Preliminary Results on a New Fuzzy Cognitive Map Structure

John T. Rickard[1], Janet Aisbett[2], Ronald R. Yager[3], and Greg Gibbon[2]

[1] Distributed Infinity, Inc. Larkspur, CO, USA
terry.rickard@reagan.com
[2] The University of Newcastle, Callaghan, NSW, Australia
{janet.aisbett,greg.gibbon}@newcastle.edu.au
[3] Machine Intelligence Institute, Iona College, New Rochelle, NY, USA
yager@panix.com

Abstract. We introduce a new structure for fuzzy cognitive maps (FCM) where the traditional fan-in structure involving an inner product followed by a squashing function to describe the causal influences of antecedent nodes to a particular consequent node is replaced with a weighted mean type operator. In this paper, we employ the weighted power mean (WPM). Through appropriate selection of the weights and exponents in the WPM operators, we can both account for the relative importance of different antecedent nodes in the dynamics of a particular node, as well as take a perspective ranging continuously from the most pessimistic (minimum) to the most optimistic (maximum) on the normalized aggregation of antecedents for each node. We consider this FCM structure to be more intuitive than the traditional one, as the nonlinearity involved in the WPM is more scrutable with regard to the aggregation of its inputs. We provide examples of this new FCM structure to illustrate its behavior, including convergence.

1 Introduction

Fuzzy cognitive maps (FCM) [1-4] are fuzzy signed di-graphs whose nodes correspond to high-level descriptive concepts and whose links have weights corresponding to the causal relationships (positive or negative) between these concepts. Associated with each node is a fuzzy value indicating the degree to which the corresponding concept is activated as a function of the activations of the other nodes that link into it. FCMs are implemented as dynamical systems that enable the modeling of first-order feedback relationships in complex networks. Typically, certain nodes are initialized and held to fixed activation strengths, and then the network is iterated to determine the evolution of activations of the remaining nodes. The asymptotic behavior of these activations reflects the coupling of causal relationships among the nodes. FCMs have been the subject of a great deal of research interest in recent decades and have proven useful in modeling numerous applications, as surveyed in [4].

The dynamic structure of traditional FCMs is based upon a neural network model, where at each iteration the activation level of a given node is computed as a weighted

M. Jamshidi et al. (eds.), *Advance Trends in Soft Computing WCSC 2013*,
Studies in Fuzziness and Soft Computing 312,
DOI: 10.1007/978-3-319-03674-8_17,

sum of the activations of its antecedent nodes at the previous iteration, which is then normalized by a sigmoid-type "squashing function" that maps this sum into either the interval $[-1,1]$ or the interval $[0,1]$ Starting from an initial state, repeated iteration of the node activations of FCMs employing this structure are known to follow trajectories resulting in either a fixed point, a periodic limit cycle, an aperiodic attractor or a chaotic attractor [2,5].

Aside from this variable convergence (or in some instances, non-convergence) behavior associated with the traditional FCM architecture, there is a more fundamental issue that detracts from its use in models of real-world conceptual relationships. The simple, biologically-inspired neuronal model employed in this architecture is a serial combination of two functions: 1) a linear weighted arithmetic average to aggregate antecedent node activations, and 2) a nonlinear mapping of this aggregate output back into the interval $[-1,1]$ or $[0,1]$. The serial combination of these two functions results in a somewhat inflexible and inscrutable mathematical transformation, as it imposes the limitation of a linear combination of the input activations, followed by a nonlinearity that is chosen primarily for its normalization properties rather than for its logical significance. The entanglement between these two operations complicates the cognitive interpretation of the overall transformation.

While this neuronal model has proven useful in many neural network applications involving the processing of relatively low-level features such as those derived from time series or pixel values, where a cognitive interpretation of the operation is perhaps of less concern, we question its efficacy in modeling the relationships between node activations involving the higher-level conceptual features typically encountered in FCM models. A more intuitive and scrutable aggregation operator is desirable in these applications.

This has led us to investigate alternative FCM architectures. We are especially interested in the class of mean operators [6,7] for use as the aggregation operator for the antecedent activations in the nodes of the FCM, and in this paper we consider in particular the weighted power mean (WPM) operator [8-11] acting separately on the positively and negatively causal antecedents to a given node, followed by taking the difference between the positive and negative aggregates, which is then simply shifted and scaled to produce a resulting node activation in $[0,1]$.

