

OPTIMIZED CLASSIFICATION OF METAPHASE CHROMOSOMES USING FUZZY RULE & NON-LINEAR CLASSIFIERS

¹R.SATHYAPRIYA, ²K.SAKTHIVEL

^{1,2}Assistant professor, Dept. of EEE,
BIST, BIHER, Bharath university, Chennai-73
r.Sathyapriya2013@bharathuniv.ac.in

Abstract: A new feature extraction algorithm for chromosome classification system based on fuzzy rule and non-linear classifier are proposed. Our hypothesis is that by selecting most effective feature sets and adaptively optimizing classifiers for the different groups of chromosomes with similar image characteristics, we will be able to achieve optimized classification result. The algorithm is based on two-layer classification platform. In the first layer of the scheme, fuzzy logics are employed to classify 24 chromosomes into seven classes. In the second layer, non-classifiers were employed for seven classes to identify individual chromosomes. It demonstrates that this new scheme achieves higher and robust performance in classifying chromosomes. Thus we can reduce the complexity of automated karyotyping scheme and improve its performance and robustness.

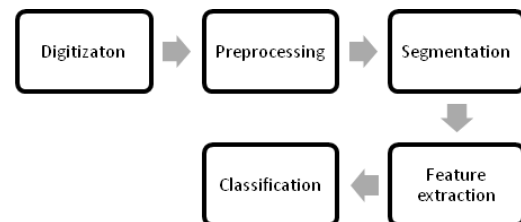
INTRODUCTION:

Chromosome karyotype analysis is an important screening procedure routinely utilized in clinical cytogenetic labs with the aim of early diagnosing cancers and genetic aberrations. Chromosomes are first stained with a fluorescent dye [1-6], and then imaged through a microscope for subsequent analysis and classification. Each chromosome in the image has to be identified and assigned to one of 24 classes. The result of this procedure is the so-called karyotype image [7-11], in which all chromosomes in a cell are graphically arranged according to an international system for cytogenetic nomenclature (ISCN) classification.

METHODOLOGY:

- Digitization
- Pre processing
- Segmentation
- Feature extraction
- Non-linear Classification
 - ✓ SVM based kernel classifier
 - ✓ Similarity classifier
 - ✓ Neuro-fuzzy based classifier

Block diagram:



DIGITIZATION:

This is generally achieved with flying spot microscope which converts the chromosome pattern into a density array in the computer. The system will start pixel by pixel scanning to find the densities above the background density. Thus the position of each chromosome will be defined to identify them in pairs [12-19].

PRE PROCESSING:

The acquired image is preprocessed in order to reduce noise. First the image is converted into gray scale format followed by Area normalization [20-26], Intensity normalization,

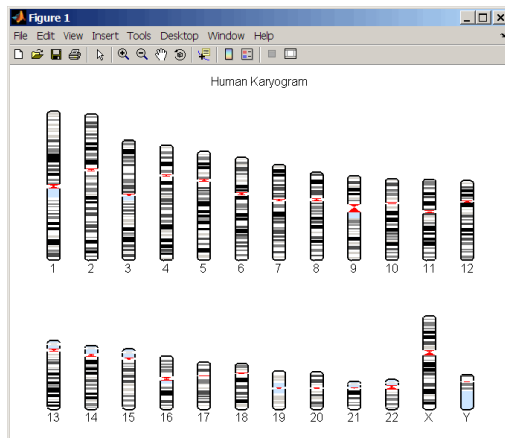
Contrast enhancement, Erosion & Edge detection [27-35].

Ideogram:

Ideogram chart is used for the identification of chromosomes. The chromosomes get their banding pattern due to the varying amount of absorption of the stain by the proteins present in them. The ideogram shown is for 400 banding - 800 banding will have more number of bands [36-41].

Geneticist identifies a chromosome based on Length, Banding pattern and position of the constriction (centromere) present in it.

