International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
International Conference on Humming Bird (01st March 2014)

RESEARCH ARTICLE OPEN ACCESS

A new metric based on Motifs for Animal Recognition

D.Samson
Assistant Professor, Department of Computer Science and Engineering,
Cape Institute of Technology ,Levengipuram, Rajakrishnapuram P.O, Tirunelveli-627114,Tamil Nadu, India

Abstract
Recent years have witnessed new research interest in the study of network science, in domains like biological systems, social networks etc. Seminal works covering each of these systems have appeared in high impact journals like Nature, Science, etc. Unifying principles have emerged and helped in gaining new understanding in a domain by extending the understanding gained in other domains. These developments in network science open up possibilities in the research into image processing. We abstract the face image of different Canis lupus familiaris *(*dogs) as a network/graph, where the nodes/vertices correspond to rows/columns represented by the pixels and edges correspond to presence of any non-zero value present in the pixel represented by {row i, column j}. Network science research in biology defines motifs as recurring sub-graphs from which the network is built. They also argue motifs as simple building blocks of complex networks, offering a way to understand the basic functionality of a system. In this paper, we explore 120 Canis lupus familiaris (dog) breed characteristic images (face) from literature [Stanford dog dataset] for the study of motifs. We discover motifs within each characteristic and also interesting motif templates across them. We use this idea and propose a new low level classifier for vision to conclude that motifs could be used for animal recognition.
Keywords - Motifs, Motif profile, Face, Canis lupus familiaris

I.Introduction

Recent years have witnessed new research interest in the study of network science, in domains like biological systems, social networks etc. [Duncan J Watts , Newman MEJ]. Unifying principles have emerged and helped in gaining new understanding in a domain by extending the understanding gained in other domains [Boccaletti S et al]. Researchers in other areas have commented on the hesitation of researchers in complex engineering systems to look at their problems in the light of emerging ideas in complex systems in general. “Engineering should be at the centre of these developments, and contribute to the development of new theory and tools” [J.M. Ottino]; “Engineers seem a little bit indifferent as if engineering is at the edge of the science of complexity” [Zhi-Qiang Jiang et all].

The dictionary definition of complexity refers to – consisting of interconnected/interwoven components. Complexity of a system scales with the number of components, number of interactions, complexities of the components & complexities of interactions [Edward Crawley et all]. Biometric systems are considered complex systems. Complex animal characteristic (face) is represented by a large number of pixels. We abstract the face image image of different dogs as a network/graph, where the nodes/vertices correspond to rows/columns represented by the pixels and edges correspond to presence of any non-zero value present in the pixel represented by {row i, column j}.

Network science research in biology defines motifs as recurring sub-graphs from which the network is built. In biology, the analysis of network motifs has led to interesting insights in the areas of protein-protein interaction prediction [Albert L and Albert R] and analysis of temporal gene expression patterns [M Ronen et all , S.S. Shen-Orr et all]. Research in biology also argues motifs as simple building blocks of complex networks whose selection may possibly be one way to understand the basic functionality of a system. In this paper the words system and images are used interchangeably.

II.Motifs

Motifs are considered to be functional building blocks of a network. “Motifs are recurring sub-graphs of interactions from which the networks are built” [Milo R et all]. These are patterns of interconnections occurring in real networks in numbers that are considered significant. . Motifs can be of any size from n=2 to N-1, where N is the total number of nodes in the network. Let us consider a directed network with N nodes and look for motifs of size n=3. There are ${}^{N}C_3$ different combinations of triplets of nodes in an N-noded network. Some triplets out of ${}^{N}C_3$ need not form a connected graph, and are not sub-graphs (an example is when out of 3 nodes 2 nodes are connected to each other and the third does not have an edge with the first two). A connected triplet is a 3-noded sub-graph. For a 3-noded sub-graph there are 13 patterns possible as shown in Fig 3.1.

