Determining senatorial voting archetypes through hyperspectral unmixing

Crizzia Mielle M. de Castro* and May T. Lim

National Institute of Physics, University of the Philippines Diliman, Philippines

*Corresponding author: cdecastro@nip.upd.edu.ph

Abstract

Dimensionality reduction and hyperspectral unmixing techniques are commonly applied to process and analyze hyperspectral satellite images. In this work, we apply hyperspectral unmixing on the partial results of the Philippine senatorial elections last 2016 to determine possible voting archetypes. HySime was used for dimensionality reduction and determining the number of archetypes. VCA and MVSA, and SUnSAL were used to estimate the mixing matrix $\mathbf{M}$, and the abundance matrix $\mathbf{S}$, respectively. From the six archetypes obtained, the first archetype best describes votes for the 12 winning candidates. The second and third best describe the candidates that gained the most media coverage. The fourth archetype could represent celebrity voting. The fifth archetype could represent voters that choose candidates who have already run locally.

Keywords: dimension reduction, hyperspectral unmixing, archetypes, senatorial elections

1 Introduction

In the 09 May 2016 Philippine National Elections, Filipinos voted for a new president and vice president, 12 senators, one district representative and party list representative, and provincial/city/municipal officials. It was also the third time the Philippines used an automated voting technology since 2010 [1]. Using partial unofficial results tallied by NAMFREL, we aim to determine and analyze possible senatorial voting archetypes using hyperspectral linear unmixing, a technique commonly applied to satellite images.

The Philippines can be compared to a single hyperspectral satellite image where each pixel is one clustered precinct. Each pixel in a satellite image has a corresponding emission spectrum. Each clustered precinct also an intensity profile corresponding to the fraction of votes each candidate garnered. Philippine clustered precincts and hyperspectral satellite images are structurally similar, so it would make sense to apply the same algorithms to both.

Linear unmixing assumes that the emission spectrum of each pixel is just a linear combination of archetypes. Simple examples of possible archetypes for a satellite image are trees, land, and water. Similar to this, the fraction of votes in each clustered precinct may be described using a linear combination of certain voting archetypes. A major assumption in this work is that voting archetypes exist.

Subspace identification is an important first step to hyperspectral unmixing. Hyperspectral Signal identification by minimum error (HySime) [2], developed by Bioucas-Dias and Nascimento, is used for dimensional reduction and determining the number of archetypes. HySime estimates the signal and noise correlation matrices, and then selects the eigenvalue subset that best represents the signal subspace.

Hyperspectral unmixing algorithms aim to estimate $\mathbf{M}$ and $\mathbf{A}$ in 1:

$$\mathbf{Y} = \mathbf{MA} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y}$ is the observed spectral matrix, $\mathbf{M}$ is the mixing matrix, $\mathbf{A}$ is the abundance matrix, and $\mathbf{N}$ is the noise matrix [2–5]. The solution is subject to the positivity constraint $\mathbf{A} \succeq 0$ and the sum constraint $\mathbf{1}_p^T \mathbf{A} = \mathbf{1}_n$.

Minimum Volume Simplex Analysis (MVSA) [4] is a geometric linear unmixing algorithm without the pure pixel assumption used to estimate $\mathbf{M}$. MVSA fits a minimum volume simplex to the data by constraining the abundance fractions. The vertices of this simplex corresponds to the endmembers or archetypes of the data. Identifying these vertices estimates the mixing matrix.

Vertex Component Analysis (VCA) [3] is also a geometric linear unmixing algorithm, but with the pure pixel assumption. VCA exploits the fact that the affine transformation of a simplex is also a simplex. It projects data onto a direction orthogonal to the subspace spanned by the current endmembers until all possible endmembers are exhausted.

Sparse Unmixing via Variable Splitting and Augmented Lagrangian (SUnSAL) [5] is a sparse regression algorithm that fits the observed data $\mathbf{Y}$ with linear mixtures of known spectral signatures $\mathbf{M}$. SUnSAL is used to estimate the abundance matrix **A**. It decomposes a complicated problem to a sequence of simpler ones. An augmented Lagrangian method is used to solve the resulting constrained problem.

2 Methodology

The partial election results as of May 9, 2016 5:42 pm were retrieved from http://elections.org.ph. We created a matrix with rows corresponding to the results of a precinct and columns corresponding to the fraction of votes of each senatorial candidate with respect to the total number of ballots in said precinct. This matrix is $\mathbf{Y}$ in Eq. 1.