The WPM operator provides a scrutable aggregation of antecedent activations, incorporating both importance weighting of the antecedents and the ability to take a continuously variable perspective on the input contributions to the aggregation, ranging from the most pessimistic (corresponding to the minimum activation amongst the input antecedents) to the most optimistic (corresponding to the maximum activation). This perspective is determined by the selection of the power exponent p used in the WPM. Various choices for p correspond to well-known aggregation operators, e.g., min ($p=-\infty$), harmonic mean ($p=-1$), geometric mean ($p=0$), arithmetic mean ($p=1$), root-mean-square ($p=2$), and max ($p=+\infty$).

The normalization of the output of the WPM operator is implicit in its structure, resulting in values lying in the unit interval. Thus the use of the WPM as an aggregation operator for the positively (negatively) causal antecedents to a given node in an FCM enables us more intuitively to specify how the activations of these antecedent nodes exert a corresponding positive (negative) influence on the activation of the subject node.

Another advantage of the WPM is that, using our results in [10], one can feasibly compute type-2 fuzzy WPM aggregations of type-2 fuzzy inputs. This enables the FCM architecture employing the WPM to be generalized to the perceptual computing paradigm of [12], in similar fashion to that described in [13] for social networks. Indeed, our interest in the application of the WPM aggregator to FCM structure was originally motivated by the insights gained from its application to social network analysis. In this case, rather than using scalar values for the WPM weights, exponents and activation values of the nodes in the FCM, we can use interval type-2 (IT2) fuzzy membership functions corresponding to a set of vocabulary words as in [12,13], which enables us to account for imprecise knowledge of these parameters.

In Section II of this paper, we first describe the FCM architecture constructed from mean operators in general terms. We then detail this structure in the case of the WPM operator and illustrate convergence behavior. Section III provides examples of this new FCM architecture, and Section IV concludes. We stress that our research in this area is ongoing, and thus the results in this paper are preliminary.

2 FCMs Constructed from Mean Operators

2.1 General Structure

For a traditional FCM, at time k, the state of the i^{th} node attribute is given by

$$A_i(k) = f\left(\sum_{j=1}^{n} W_{ij} A_j(k-1)\right), \tag{1}$$

where $W_{ii} \neq 0$ admits the case of self-feedback and $f(x)$ is a transfer or 'squashing' function that maps the inner product back into the interval $[-1,1]$, e.g., $f(x) = \tanh(cx) = \left(e^{2cx}-1\right)/\left(e^{2cx}+1\right)$ or into the interval $[0,1]$, e.g., $f(x) = \left(1+e^{-cx}\right)^{-1}$.

In this paper we consider updating node states using the shifted and scaled aggregations of positively and negatively causal antecedent node states obtained through a weighted mean-type aggregation operator $L(\mathbf{x})$, which for all points $\mathbf{x} = \left(x_1,\dots,x_n\right)$ in the state n-cube $[0,1]^n$ satisfies

$$\min\left(x_1,\dots,x_n\right) \le L\left(\mathbf{x}\right) \le \max\left(x_1,\dots,x_n\right). \tag{2}$$

Such mean aggregation operators include the familiar weighted power means (WPM), the exponential means and the ordered weighted averages. They also include new classes of thresholding mean type aggregation operators introduced in [14,15].

Specifically, we consider dynamic systems for which the state at node i at the k^{th} time interval is given by

$$x_i(k) = L_i\left(\mathbf{x}(k-1)\right) = \ldots = L_i^k\left(\mathbf{x}(0)\right) \tag{3}$$

for some mean aggregation operator $L_i : [0,1]^n \to [0,1]$ and for $\mathbf{x}(k)$ the vector of node states $x_i(k)$ at time k.

2.2 FCM Using the Weighted Power Mean

Consider a FCM in which the state of the i^{th} node at time k is given by the following expression using the WPM:

$$x_i(k+1) = 0.5\left[\begin{array}{l}\left(\sum_{j=1}^{n} W_{ij}^{+} x_j(k)^{p_i^+}\right)^{\frac{1}{p_i^+}} \\ -\left(\sum_{j=1}^{n} W_{ij}^{-} x_j(k)^{p_i^-}\right)^{\frac{1}{p_i^-}} + \delta_i + (1-\delta_i)\, x_i(k)\end{array}\right], \tag{4}$$

where $W_{ij}^{+} \geq 0$, $W_{ij}^{-} \geq 0$, $W_{ij}^{+} W_{ij}^{-} = 0$, and where $\sum_{i=1}^{n} W_{ij}^{+} = 1$ or 0 and $\sum_{i=1}^{n} W_{ij}^{-} = 1$ or 0, and

$$\delta_i = \begin{cases} 0, & W_{ii}^{+} = 1, \sum_{i=1}^{n} W_{ij}^{-} = 0 \\ 1, & \text{otherwise} \end{cases}.$$