- ✓ Chromosomes 1- 22 are called autosomes.
- ✓ X and Y chromosomes are called as Sex chromosomes as they determine the sex of a person.



SEGMENTATION:

The digitized image is segmented using OTSU algorithm. OTSU's method is used to automatically perform histogram shape-based image thresholding or the reduction of gray level image into binary image. So, the algorithm assumes that the image to be thresholded contains two classes of pixels

(e.g. foreground and background) then calculates the optimum threshold separating those two classes. The extension of this to the multilevel thresholding is referred to as multi OTSU method [42-45].

Algorithm:

1. Compute histogram and probabilities of each intensity level
2. Set up initial $\omega_i(0)$ and $\mu_i(0)$
3. Step through all possible thresholds $t = 1 \dots \text{maximum intensity}$
 1. Update ω_i and μ_i
 2. Compute $\sigma_b^2(t)$
4. Desired threshold corresponds to the maximum $\sigma_b^2(t)$
5. You can compute two maxima (and two corresponding thresholds). $\sigma_{b1}^2(t)$ is the greater max and $\sigma_{b2}^2(t)$ is the greater or equal maximum
6. Desired threshold =
$$\frac{\text{threshold}_{d1} + \text{threshold}_{d2}}{2}$$

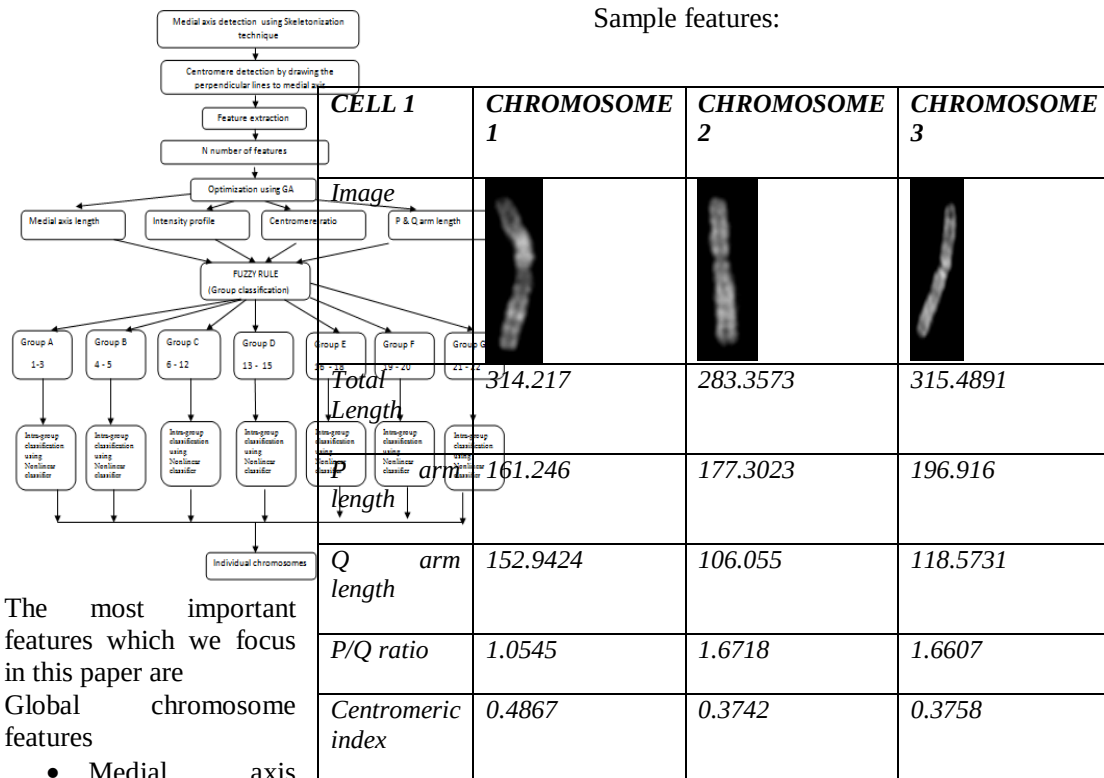
FEATURE EXTRACTION:

We have extracted the feature by using Medial Axis Transform (MAT Algorithm).