The MATLAB code for HySime, VCA, MVSA, and SUnSAL provided by J. Li et al. [4] were used to determine the number of archetypes (HySime), estimate the mixing matrix (VCA and MVSA), and estimate the abundance matrix (SUnSAL).

The resulting archetype signatures from the mixing matrix were presented as bar graphs ordered from the candidate with most votes to the candidate with the least. A histogram of the corresponding normalized weights per archetype obtained from the abundance matrix was also presented.

3 Results and Discussion

Using HySime, the sentorial voting data set was determined to have six archetypes (Fig. 1). A summary of candidates whose intensities are greater than or equal to the archetype mean plus one standard deviation is presented in Table 1. Majority of the candidates that reach the cut-off are those in the top 20, while those below may be considered as nuisance candidates. The only exception occurs in the 2nd archetype.

The first archetype best describes votes that represent the winning candidates, since majority of the top 12 candidates (Villanueva to De Lima) have relatively high intensities.

In the second and third archetypes, Manny Pacquiao and Martin Romualdez have very high intensities (> 1) compared to the other candidates. These two candidates received the most printed coverage from news outlets during the first two weeks of the campaign period last 2016 [6]. Although, it is important to note that half of the news articles about Romualdez came from the Manila Standard, a news company he owns.

Looking at the three candidates with the highest intensities in the fourth archetype, we have Vicente Sotto III, Manny Pacquiao, and Mark Lapid- best described as celebrity voting, since these three candidates are either Philippine celebrities or are closely associated with other celebrities [7]. Francis Pangilinan, the fifth highest candidate, is commonly associated with his wife, Sharon Cuneta, another Philippine celebrity. Francisco Domagoso, popularly known by his screen name *Isko Moreno*, also reached the cut-off.

Martin Romualdez and Jericho Petilla have the highest intensities (> 1) in the fifth archetype. Romualdez served as representative of the 1st district of Leyte for three terms, while Petilla served as the governor of Leyte from 2004-2012. The purest pixel or precinct corresponding to this archetype was found to be a clustered precinct in Leyte.

4 Conclusions

Our use of hyperspectral image processing techniques on election results to determine possible senatorial voting archetypes was motivated by the empirical notion that there are voting patterns in Philippine elections. Our results, so far, seem to be supporting this notion.

Acknowledgments

M.L. acknowledges the support provided by the ICTP Associate Scheme. The authors received support from the Commission on Higher Education-PCARI project DARE: Data Analytics for Research and Education (IIID-2016-006).

References

[1] M. Palatino, 7 things to know about the 2016 Philippine elections (2016), last accessed 09 April 2019, https://thediplomat.com/2016/01/7-things-to-know-about-the-2016-philippine-elections/.

[2] J. M. Bioucas-Dias and J. M. P. Nascimento, Hyperspectral subspace identification, *IEEE Transactions on Geoscience and Remote Sensing* **46**, 2435 (2008), ISSN 0196-2892.

