ISBN: 978-1-948012-17-1

Ethics and Information Technology (ETIT)

DOI: <http://doi.org/10.26480/etit.02.2020.138.141>

CrossMark

MODELING OF OPINION FORMATION UNDER INFLUENCE OF SOCIAL MEDIA FOR ANALYZING HISTORICAL EVENTS WITH BENGAL PARTITION AS CASE STUDY

Priyadarsini Sinha^a, Milan Mukherjee^b, Souvik Roy^c, Abhik Mukherjee^a^a CST Dept, IIST Shibpur^b Hyland Software^c IT Dept, IIST Shibpur*Corresponding Author Email: georgenelson5006@gmail.com

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 25 October 2020

Accepted 26 November 2020

Available online 03 December 2020

ABSTRACT

Error correcting codes are important for communication systems to retrieve messages transmitted over noisy channels. Recent advancement of low density parity check (LDPC) codes has enabled reliable communication in highly noisy environment of deep space or underwater. The LDPC decoder algorithm employs localized neighbourhood based parity check equations to improve upon the error correcting capability. It is shown in this work that such principle can be implemented on bi-partite graph to simulate the connectivity among people of the society who form opinion through influence from their connected neighbourhood. With the advent of social media, this concept of neighbourhood has broken the localized boundaries so that influencing opinions can be far spread. Historical event like Partition of Bengal in 1947, when revisited in light of this model, provides alternate course of history.

KEYWORDS

online social media, graph algorithm, error correcting codes.

1. INTRODUCTION

Error correcting codes is a well studied subject in communication theory (Lin and Costello, 1999). Low Density Parity Check (LDPC) codes are one of the best error correcting codes (apart from Turbo codes) till now for practical applications like wireless communications and deep space missions. The decoders of LDPC codes are, by nature, some generic iterative decoders which can be used to decode any block code for a given channel. Iterative decoders are different from the standard algebraic decoders that settle in one pass for the codeword situated at global minimum distance from the received word. Rather, they care for the connectivity established by the individual bits of the parity check equations and initiate corrections so that all parity check equations are satisfied. The process does not often terminate in one pass and may require a good number of iterations before convergence of the parity check equations. The process may terminate without being decisively able to resolve the ambiguity even after running large number of iterations.

In the present work, individuals of the society are considered to be represented by the bits who participate in parity check equations where they try to settle to some convergent opinion by changing their own opinion based on majority of opinions they receive.

The binary content of the bit is suitable to implement the two possible states of being agitated (1) or at peace (0) with the issue in question. The codeword has two parts, the message bits and check bits. Here the check bits are considered to be the influencing members of society who matter to the individuals while forming their opinion through the satisfying of parity check equations. Analogy may be drawn with how individual opinions can converge to some opinion about some societal issue.

The main contention of this paper is to show how more influencing from remotely located individuals could dominate over local influences or global trends resulting in convergence to different state. This arguably is one of the main feature today with the exodus of social media in the life of an individual. For this, we have chosen the historical event of Bengal partition (Chatterjee, 1994) that was decided largely through localized political interest groups. We have argued that an alternate course of history could have been arrived at if social media was present in that era. This is hypothetical reasoning but nevertheless useful for understanding how opinions form in the present and what the future holds.

2. ERROR CORRECTING LDPC CODES

Early variant of LDPC code was discovered by Robert in early 1960s (Gallager, 1962). Largely suppressed from further research due to technical limitations. Tanner's graph theoretic view (Tanner, 1981) on block codes presented in 1981 was also ignored. Late 1990s came up with the same idea of LDPC code when a researchers and other eminent researchers found out these codes independently while investigating on graph based iteratively decodable codes (Davey et al., 1998; MacKay, 1999).