Identification of centromere:

- The shortest segment perpendicular to the Medial axis gives the location of the centromere.
- The intersection of the shortest segment and Medial axis is taken as the point separating the P and Q arms of the chromosome.

FLOWCHART:



The most important features which we focus in this paper are Global chromosome features

- Medial axis length or Total length (Normalized)
- Contour length (Normalized)
- Area (Normalized)

Chromosome's centromeric features:

- Medial axis length ratio of P/Q (Normalized)
- Contour length P/Q ratio (Normalized)
- Area P/Q ratio (Normalized)
- Centromeric Index ($Q/(P+Q)$)

Number of real bands: Number of maxima & minima in profile.

Real gray level banding: These are the average gray level of the real band.

Distances between centers of bands: These are the distances between the bands centers.

Real band thickness: These are the real banding thickness.

CLASSIFICATION:

We are going to adopt 2 phase method of classification.

Phase 1: Inter Group Classification(5474 samples-approx 2600 as training data & 2600 as testing data)

Phase 2: Intra Group Classification(approx 714 samples for each group)

Inter Group Classification:

In the first phase the diploid set of chromosomes are karyotyped into seven groups based on Denver's system of classification.

Denver's system:

Chromosome class	size	Relative position of centromere
Group A (1–3)	Large	Metacentric
Group B (4–5)	Large	Sub metacentric
Group C (6–12,X)	Medium	Sub metacentric
Group D (13–15)	medium	Acrocentric
Group E (16–18)	Relatively short	Sub metacentric
Group F (19–20)	short	Metacentric or Sub metacentric
Group G (21–22,Y)	Short	Acrocentric

Phase 1 classification can be done by estimating the range of Centromeric Index. Thus Fuzzy rule based classifier is used in this phase to perform the classification with robustness.

FUZZY BASED CLASSIFIER:

Fuzzy classification is the process of grouping individuals having the same characteristics into a **fuzzyset**. A fuzzy classification corresponds to a membership function that indicates whether an individual is a member of a class, given its fuzzy classification predicate.

Mathematical background:

Normal fuzzy set: A fuzzy subset A of a classical set X is called normal if there exists an $x \in X$ such that $A(x) = 1$. Otherwise A is subnormal.

α -cut: An α -level set of a fuzzy set A of X is a non-fuzzy set denoted by $[A]_\alpha$ and is defined by

$$[A]_\alpha = \{t \in X | A(t) \geq \alpha\} \text{ if } \alpha > 0$$

$$cl(A) \text{ if } \alpha = 0$$

where $cl(A)$ denotes the closure of the support of A .

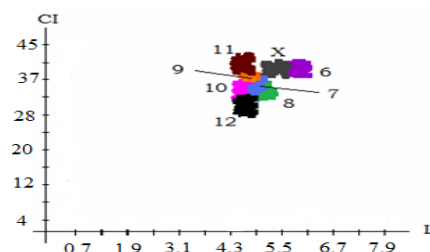
Intra Group Classification:

This is the core part of our process where we need to concentrate on non linear classifiers. So that overlapped data can be easily estimated in aligned in a particular domain.

Why non-linear classifier?

The main idea for going for non-linear is that each class vector contains overlapping features in it. Thus it makes impossible to karyotype the intragroup cells using linear function. So that the Kernel function is used to convert the linear data into the non linear format and feature vectors are aligned into the particular domain class by using nearest neighbouring technique.

Limiting factor:



spread plot of group C chromosome

SVM BASED KERNEL CLASSIFIER:

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships.

Illustration:

A *kernel function* is defined as a function that corresponds to a dot product of two feature vectors in some expanded feature space

$$K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$$

The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a **non-probabilistic binary linear classifier**.