The conditions insure that a given antecedent node may be positively or negatively causal (or neither) for a consequent node, but not both, while setting $\delta_i = 0$ covers cases where a node's activation is fixed at its initial value, since then

$$x_i(k+1) = 0.5\left[\left(x_i(k)^{p_i^+}\right)^{\frac{1}{p_i^+}} + x_i(k)\right] = x_i(k).$$

The two WPM operators in (4) admit separate sets of importance weights on their respective antecedents and also admit separate exponents in the WPMs, which yields separate perspectives on the aggregations of positively and negatively causal antecedents. Thus (4) provides a very general and logically intuitive inferencing structure for specifying the FCM node dynamics.

Note further from (4) that if a node's activation is not fixed and if all of its positively causal antecedent nodes have unity activations and all of its negatively causal antecedent nodes have zero activations, then the two WPM terms within the brackets take the values 1 and 0, respectively, and $x_i(k)$ takes the value 1. On the other hand, if all positively causal antecedent nodes have zero activation and all negatively causal antecedent nodes have unity activation, then $x_i(k)$ takes the value zero. Finally, if the two WPM terms produce equal values then $x_i(k)$ takes the neutral value of 0.5. Thus our FCM structure exhibits intuitively desired behaviors at both the contra extremes and the equal-valued activations of the positively and negatively causal aggregations.

We observe that the matrices W_{ij}^+ and W_{ij}^- in some instances are right stochastic matrices, i.e., when they have at least one positive entry in each row, since then all of their row sums are equal to unity [16-18]. This type of matrix is also termed a probability matrix, transition matrix or Markov matrix, and is ubiquitous in the analysis of Markov chains. However, this is not always the case, particularly as the FCM node interconnections become more sparsely populated. In the latter instances, there may be one or more rows in either W_{ij}^+ or W_{ij}^- having only zero entries.

Another feature to note from (4) is that, when all rows of *both* W_{ij}^+ and W_{ij}^- have at least one positive entry and none of the node values is held fixed (i.e., there is an off-diagonal positive entry in at least one of the matrices for each row), then the stationary value of this equation is $\forall_i \lim_{k\to\infty} x_i(k) = 0.5$. Under these assumptions, since all row sums of W_{ij}^+ and W_{ij}^- then equal unity, the first two terms in the brackets in (4) cancel each other when $\forall_j\, x_j(k) = 0.5$. This leaves only $\delta_i = 1$ within the brackets, and thus results in the identity $\forall_i\, x_i(k) = 0.5$. Since in most applications of FCMs we are interested in their dynamics when one or more node activations are fixed, this case is of little practical interest.

2.3 Convergence Properties

The convergence properties for nonlinear iterations of the form in (4) can be notoriously difficult to prove in the absence of being able to demonstrate that the nonlinearity represents a contraction mapping. There are certain cases where the WPM FCM cycles repeatedly between values, so (4) clearly is not a contraction mapping. A simple example of this is seen by choosing the following matrices for W_{ij}^+ and W_{ij}^-:

$$W^+ = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad W^- = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{5}$$

in which case each $x_i(k),\ i = 1,2$ in (4) cycles between two values. One can artificially construct other FCMs in higher dimensions that also exhibit this cyclical behavior.

However, in hundreds of thousands of simulations of WPM FCMs having a more realistic structure, we have observed overall exponential convergence with only a tiny fraction, i.e. $O(10^{-4})$ exceptions. These simulations were conducted by generating uniform random entries lying in [0,1] for the initial vector $\mathbf{x}(0)$ and the matrices W_{ij}^{+} and W_{ij}^{-}, along with uniform random entries lying in $[-10,10]$ for the individual values of p_j^{+} and p_j^{-} in (4). We then randomly zeroed out entries in W_{ij}^{+} and W_{ij}^{-} with varying probabilities, ranging up to 0.5, which produced sparser non-zero entries in these matrices, including instances where one or more entire rows of either matrix had all zero entries. We then iterated (4) for a maximum of 1500 iterations or until the squared norm of the successive differences $\|\mathbf{x}(k+1)-\mathbf{x}(k)\|_2$ was less than 10^{-12}. Figure 1 is a histogram of the number of iterations taken, drawn from 100,000 WPM FCMs using such randomly generated weight matrices, WPM exponent vectors and initial activations. The highest count to achieve convergence in this particular simulation was 667 iterations, but this obviously was an outlier. In other simulations, we have observed, in the above-noted tiny fraction of cases, periodic cycling between two values for $\mathbf{x}(k+1)$ and $\mathbf{x}(k)$.