An LDPC code is defined with respect to its parity check matrix. They are a subclass of block codes in general. The term low density signifies a sparse parity check matrix in terms of the number of 1s present in the matrix. Due to this sparseness, LDPC codes possess great advantages in terms of efficient decoder implementation and storage and can approach the Shannon Limit very effectively. It has been implemented in various practical standards such as IEEE 802.16, IEEE 802.20, IEEE 802.3, DVB-

Quick Response Code	This paper was presented at	Access this article online	
	International Conference on Contemporary Issues in Computing (ICIC-2020) - Virtual IETE Sector V, Salt Lake, Kolkata From 25th-26th July 2020	Website: www.intelcomp-design.com	DOI: 10.26480/etit.02.2020.138.141

RS2 and many others. Such is the advantage that now researchers are even investigating non-binary LDPC codes and their applicabilities.

An (N, w_c, w_r) LDPC code with a parity check matrix H of size $(M \times N)$, where the Hamming weight of each column and each row are two fixed integers, w_c and w_r respectively with $w_r > w_c > 2$. Since the matrix is sparse in nature, so $w_c \ll M$ and $w_r \ll N$. The total number of 1s = $w_c \times N = w_r \times M$. Thus, $M = \frac{(N \times w_c)}{w_r}$ must be an integer. This also implies that every code bit participates in w_c parity check sums and each parity check sum involves w_r code bits. The density of this parity check matrix is defined as $\psi = \frac{w_r}{N} = \frac{w_c}{M}$. This helps to understand, both intuitively and mathematically, how much sparse H really is.

R. M. Tanner in 1981 showed the existence of a bijection between the parity check matrix of any block code and a bipartite graph (Tanner, 1981). This graph is now referred to as a Tanner graph. Each row of a $(n-k) \times n$ parity check matrix H of a block code denotes a parity check sum f_i for all $i = 0, 1, \dots, (n-k)$. Each column denotes a particular position, c_j for all $j = 0, 1, \dots, n$ of the code word with block length n . Whenever $H[i, j] = 1$, it indicates that the i^{th} parity check sum affects the j^{th} code bit position or the j^{th} code bit. From this Tanner constructed a bipartite graph where the two sets of vertices/nodes consists of the parity check sums and code bits as nodes respectively. (f_i, c_j) is an edge in the graph whenever $H[i, j] = 1$.

Construction of the generator matrix can be initiated by forming a sub-matrix H_0 of size $M \times N$ as follows : each row of H_0 has w_r 1s in such a way that the i^{th} row of H_0 contains all its w_r 1s in the columns $[(i-1).w_r + 1]$ to $i.w_r$. After this, the entire parity check matrix can be constructed by stacking any $(w_c - 1)$ column permutations of H_0 . Here $\Pi_k(H_0)$, $k \neq 1$, denotes a sub-matrix obtained by permuting the columns of H_0 randomly while $\Pi_1(H_0) = H_0$. If each permutation is chosen randomly, it can sometimes produce good code. The average minimum distance of the code increases with N for $w_c \geq 3$. Since the so designed LDPC code is not necessarily linear code, $\text{Rank}(H) \leq M$. Therefore dimension of the code is $\text{Dim}(C) \geq (N - M)$ or $K \geq (N - M)$. Hence rate of the code $R = \frac{K}{N} \approx 1 - \frac{w_c}{w_r}$. Hence the actual rate of the code is higher than the H matrix suggests as the designed rate. For practical LDPC codes N is taken to be quite large, like $N > 5000$, while the column weight is held around 3 or 4, so the density of 1s in the matrix is quite low.

3. ITERATIVE DECODER ALGORITHM

Iterative decoding principle is employed for decoding of LDPC codes. Unlike one shot method used in algebraic decoding of linear block codes, this uses multiple iterations capable of eliminating high noise. A study introduced a simple Bit-Flipping algorithm that corrected a single bit in every iteration (Gallager, 1962). If multiple bits are allowed to flip in one iteration, then the algorithm is called a Multi-Bit-Flipping (MBF) Algorithm.