Mathematical Background:

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is **as wide as possible**. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVMs can efficiently perform non-linear classification using what is called the **KERNEL TRICK**, implicitly mapping their inputs into high-dimensional feature spaces.

$$g(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b = \sum_{i \in SV} \alpha_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + b$$

SIMILARITY CLASSIFIER:

The term **similarity-based classification** implies whether the pair wise relationship is a similarity or dissimilarity.

Illustration:

Similarity-based classifiers estimate the class label of a test sample based on the similarities between the test sample and a set of labeled training samples, and the pair wise similarities between the training samples. Similarity-based classification does not require direct access to the features of the samples, and thus the sample space can be any set, not necessarily a Euclidean space, as long as the similarity function is well defined for any pair of samples.

Ordered Weighted Averaging Operator:

Formally an **OWA** operator of dimension n is a mapping $F: R_n \rightarrow R$ that has an associated collection of weights $W = [w_1, \dots, w_n]$ lying in the unit interval and summing to one and

$$F(a_1, \dots, a_n) = \sum_{j=1}^n w_j b_j$$

with

where b_j is the j^{th} largest of the a_i .

By choosing different W one can implement different aggregation operators. The OWA operator is a non-linear operator as a result of the process of determining the b_j .

OWA operator is characterized by the measure of orness & entropy.

Orness is used to account for the similarity of the operator with the maximum when a particular parameterization is used.

Orness(W) =

$$\left(\frac{1}{n-1}\right) \sum_{i=1}^n ((n-i) \cdot w_i)$$

The measure of dispersion/Entropy shows how much information of the arguments is being used during an aggregation. It is also called as dispersion.

Dispersion (W) =

$$-\sum_{i=1}^n w_i \cdot \ln(w_i).$$

NEURO-FUZZY CLASSIFIER:

"Neuro-fuzzy" means the employment of heuristic learning strategies derived from the

domain of neural network theory to support the development of a fuzzy system.

Illustration:

The problem of finding appropriate membership functions and fuzzy rules is often a tiring process of trial and error. Therefore the idea of applying learning algorithms to fuzzy systems was considered so called adaptive or self-organizing.

Constituents:

Neural networks- Learning and adaptation of patterns in large amounts of data. Fuzzy set theory- Imprecise reasoning using fuzzy IF-THEN-ELSE rules. Genetic or evolutionary algorithms and simulated annealing- Systematic search using non-derivative methods. Conventional AI- Ability to do symbolic manipulation.

Algorithm:

- Initialize the fuzzy rules using k-means algorithm
- Each class is given with some number of clusters
- Gaussian membership function is best suited for fuzzy systems
- Weight $\implies$ Samples
- The scaled conjugate gradient (SCG) algorithm is used to determine the optimum values of nonlinear parameters.
- The SCG is faster than the steepest descent and some second order derivative based methods.
- And it is suitable for large scale problems

Linguistic hedges:

Linguistic hedges are applied to the fuzzy sets of rules, and are adapted by SCG algorithm. By this way, some distinctive features are emphasized by power values, and some irrelevant features are damped with power values. The power effects in any feature are generally different for different classes. By using the linguistic hedges increase the recognition rates.

CONCLUSION:

A new feature extraction and classification algorithms for Meta phase chromosomes images is proposed and our hypothesis is to improve the performance and robustness of an ANN-based automated karyotyping scheme by reducing the complexity of ANN structure and selecting optimal feature sets, which has unique characteristics. First, we are going to develop a fuzzy rule based classifier where the rules to the inference are based on centromeric index and then the each individual chromosome is identified using perceptron neural network.