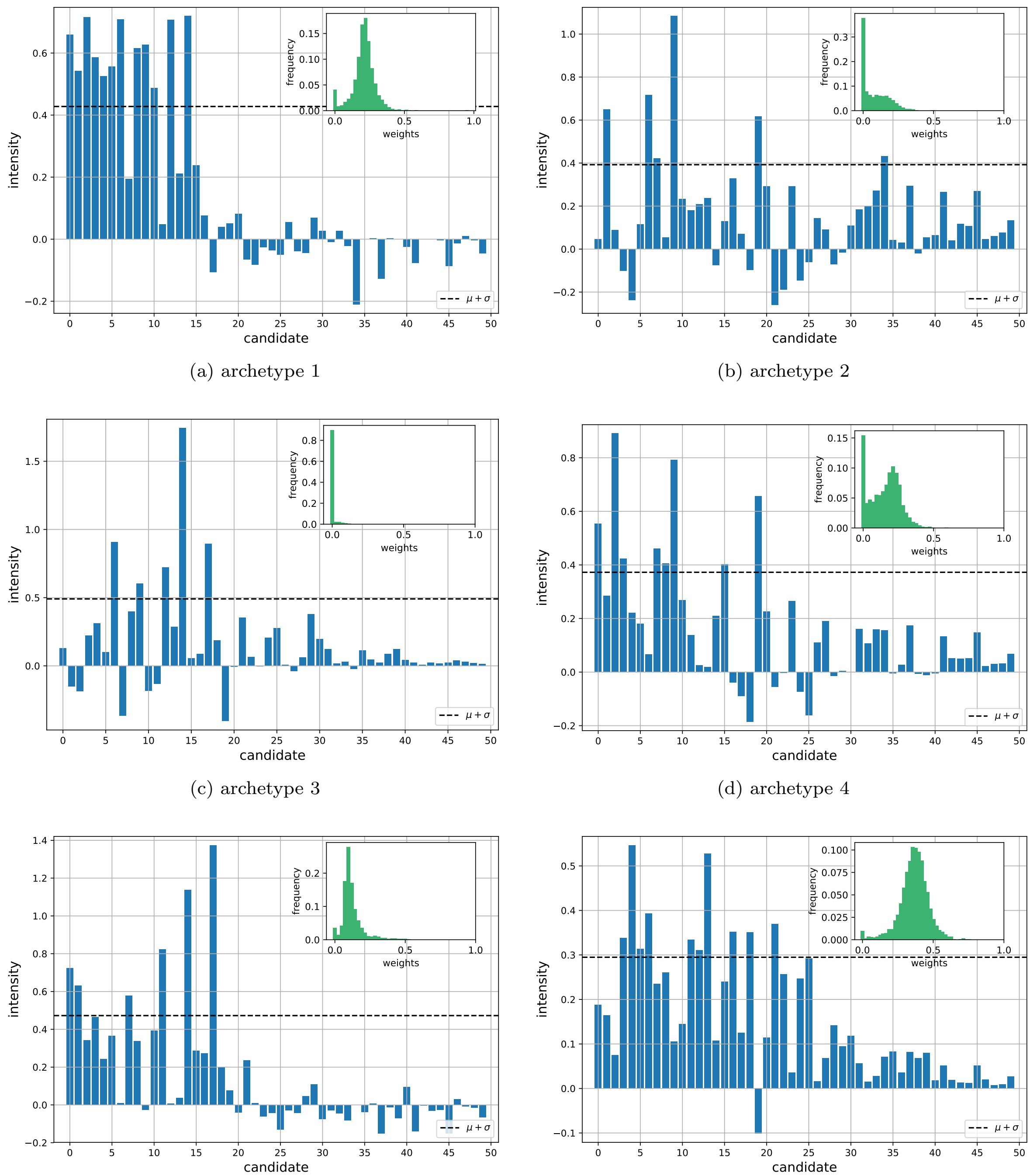
[3] J. M. P. Nascimento and J. M. B. Dias, Vertex component analysis: a fast algorithm to unmix hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing* **43**, 898 (2005), ISSN 0196-2892.

[4] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing* **53**, 5067 (2015), ISSN 0196-2892.

[5] J. M. Bioucas-Dias and M. A. T. Figueiredo, in *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (2010), 1-4, ISSN 2158-6268.

[6] Center for Media Freedom and Responsibility, Candidates for senate and party-list: Hardly noticed (2016), last accessed 23 April 2019, https://cmfr-phil.org/elections-2016/candidates-for-senate-and-party-list-hardly-noticed/.

[7] Rappler, Celebrities running, supporting candidates in the 2016 national elections (2015), last accessed 23 April 2019, https://www.rappler.com/entertainment/news/109256-celebrities-running-support-candidates-2016-national-elections.



(a) archetype 1

(b) archetype 2

(c) archetype 3

(d) archetype 4

(e) archetype 5

(f) archetype 6

Figure 1: Voting archetypes with normalized distribution of weights

Table 1: Candidates with intensities $\geq \mu + \sigma$

		archetype					
	candidate	1	2	3	4	5	6
0	Villanueva	/			/	/	
1	Drilon	/	/			/	
2	Sotto	/			/		
3	Lacson	/			/		/
4	Gordon	/					/
5	Hontiveros	/					/
6	Zubiri	/	/	/			/
7	Pangilinan		/		/	/	
8	Gatchalian	/			/		
9	Pacquiao	/	/	/	/		
10	Recto	/					
11	De Lima					/	/
12	Tolentino	/		/			/
13	Osmena						/
14	Romualdez	/		/		/	
15	Domagoso				/		
16	Guingona						/
17	Petilla			/		/	
18	Colmenares						/
19	Lapid		/		/		
20	Manzano						
21	Romulo						/
22	Ople						
23	Lacsamana						
24	Belgica						
25	Alunan						
26	Gadon						
27	Langit						
28	Kapunan						
29	Pagdilao						
30	Santiago						
31	Napenas						
32	Chavez						
33	Montano						
34	Ambolodto		/				
35	Bello						
36	Palparan						
37	Kiram						
38	Liban						
39	Cam						
40	Paez						
41	Albani						
42	Maganto						
43	Montano						
44	Arquiza						
45	Ali						
46	Baligod						
47	Valeroso						
48	Dorona						
49	Kabalu						