In every iteration of this algorithm the code word bit(s) with the highest number of a metric are flipped. Metric is characterized by the type of algorithm: Hard-Decision and Soft-Decision. Hard decision metric is the number of unsatisfied parity check equations. Soft decision metric consists both the number of unsatisfied and satisfied parity check sums, since reliability information is present. An algebraic decoding algorithm in this context can only correct up to its random error correcting capability. But due to the sparse nature of H , the MBF algorithm may correct error patterns whose number of errors exceeds the error correcting capability of the code. Since the present work relies on hard decision message passing only, the details of soft decision algorithm that exploits reliability information is not introduced.

Message Passing Algorithm (MPA) in context of LDPC codes can be visualized more easily if we refer to the Tanner graph of the associated code. Messages are passed between message nodes and check nodes along the edges of the Tanner graph. The information that is sent from a node v_i to v_j in MPA should not get influenced by the node v_j i.e., in other words, the data that is being sent to a node must depend only on the other nodes. This type of message is called an "extrinsic message" or "extrinsic information".

The generic message passing structure is as follows :

- 1) Initialize the decoder by sending the received values of the message nodes to check nodes.
- 2) In each check node, an extrinsic message for each neighbor variable node is calculated and sent.
- 3) Each message node calculates a new value for itself depending on the previous received values and the information from the check nodes. The output is then checked by all the parity check equations and if all of them agree, decoding stops. Otherwise, an extrinsic message for each neighbor check node is calculated and sent.
- 4) Repeat steps 2 and 3 until a maximum number of iterations is reached.

4. ANALOGY WITH SOCIETAL OPINION FORMATION

4.1 Dynamics of opinion formation

One large binary string with a number of zeros and ones can be considered as a received word. Each bit can be considered as an individual having a binary state of mind where one represents agitated state and zero represents peaceful state. Some of these individuals may be considered to be influential and thus considered as check bits. The bipartite graph can then be constructed with edges drawn between individual nodes on one side and the influencing nodes on the other side.

Analogy can be drawn between this construction and the MPA decoder of LDPC code discussed in the previous section. Each individual node is connected to a number of influencing nodes.

Thus they receive message of being in agitated or peaceful state from each such check node. Then they arrive at the majority and accordingly retain or flip the bit, implying that they may go from agitated to peaceful state or vice versa based on collected majority opinion.

This process may continue over a number of iterations with the terminating condition described as reaching all zero codeword or exceeding a certain number of iterations. Check bits also appear as individuals thereby allowing flexibility of opinion for all.

The main argument is that with localized influence it is not possible to get back from a number of ones and zeros in the received word back to the all zero codeword. Along similar lines, unless the individual nodes get connected to some remote influencing nodes, it is not at all possible for the MPA of the iterative decoding algorithm to get to the all zero state. One important flavour about the success of MPA is that the parity check matrix is sparse yet it provides a spread of the check nodes.

4.2 Simulation results for message passing

The 20, 3, 4 code has been tested by creating bit strings where the all zero code word resembling peaceful state is contaminated with number of 1s (Gallager, 1962). The message passing based iterative decoder algorithm is run on these simulated received words.

Success and failure along with the number of 1s corrected to 0s is reported in Table 1 along with number of iteration needed to achieve the success. It can be seen that for some cases, even after 5000 iterations, the error does not get corrected. It depends on the distribution of the error with respect to the generator matrix used for the code. The corresponding bipartite graph is shown in Figure 1.

Figure 1: Bipartite graph showing the interaction among the nodes

The generator matrix is such that for the first five rows the influence is localized, so that f_i influences c_{4i+1} to c_{4i+4} . However, apart from f_1 , c_1 is also influenced by f_8 and f_{14} .

Likewise, c_2 is influenced by f_1 , f_9 and f_{12} . In this manner, the localized influence is perturbed and the maximum likelihood or minimum distance decoding philosophy gets altered.