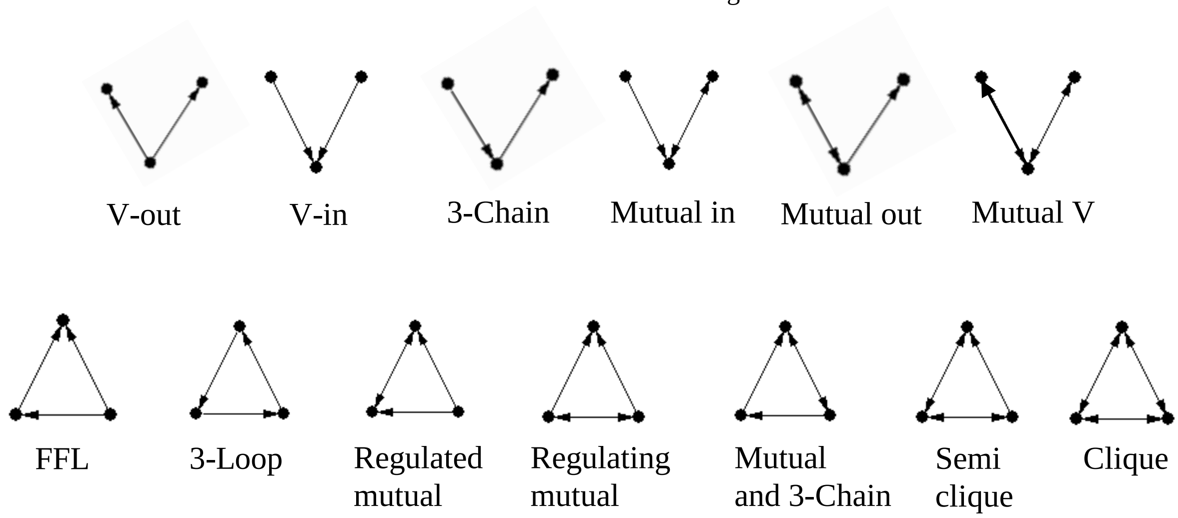

Fig 1 Motifs

Each of the ${}^{N}C_3$ triplets, if it is a sub-graph, will assume one of the 13 patterns. One can count the occurrence of each pattern for all ${}^{N}C_3$ triplets and define a vector, $\mathbf{P}_{real}$, of size 13. In a network the count for a particular pattern may be high, which by itself is not considered important. It is possible that such high count for that pattern is unavoidable for a network synthesized using the N nodes that preserve the degree distribution of the real network. To investigate this, randomized networks are created [Milo R et all] using same N nodes, ie. number of nodes and their degree distribution is preserved. Each randomized network defines a pattern count vector, $\mathbf{P}_{rand\text{-}i}$. Large number of randomized networks (i=1 to m) will define a vector of mean, $\boldsymbol{\mu}_{rand}$ and a vector of standard deviation, $\boldsymbol{\sigma}_{rand}$, of 13 patterns. For the real network we can check the significance of j^{th} pattern by, $S_j = (P_{real\text{-}j} - \mu_{rand\text{-}j})/\sigma_{rand\text{-}j}$ for j=1 to 13. For a normally distributed random number, value of S_j greater than 3 or less than 3 implies a rare occurrence (3σ limit). Any pattern with its $S_j > 2$ is considered a motif [Milo R et all], and is an over-represented pattern. Any pattern with its $S_j < -2$ is an anti-motif, and is an under-represented pattern.

2.1 Motif Significance Profile

S is a vector of size 13 that defines significance of 13 patterns in the real network. Milo R et all argue that **S** is influenced by the size of the network and propose normalization of **S** to make it largely independent of network size. Thus, significance profile vector, **Z** is defined as $Z_j = S_j / |\mathbf{S}|$. This makes comparison of networks of varying sizes possible.

2.2 Correlation of Motif Significance Profiles

[Milo R et all] have reported similarities in significant profiles of systems. They propose the standard correlation coefficients (Pearson correlation coefficient) between **Z** vectors of two systems as a measure of similarity between their significance profiles. The correlation coefficient can vary from -1 to +1. A value of +1 implies that the 13 patterns are present to the same extent in both systems, ie if a particular pattern is over-represented (under-represented) in one system it will be over-represented (under-represented) in the other system to the same extent. A value of -1 means that if a pattern is over-represented (under-represented) in one system the same will under-represented (over-represented) in the other system.

III. Motif Experiment with Characteristic Images of Dogs

In this paper we consider 159 arbitrarily chosen Canis lupus familiaris *(Dog)* characteristic images (face) of 120 breeds from literature for the study of motifs. Figure 2 briefly identifies a sample from the 150 images of 120 breeds. We create 1000 random networks for each considered image using same N nodes, ie. number of nodes and their degree distribution is preserved. Adequacy of 1000 samples for estimating μ and σ of patterns is confirmed. For each real image we compute the significance of each of the 13 patterns of 3-noded sub-graphs, $S_j = (P_{real\text{-}j} - \mu_{rand\text{-}j})/\sigma_{rand\text{-}j}$; j=1 to 13. For example, an face image of a dog has **S** = [-1.86, -0.04, -0.90, 0.72, 0.37, -1.04, 19.49, -2.71, -0.23, 0.94, 0.2, 16.86, 0.52]. **S** vectors are in fact computed for 3-noded, 4-

noded and 5-noded sub-graphs and results are available at our website [Archive of software code]. (It may be noted that the size of **S** vector for 4-noded is 199). Further study in this paper is restricted to 3-noded sub-graphs only. The significance profiles for all 120 systems are now computed as, Zj = Sj / |S|.

