmoidal[3]).

3 Each one of them have different parameters. See, for example, Cherkassky & Ma (2004).

Chart 2 Radial basis function regression



Source: Gunn, 1998: 35

Even, mistakes are admitted into the model through a constant (c), limiting the effect of data not classifiable in the final model. The value of C is usually given experimentally. Consequently, it tends to be 0 and increases as you need greater reliability of the model without harming the classification.

Algorithms and Framing research

The potentiality of those research dynamics in the traditional study of process of framing is out of question. The development of Political Communication shows that the scholar attention to not only the theoretical implications of framing, but also to the methodological assumptions, has been prominent within the field. A macro approach to framing that examines media frames as outcomes of journalistic norms or organizational constraints, assumes that individuals cannot understand the world fully and consequently actively classify and interpret their life experience to make sense (Scheufele, 2000: 300-301). The individual´s reaction to sensory information therefore depends on schemes of interpretation called primary frameworks (Goffman, 1974). The most important implication for the field of Mass Communication research, then, is that there are various ways of looking at and depicting events in news media that depends on the framework employed by the journalists (Scheufele, 2000: 301). Hence, there is an enormous potentiality of connection between mass media coverage of political events and the framework individuals use to give interpretation to those events.

Most research in this specific realm of Political Communication has been elaborated in the last decades using content quantitative analysis as the main methodological technique. The limits of these interesting and useful resources are well known and have mainly to do with the problems coming from the subjectivity in the coding process as well as the restrictions in its consistency (reliability tests).

How to combine machine learning techniques with text analysis? Algorithms work with vectors, so the first thing is to transform documents into vectors, i.e, to a matrix of computable data (after all, a news story is a form of non-structured data).

One of the central problems in the analysis of texts, (text mining or natural language processing) is how to determine what a document is. One of the traditional approaches has been to focus on the *words*. Indeed, on many occasions the frequency of terms (tf) is used to determine the importance of a word in a text. However, there are many words whose usage frequency really tells us little, as "the" or "is", "the", "it", etc. Normally, to limit this problem are lists of words that are removed from the corpus (call *stopwords*). Another way to solve the problem, more sophisticated, are the *inverse document frequencies* (itf), which decreases the weight assigned to a word for the more common and increases for those that are least used in a series of documents.

Regarding the frequencies of terms, the main step is to create vectors of terms, where the frequencies the unit of analysis. The interesting thing, of course, is that these matrices are also inverse, calculating not only the presence of terms, but the absence of other terms that are present in other units of analysis (in our sample newspaper articles), according to those values the matrix is weighted in one direction or another. The algorithm tries to classify the data sets to homologate it to points on a plane (the matrix displayed as a spatial dataset) and it is then when establishing the classification.

Tf-itf has limitations such as not considering the synonyms and forms in the plural or in different genus (Ramos, 2003). The latter is easy to solve through the elimination of ambiguity in the words with the use of libraries intended for that, called stemming.

Examples

We have used this technique in order to illustrate some of the fields where SVM is particularly useful. They will be presented in the conference.

Conclusions

The first conclusion we want to mention here is obvious: machine learning techniques are useful for researchers in Social Sciences. We have demonstrated that to create a research on framing is possible without human coders at a fraction of the cost and the time. After all, it took the computer less than an hour to code the whole sample (after the authors coded 342 news stories "by hand"). Of course, it is not the first time this is achieved, and the scientific literature is full of quantitative analysis on framing. However, most of them used only word frequencies or other similar techniques (similarity analysis, word distance, etc.). All of those techniques had similar problems when approaching the real meaning of those words (like "terrorism"). Machine learning techniques, at least those based on supervised algorithms, are somewhat different. First, because they do not rely solely on word frequencies, since each algorithm weigh frequencies and inverse frequencies; and, second, because the researchers *teach* the algorithm qualitatively. And this algorithm tries to *learn* how to separate those groups created by the researchers. This implies that the amount of training is correlative to the complexity of the task. That is the reason behind choosing such a straightforward issue for this paper. In some cases, we are sure that the complexity of frames is going to make the technique unfeasible. But, in other cases can be an easy technique to use. But, since there are so many algorithms and other similar tools, maybe is just a matter of choosing a different one, which constitutes a nice source of research for the coming years.