Results are tabulated in Table 1, clearly showing large number of additional errors or agitated opinions getting corrected. Some of these results are

Table 1: Iterative decoding success for 20,3,4 LDPC code

Iteration count	Bits corrected	Remarks
102	12	Success
18	9	Success
9	8	Success
73	12	Success
5000	3	Failure
5000	4	Failure
5000	5	Failure
5000	9	Failure

indeed encouraging from our present perspective. Large number of individuals in riotous state (1) can be brought to pacified state (0) through the message passing algorithm that relies upon flipping of node bit based on majority of its influencing bits.

It can be seen that often large number of iterations are needed, sometimes the process fails in the long term as well. But this brings out the importance of geographical spread of influencing bits beyond localised boundaries. We found that the system inevitably terminates with failure if influence is localized, check bits are less in number and the 1's are concentrated in some locality, it is never possible to get to the all zero codeword at all.

5. CASE STUDY OF BENGAL PARTITION

5.1 Historical perspective

The first attempt towards partition of Bengal by the British in 1905 was thwarted through peoples' struggle and forced to be nullified in 1911. But the second one by the same British rulers was rather invited - by the same people in a gap of only 40 years. A spate of communal riots spread across the entire territory of undivided Bengal; with Kolkata, Dhaka, Mymensingh, Chittagong all throwing in their bit; gradually became uncontrollable.

During the initial years, the communal riots of undivided Bengal largely originated due to economic oppression. The feudal economic structure of Bengal was such that the oppressing Zamindars or landlords of Eastern Bengal were mostly Hindus and their Ryots representing the peasantry were mainly Muslims. The signature of this class character dominated till 1926 as claimed by the historians (Das, 2010). Such crowd leader dichotomy however shifted during the 1930's when the British government were clearly instigating the riots, for example the Chittagong riots of 1930. Finally, in the 1940's the political parties came to the forefront of the riotous population.

With two nation theory and its inevitable logic of Bengal partition suddenly becoming the agenda of all the political outfits, this obviously dominated the discussion forums. Under such circumstances, opinion poll of the legislature representatives were held regarding the partition in a state where the democratic institution was also cut across communal lines as part of British policy of representation (Chatterjee, 1994). As expected, the opinion was biased, driven by narrow interests of the representatives and ignorant of public opinion altogether. Even, the decision of Eastern Bengal legislature was not binding on the West. Hence the voting pattern was highly communal and localized.

The localization can be understood from the distribution of minority population on either side of the border as shown in Figure 2(L) (Chatterjee, 1947). However, organized working class existed in the Railways and to an extent in the trade unions of Jute industry and in these spheres, there was a small spread cutting across the local barriers as depicted in Figure 2(R).

But this was too weak compared to the geographical spread enjoyed by the online social media of today. Recent Brexit debate has seen huge number of such cross votings mainly from Conservative party backbenchers and few Labour Party MPs as well in the British parliament, mainly because of various extraneous influences (Aidt et al., 2019).

Figure 2: Map showing minority population on either side of the border, indicator of local influence (Left) Map showing manufacturing industry on either side of the border, indicator of remote influence (Right)

5.2 Presence of social media

To revisit the Bengal partition, the influence and spread of social media in Bangla speaking regions is considered. Whatsapp or Facebook connectivity depends more on acquaintance like school or office, such friendship is sparse across the border. Friendship groups do exist but their presence is formal. For example, the information shared by Bangladesh India Friendship Association, a community in Facebook, are either visa related or about health care services.

We therefore looked for some social media that binds on language and discusses on convergent issues. Quora is one such platform and analysis of Quora discussion forums can be found in literature as well (Wang et al., 2013; Maity et al., 2018). So for the present work, Quora Bangla dataset has been taken and the spread of interactions on Quora Bangla posts amongst the Bangla speaking community across the border is explored.

Quora Bangla data was collected for 10 questions raised in the platform with 3006 responding users along with the users profile location. Figure 3 shows the binding of both sides of the partition. The nodes depict the user profile (blue for users from Bangladesh and red for users from West Bengal) and two nodes are connected via edges depending on *upvote* and *comments* response. The group attribute layout of the graph is shown depending on country attribute which shows how engrossed both sides of the partition are in this respect.