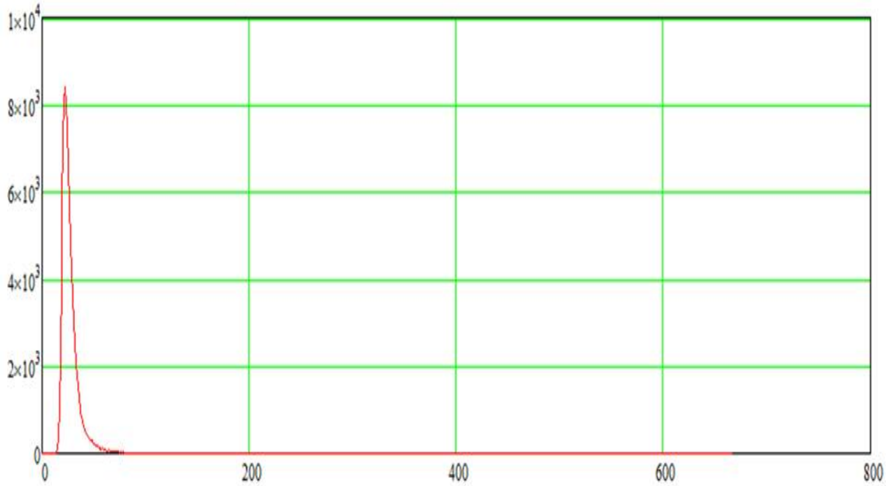

Fig. 1. Histogram of number of iterations required to reach the convergence criterion of a squared norm difference between successive activations of less than 10^{-12}. Horizontal axis is iteration count. Data from 100,000 simulations.

For an 8-node FCM, convergence from arbitrary initial node activations to "interesting" final values that are dependent only upon the system parameters and the values of any fixed node activations generally occurs in $O(10)$ iterations (excepting the previously mentioned case where no node activations are fixed and all rows of W_{ij}^{+} and W_{ij}^{-} have at least one non-zero entry, whereupon all node activations converge to 0.5).

We also tested the extreme cases where all exponents in the WPMs take values of positive or negative infinity, and observed similar results. Thus the WPM FCM system appears to be quite stable for a wide range of realistic parameter values, unlike the traditional FCM structure, which exhibits variable and generally unpredictable limiting behaviors.

We have further work to do on the WPM FCM, both in determining possibly stricter analytical conditions on convergence and in the analysis of the logical implications of the converged values. However, we believe that the initial results obtained recommend themselves to exploitation of this more scrutable structure for modeling the causal relationships between higher-level concepts.

3 Examples

We present some examples in this section that illustrate the behaviors of the WPM FCM for various system parameters and initial states. The examples are chosen to illustrate the effects of successive constraints on the initial node activations, beginning with the unconstrained case. For these examples, we could have selected a particular FCM from the numerous ones that have been studied in the literature (e.g., see 4). However, our purpose in this series of examples is to illustrate the impact of incremental changes in the WPM FCM structure in order to demonstrate the intuitive logical consistency of this structure, which is one of its primary benefits relative to the traditional structure. In future work, we shall perform comparisons between these two structural alternatives on previously studied applications.

3.1 Example 1

Let the transposes of the WPM exponent vectors $\mathbf{p}^{\pm}$ be given by

$$\begin{aligned} \mathbf{p}^{+} &= \begin{bmatrix} 9.651 & 0.271 & 7.961 & -3.471 & 1.416 & -3.037 & -4.197 & 8.456 \end{bmatrix}^{T} \\ \mathbf{p}^{-} &= \begin{bmatrix} 6.14 & 3.734 & 5.337 & 4.435 & -3.169 & -0.314 & 6.764 & 0.812 \end{bmatrix}^{T} \end{aligned} \tag{6}$$