The most important consequence, at least from our perspective, is the *black box* effect. Those algorithms are so complex that it is needed a lot of knowledge to know what they really do. It can be a minor problem in supervised algorithms, like SVM, since it is the researcher who tells the tool what to look for. But a major problem in non-supervised algorithms (like LDA). However, the danger of using tools without a complete understanding is still there and can be a real issue when dealing with big databases where a small mistake could lead to a big error. An example can be the parameters of each kernel in SVM. Most kernels have different parameters according to the mathematics in use. For example, we used a radial basis kernel in this research. Radial basis kernels have two main parameters, C, already explained, and *gamma*. Both parameters have default values but if we change them the effectivity of the model changes without knowing the exact changes in it. Furthermore, "the correct choice of kernel parameters is crucial for obtaining good results, which practically means that an extensive search must be conducted on the parameter space before results can be trusted, and this often complicates the task" (Meyer, 2007: 8).

To conclude, although challenging, the authors believe that machine learning techniques are going to determine the future of political communication. Therefore, the academic community should pay them more attention to the dynamics mentioned before, as many other fields of study are doing already.

References

Bennett, W. Lance & Iyengar, Shanto (2008). A new era of minimal effects? The changing foundations of Political Communication. *Journal of Communication*, *58*, pp. 707–731.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 27.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, *17*(1), 113-126.

Chum, O., Philbin, J., & Zisserman, A. (2008, September). Near Duplicate Image Detection: min-Hash and tf-idf Weighting. In *BMVC* (Vol. 810, pp. 812-815).

Churchland, Patricia (2008) "The impact of neuroscience on Philosophy", *Neuron* Vol. 60, Issue 3, pp. 409-411.

Collins, A. (2007). *Contemporary security studies*. Oxford: Oxford University Press.

Cong, Y., Chan, Y. B., Phillips, C. A., Langston, M. A., & Ragan, M. A. (2017). Robust inference of genetic exchange communities from microbial genomes using TF-IDF. *Frontiers in Microbiology*, *8*.

Forman, G. (2008, October). BNS feature scaling: an improved representation over tf-idf for svm text classification. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 263-270). ACM.

Grün B and Hornik K (2011). "topicmodels: An R Package for Fitting Topic Models." *Journal of Statistical Software*, **40**(13), pp. 1–30. doi: 10.18637/jss.v040.i13.

Gunn, S. R. (1998) Support Vector Machines for Classification and Regression. *University of Southhampton*.

Karpf, D.A., Kreiss, D., & Nielsen, R.K. (2014). A new era of fieldwork in Political Communication research? In L. Lievrouw (Ed.), Challenging Communication Research (). New York: Peter Lang, pp. 43-57

Scholl, P. y Peyrin, J. P. (1991). *Esquemas algorítmicos fundamentales: secuencias e iteración*. Masson.

Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170.

Luengo, Óscar G. (2016). "Comunicación Política: de la propaganda masiva a las neurociencias", en Colino, César et. al. (comp.) *Ciencia Política: Una aventura Vital*, Valencia: Tirant Lo Blanch, pp. 721-740.

Meyer, D., & Wien, F. T. (2001). Support vector machines. *R News*, *1*(3), 23-26.

Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of statistical software*, *25*(5), 1-54.

Nielsen, Rasmus K. (2014). Political Communication Research: New Media, New Challenges, and New Opportunities, *MedieKultur*, 56, pp. 5-22.

Onghena, Y. (2015) "Migrant or refugees" *CIDOB Opinion 355*, 2015

Potthast, M., Hagen, M., Gollub, T., Tippmann, M., Kiesel, J., Rosso, P., ... & Stein, B. (2013). Overview of the 5th international competition on plagiarism detection. In *CLEF Conference on Multilingual and Multimodal Information Access Evaluation* (pp. 301-331). CELCT.

Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*.

Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of Political Communication. *Mass Communication & Society*, *3*(2-3), 297-316.

UNCHR (2016) *Global Trends 2015. Force Displacement in 2015*, Geneva: UNCHR publications (http://www.unhcr.org/576408cd7.pdf)

Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF* IDF, LSI and multi-words for text classification. *Expert Systems with Applications*, *38*(3), 2758-2765.