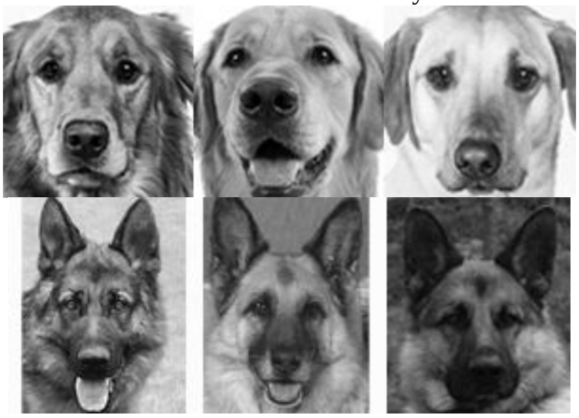

Fig. 2 Example images from the datasets

Similarities in significance profiles across all 120 breeds are now investigated by computing correlation coefficient between each image. The results are shown in Fig 3and Fig 4

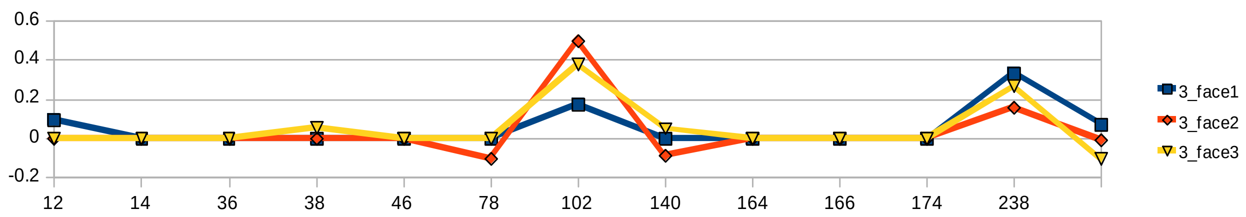

Fig. 3 Motif Profile for Retriever breed Canis lupus familiaris

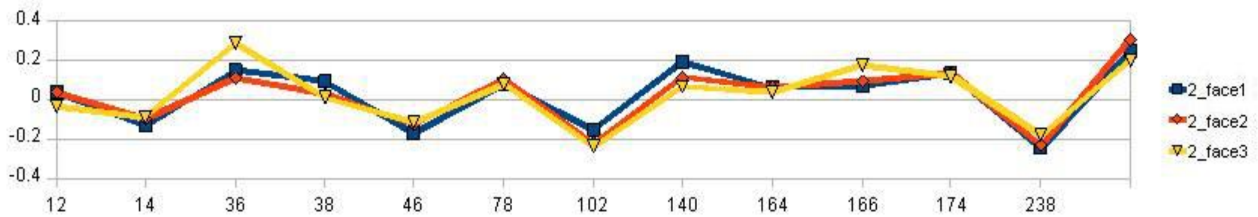

Fig. 4 Motif Profile for German shepherd breed Canis lupus familiaris

4.Discussion

All images from Fig 3are positively correlated to each other with correlation coefficients that average at 0.96. Similarly, all the images from Fig 4 are positively correlated with each other with correlation coefficients that average at 0.97. Interestingly images in Fig 3 are retriever face images where as images in Fig 4 are German shepherd images. Also, the face images of one breed has a high correlation when compared to the face image of another breed. This implies that, motifs shall be used for animal recognition.

IV.Conclusion & Directions

Ideas related to network science may give insight into previously complex and poorly understood phenomena in biological domains. Albert Barabasi argues that, "The science of networks is experiencing a boom. But despite the necessary multidisciplinary approach to tackle the theory of complexity, scientists remain largely compartmentalized in their separate disciplines" [Albert László Barabási]. The application of this network science based ideas is still in infancy and has very recently entered into study of engineering systems. This paper is probably the first paper to apply the ideas related to motifs for animal recognition. This paper has calculated motifs and significance profile for 120 breeds of Canis lupus familiaris *(Dog)* . Interesting motifs are seen in all systems. This study has thrown some insights about motif being a possible building block/classifier to understand animal recognition system.

V.Acknowledgement

We thank our brother Dr. A. S. Shaja (PhD, IIT Bombay) for his help in providing written permission to us to use his code related to motifs and

also to use some contents from his research article without any issues.