and consider the matrices W_{ij}^{+} and W_{ij}^{-} given by

$$W^{+} = \begin{bmatrix} 0.393 & 0 & 0 & 0 & 0 & 0.455 & 0.152 & 0 \\ 0.219 & 0 & 0.7 & 0 & 0 & 0 & 0.08 & 0 \\ 0.37 & 0 & 0.184 & 0 & 0 & 0 & 0 & 0.446 \\ 0 & 0.055 & 0 & 0 & 0 & 0 & 0 & 0.945 \\ 0 & 0.983 & 0 & 0.017 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{7}$$

$$W^- = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0.891 & 0 & 0.109 & 0 & 0 \\ 0 & 0 & 0 & 0.244 & 0.756 & 0 & 0 & 0 \\ 0 & 0 & 0.342 & 0.131 & 0.172 & 0 & 0.355 & 0 \\ 0 & 0 & 0 & 0 & 0.108 & 0.265 & 0.387 & 0.24 \\ 0.402 & 0.221 & 0 & 0 & 0 & 0.377 & 0 & 0 \\ 0.407 & 0.055 & 0.159 & 0.057 & 0 & 0.323 & 0 & 0 \\ 0.537 & 0.454 & 0 & 0 & 0 & 0 & 0 & 0.009 \end{bmatrix} \quad (8)$$

Note that row 7 of W^+ in (7) has all zeroes, indicating no positively causal inputs to node 7, whereas all other nodes have both positively and negatively causal inputs.

From an initial activation state $\mathbf{x}(0)$, the WPM FCM converges in 56 iterations to the final state shown below:

$$\mathbf{x}(0) = \begin{bmatrix} 0.99 & 0.037 & 0.761 & 0.054 & 0.813 & 0.344 & 0.648 & 0.294 \end{bmatrix}^T \quad (9)$$

$$\mathbf{x}(56) = \begin{bmatrix} 0.619 & 0.529 & 0.476 & 0.426 & 0.626 & 0.368 & 0.223 & 0.324 \end{bmatrix}^T \quad (10)$$

Note from (6) that p_7^- for the negatively causal WPM is relatively large and positive, so this WPM tends toward the maximum of its inputs. Since there is zero contribution from the positively causal WPM for this node, this causes the converged value of node 7 to be relatively small (0.223).

This value also happens to be the sole positively causal input for node 8, whereas its primary negatively causal inputs from nodes 1 and 2 have activations above the neutral value of 0.5, and their corresponding WPM has a positive exponent (0.812). This causes the converged activation for node 8 also to be relatively small. Examining the other converged activations, we conclude that they appear to be consistent with the system parameters.

3.2 Example 2

Suppose that we now replace the first rows of W_{ij}^+ and W_{ij}^- in Example 1 with all zeroes, i.e., so that the activation of node 1 is held fixed, with the remaining rows of these matrices, the WPM exponents and the initial activations unchanged. The FCM now converges in 19 iterations to:

$$\mathbf{x}(19) = \begin{bmatrix} 0.99 & 0.647 & 0.568 & 0.275 & 0.778 & 0.271 & 0.066 & 0.122 \end{bmatrix} \quad (11)$$

The activation of node 1 remains constant as expected, and its high value (virtually the maximum of 1) coupled with its significant weight in $W_{7,1}^-$ and the large positive WPM exponent p_7^- for the negatively causal inputs to node 7 (with no positively causal input) results in a very low converged activation for node 7, exactly as would be anticipated.

3.3 Example 3

Suppose now that both the first and fourth nodes' activations are held fixed in Example 1, so that the corresponding rows in W_{ij}^{+} and W_{ij}^{-} are all zero, with the WPM exponents and initial activations unchanged. Then the FCM activations converge in 25 iterations to:

$$\mathbf{x}(25) = \begin{bmatrix} 0.99 & 0.705 & 0.555 & 0.054 & 0.806 & 0.267 & 0.065 & 0.108 \end{bmatrix} \qquad (12)$$

Comparing this vector with the converged activations (11) of the previous example, we see that fixing the activation of node 4 at its low initial value of 0.054 has caused only minor rebalancing of the free nodes' activations. On examining the fourth column of W^{+} in (7), we see that node 4's activation contributes nothing in the way of positively causal influence to any node except node 5, and to this one only to a very small degree since $W_{5,4}^{+} = 0.017$. Thus virtually all of the effect of fixing node 4's activation is accounted for in the negatively causal inputs, which is most prominent for node 2. Again, the results are consistent with what would be expected of the logic.

4 Conclusion

We introduced a new FCM structure in this paper that has a more scrutable interpretation of the aggregations that go into the activations of its nodes, by employing the WPM as the aggregation operator in place of the traditional approach using a linear weighted average followed by a non-linear squashing function. While purely periodic cycling between values of the activations can occur, individual components converged in simulations of realistic structures.