Typical location specific results are shown in Figure 3(R) which depicts response from Bangladesh users (A: Dhaka & Southwards; B: Mymensingh & Northwards; C: Chittagong & Eastwards; and D: Bangladesh other area) for the posts from West Bengal users and response from West Bengal (A: Kolkata & Southwards; B: Siliguri & Northwards; C: North East India) for the posts from Bangladesh users.

Figure 3: Left: Quora response graph with Attribute view; Right: Response statistics

5.3 Simulation perspective

We look back at the excitors and inhibitors of communal tension in terms of the propaganda on either side of the divide (Chatterjee, 1994). The bipartite graph based model can be useful with edge being drawn between groups of the population based on their roles as excitors and inhibitors. The graph would be sparse but with strong possibility of cycles. We consider binary opinion for the legislative representatives. As political groups, they employ local connectivity. Their initial state is highly biased as most Congress and Hindu Mahasabha legislators with majority located in West Bengal were in favour of partition (state 1) and the Muslim League

and Praja Krishak Party legislators with majority in East Bengal wanting to avoid partition (state=0). Purely local connectivity can never result in convergence to the all zero state.

Now, the LDPC code example discussed in previous section allows remote influence of check bits on code bits. In some form, social media was present, like the newspaper or leaflet to circulate opinions. But that is not adequate enough to guarantee the sort of remote connectivity needed to make the mark. Updating of the state at each node is based on the majority rule applied by considering all opinions that reach it. Now, some of the representatives are allowed remote connectivity outside their party opinion to an extent as per the permutation based construction of the parity check matrix described in previous section (see Figure 1). Iterative decoding can be employed and the system is able to reach sometimes the stable all zero state which implies thwarting of partition. Hence with the kind of influencing Quora Bangla provides, the chances of alternate courses of history might have been realized.

6. CONCLUSION

There is no way to go back in time and change the course of historical events. However, it is worthwhile to learn lessons from history and deconstruct the events in light of modernity to explore new possibilities. This work looks back at the Bengal partition in light of the advent of online social media,

This leads us to the conclusion that there was lack of enough communication amongst the people of undivided Bengal to the detriment of different options reaching them and influencing their decision. The probability of alternative discourse would have been higher in today's connected world. However, here only some random simulation perspective is given. To present more realistic picture of history, we have plans to incorporate the traits of legislators who actually participated in the process.

ACKNOWLEDGEMENT

This work received support of MHRD/ICSSR IMPRESS project.

REFERENCES

- Aidt T., Grey F, Savu A., 2019, The meaningful votes: Voting on Brexit in the British House of Commons. Public Choice (Springer).
- Chatterjee J, 1994, Bengal divided Hindu communalism and partition 1932-1947, Cam-bridge University Press.
- Chatterjee S. P., 1947, Bengal in maps, NATMO.
- Das S, 2010. Communal riots in Bengal 1905-47, Oxford University Press.
- Davey M. C., MacKay D. J. C., 1998, Low Density Parity Check Codes Over GF(q), IEEE Com. Letters, 2(6).
- Gallager, 1962, Low density parity check codes, IEEE Transactions on Information Theory, IT-8, 21-28.
- Lin and Costello, 1999. Error Correcting Codes, Prentice Hall.
- MacKay D. J. C., 1999. Good Error-Correcting Codes Based on Very Sparse Matrices, IEEE Trans. Information Theory, 45, 2, 399-431.
- Maity S.K., Kharb A., Mukherjee A., 2018. Analyzing the linguistic structure of question texts to characterize answerability in Quora; IEEE Trans. Comp. Soc. Syst. 5(3), 816-828.
- Tanner R. M., 1981, A Recursive Approach to Low Complexity Codes," IEEE Trans. Information Theory, IT-27, 533-547.
- Wang G., Gill K., Mohanlal M., Zheng H., Zhao B.Y., 2013. Wisdom in the social crowd: an analysis of Quora; WWW, pp. 1341-1352; ACM 2013.