References

[1] Albert László Barabási. (2005), Taming Complexity. Nature Physics (1): 162, 68 – 70.
[2] Albert L and Albert R. (2004), Conserved Network Motifs Allow Protein-Protein Interaction Prediction, Bioinformatics (20): 18, pp 3346-3352.
[3] Amro M. Farid and Duncan C. McFarlane, (2006), An approach to the application of the design structure matrix for assessing reconfigurability of distributed manufacturing systems, Proceedings of the IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications (DIS'06) .
[4] ANSI IEEE Standard 1471. http://www.iso-architecture.org/ieee-1471/
[5] Boccaletti S, Latora V, Moreno Y, Chavez M. and Hwang D. (2006), Complex networks: Structure and dynamics, Physics Reports, 424 (4): 175-308.
[6] Dan Braha and Yaneer Bar-Yam. (2007), Statistical Mechanics of Complex Product Development, Management Science, 53 (7): 1127-1145.
[7] Duncan J Watts. (2004), Six Degrees: The Science of a Connected Age. Norton and Company, NewYork.
[8] Edward Crawley, Olivier de Weck, Steven Eppinger, Christopher Magee, Joel Moses, Warren Seering, Joel Schindall, David Wallace and Daniel Whitney. (2004), The Influence of Architecture in Engineering Systems. MIT Engineering System Division Monograph.
[9] Homas U. Pimmler and Steven D. (1994) Eppinger Integration analysis of product decompositions, ASME Design Theory and Methodology Conference Minneapolis.
[10] ISCAS High level models: http://www.eecs.umich.edu/~jhayes/iscas.restore/
[11] ISCAS'89 benchmark data : http://www.pld.ttu.ee/~maksim/benchmarks/iscas89/bench/
[12] J.M. Ottino. (2004), Engineering Complex Systems, Nature, **427**, 399 .
[13] Manuel E Sosa, Steven D Eppinger and Craig M. Rowles. (2003), Identifying Modular and Integrative Systems and Their Impact on Design Team Interactions, Journal of Mechanical Design, 125 (2): 240-252.
[14] Mathworld website, Pearson Clustering Coefficient: http://mathworld.wolfram.com/CorrelationCoefficient.html
[15] Milo R, Shen-Orr, Itzkovitz S, Kashtan N, Chklovskii D and Alon. (2002) , Network Motifs: Simple Building Blocks of Complex Networks. Science, : 298, 824-827.
[16] M Ronen, R Rosenberg, B.I. Shraiman and U. Alon. (2002), Assigning Numbers to the Arrow: Parameterizing a Gene Regulation Network by Using Accurate Expression Kinetics, Proceedings of the National Academy of Sciences (99) : 16, 10555-10560.
[17] Newman MEJ. (2003), The structure and function of complex networks, SIAM Review, 45 (2): 167-256.
[18] L Kaufman, P.J Rousseeuw, (1990), Finding Groups in Data: An Introduction to cluster Analysis, New york, John wiley & Sons Inc.
[19] Preferential Attachment Wiki: http://en.wikipedia.org/wiki/Preferential_attachment
[20] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, Uri Alon. March (2004) Superfamilies of Evolved and Designed Networks, Science (303): 5663, pp. 1538 – 1542.
[21] Shaja AS, Sudhakar K, CASMots: http://www.casde.iitb.ac.in/complexsystems/motifs/
[22] Snijders T and Borgatti S. (1999), Non-Parametric standard errors and tests for network statistics, Connections, 22 (2):161–170.
[23] S.S. Shen-Orr, R. Milo, S. Mangan and U. Alon. (2002), "Network Motifs in the Transcriptional Regulation Network of Escherichia Coli" Nature Genetics (31): 1, pp 64-68.
[24] Software graph data for specified software systems: http://www.tc.cornell.edu/~myers/Data/SoftwareGraphs/index.htm
[25] Tyson R Browning. (2001), Applying the Design Structure Matrix to System Decomposition and Integration Problems: A Review and New Directions, IEEE Transactions on Engineering Management, 48 (3) : 292-306.
[26] Thomas U. Pimmler and Steven D. Eppinger,Integration analysis of product decompositions, ASME Design Theory and Methodology Conference
[27] Vito Latora and Massimo Marchiori. (2001), Efficient Behavior of Small-World Networks. Physical Review Letters: 87, 198701.
[28] Watts DJ. (1999), Networks, dynamics, and the small-world phenomenon. American Journal of Sociology, 10 (2) : 493-527.
[29] Yael Artzy-Randrup, Sarel J. Fleishman, Nir Ben-Tal, and Lewi Stone. August (2004), Comment on Network Motifs: Simple Building Blocks of Complex Networks and Superfamilies of Evolved and Designed Networks, Science,1107c.
[30] Zhi-Qiang Jiang, Wei-Xing Zhou, Bing Xu and Wei-Kang Yuan. (2007), Process flow diagram of an Ammonia Plant as a Complex Network, AIChE Journal, 53 (2): 423-428.
[31] http://vision.stanford.edu/aditya86/ImageNetDogs/

VI.Biography

Samson D is an Assistant professor Department of Computer Science and Engineering at Cape Institute of Technology, near Kanyakumari, Tamilnadu, India. His research interest includes Animal science, Animal welfare and Image processing.