We illustrated this new FCM structure using examples both where the node activations are unconstrained and where one or two of the initial activations are held fixed. We demonstrated that the converged activations obtained from the iterations are consistent with the implied logic of the WPM aggregations of the positively and negatively causal inputs to the nodes, which lends empirical evidence to the utility of this new structure.

In addition to its scrutability, this new structure can be extended to IT2 representations of both the linkage weights and the node activations using the results in [10] and [13], which enables us to compute the successive IT2 membership functions of the node activations as the iterations proceed. Thus we can employ the "perceptual computing" paradigm of [12] to account for imprecise word-based descriptions of the causal relationship strengths, WPM exponents and initial activation levels in the FCM. This represents a major extension to the modeling capability of FCMs.

There obviously remains much work to do on both the analytical and practical aspects of this new FCM structure. Given the preliminary and ongoing nature of our research, we have not yet applied it to specific modeling problems, nor have we yet compared our results to traditional FCM structures. However, we intend to do so in future work.

References

1. Kosko, B.: Fuzzy cognitive maps. Int. J. Man-Mach. Stud. 24(1), 65–75 (1986)
2. Kosko, B.: Fuzzy Engineering. Prentice-Hall, Englewood Cliffs (1997)
3. Glykas, M. (ed.): Fuzzy Cognitive Maps. STUDFUZZ, vol. 247. Springer, Heidelberg (2010)
4. Papageorgiou, E.I., Salmeron, J.L.: A review of fuzzy cognitive maps research during the last decade. IEEE Trans. Fuzzy Syst. 21(1), 66–79 (2013)
5. Boutalis, Y., Kottas, T.L., Christodoulou, M.: Adaptive estimation of fuzzy cognitive maps with proven stability and parameter convergence. IEEE Trans. Fuzzy Syst. 17(4), 874–889 (2009)
6. Yager, R.R.: A general approach to criteria aggregation using fuzzy measures. International J. Man-Machine Studies 38, 187–213 (1993)
7. Yager, R.R.: On mean type aggregation. IEEE Trans. Systems, Man, Cybernetics—Part B: Cybernetics 26, 209–221 (1996)
8. Dujmović, J., Larsen, H.L.: Generalized conjunction/disjunction. J. Approximate Reasoning 46, 423–446 (2007)
9. Dujmović, J.: Continuous preference logic for system evaluation. IEEE Trans. Fuzzy Syst. 15(6), 1082–1099 (2007)
10. Rickard, J.T., Aisbett, J., Yager, R.R., Gibbon, G.: Fuzzy weighted power means in evaluation decisions. In: Proc. 1st World Symposium on Soft Computing, Paper #100, San Francisco, CA (2010)
11. Rickard, J.T., Aisbett, J., Yager, R.R., Gibbon, G.: Linguistic weighted power means: comparison with the linguistic weighted average. In: Proc. FUZZ-IEEE 2011, 2011 World Congress on Computational Intelligence, Taipei, Taiwan, pp. 2185–2192 (2011)
12. Mendel, J.M., Wu, D.: Perceptual Computing. John Wiley & Sons, Hoboken (2010)
13. Rickard, J.T., Yager, R.R.: Perceptual computing in social networks. In: Proc. 2013 International Fuzzy Systems Association World Congress/North American Fuzzy Information Processing Society Annual Mtg. Paper #9, Edmonton, Alberta, Canada (2013)
14. Rickard, J.T., Aisbett, J.: New classes of threshold aggregation functions based upon the Tsallis q-exponential. In: Rickard, J.T., Aisbett, J. (eds.) 2013 International Fuzzy Systems Association World Congress/North American Fuzzy Information Processing Society Annual Mtg, paper #8, Edmonton, AB, Canada (2013)
15. Rickard, J.T., Aisbett, J.: New classes of threshold aggregation functions based upon the Tsallis q-exponential with applications to perceptual computing. IEEE Trans. on Fuzzy Syst. (accepted for publication, 2013)
16. Meyer, C.: Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia (2001)
17. Gantmacher, F.R.: Theory of Matrices, Vol. 2. AMS Chelsea Publishing, Providence, RI (1989), http://bookos.org/g/F.%20R.%20Gantmacher
18. Encyclopedia of Mathematics entry under “Stochastic matrix”, http://www.encyclopediaofmath.org/index.php/Stochastic_matrix