REFERENCE:

1. Nimal, R.J.G.R., Hussain, J.H., Effect of deep cryogenic treatment on EN24 steel, International Journal of Pure and Applied Mathematics, V-116, I-17 Special Issue, PP-113-116, 2017
2. Parameswari, D., Khanaa, V., Deploying lamport clocks and linked lists, International Journal of Pharmacy and Technology, V-8, I-3, PP-17039-17044, 2016
3. Parameswari, D., Khanaa, V., Case for massive multiplayer online role-playing games, International Journal of Pharmacy and Technology, V-8, I-3, PP-17404-17409, 2016

4. Parameswari, D., Khanaa, V., Deconstructing model checking with hueddot, International Journal of Pharmacy and Technology, V-8, I-3, PP-17370-17375, 2016
5. Parameswari, D., Khanaa, V., The effect of self-learning epistemologies on theory, International Journal of Pharmacy and Technology, V-8, I-3, PP-17314-17320, 2016
6. Pavithra, J., Peter, M., Gowtham Aashirwad, K., A study on business process in IT and systems through extranet, International Journal of Pure and Applied Mathematics, V-116, I-19 Special Issue, PP-571-576, 2017
7. Pavithra, J., Ramamoorthy, R., Satyapira Das, S., A report on evaluating the effectiveness of working capital management in googolsoft technologies, Chennai, International Journal of Pure and Applied Mathematics, V-116, I-14 Special Issue, PP-129-132, 2017
8. Pavithra, J., Thooyamani, K.P., A cram on consumer behaviour on Mahindra two wheelers in Chennai, International Journal of Pure and Applied Mathematics, V-116, I-18 Special Issue, PP-55-57, 2017
9. Pavithra, J., Thooyamani, K.P., Dkhar, K., A study on the air freight customer satisfaction, International Journal of Pure and Applied Mathematics, V-116, I-14 Special Issue, PP-179-184, 2017
10. Pavithra, J., Thooyamani, K.P., Dkhar, K., A study on the working capital management of TVS credit services limited, International Journal of Pure and Applied Mathematics, V-116, I-14 Special Issue, PP-185-187, 2017
11. Pavithra, J., Thooyamani, K.P., Dkhar, K., A study on the analysis of financial performance with reference to Jeppiaar Cements Pvt Ltd, International Journal of Pure and Applied Mathematics, V-116, I-14 Special Issue, PP-189-194, 2017
12. Peter, M., Dayakar, P., Gupta, C., A study on employee motivation at Banalari World Cars Pvt Ltd Shillong, International Journal of Pure and Applied Mathematics, V-116, I-18 Special Issue, PP-291-294, 2017
13. Peter, M., Kausalya, R., A study on capital budgeting with reference to signware technologies, International Journal of Pure and Applied Mathematics, V-116, I-18 Special Issue, PP-71-74, 2017
14. Peter, M., Kausalya, R., Akash, R., A study on career development with reference to prem heera surgicals, International Journal of Pure and Applied Mathematics, V-116, I-14 Special Issue, PP-415-420, 2017
15. Peter, M., Kausalya, R., Mohanta, S., A study on awareness about the cost reduction and elimination of waste among employees in life line multispeciality hospital, International Journal of Pure and Applied Mathematics, V-116, I-14 Special Issue, PP-287-293, 2017
16. Peter, M., Srinivasan, V., Vigneshsa, A., A study on working capital management at deccan Finance Pvt Limited Chennai, International Journal of Pure and Applied

- Mathematics, V-116, I-14 Special Issue, PP-255-260, 2017
17. Peter, M., Thooyamani, K.P., Srinivasan, V., A study on performance of the commodity market based on technical analysis, International Journal of Pure and Applied Mathematics, V-116, I-18 Special Issue, PP-99-103, 2017
 18. Philomina, S., Karthik, B., Wi-Fi energy meter implementation using embedded linux in ARM 9, Middle - East Journal of Scientific Research, V-20, I-12, PP-2434-2438, 2014
 19. Philomina, S., Subbulakshmi, K., Efficient wireless message transfer system, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-289-293, 2017
 20. Philomina, S., Subbulakshmi, K., Ignition system for vehicles on the basis of GSM, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-283-286, 2017
 21. Philomina, S., Subbulakshmi, K., Avoidance of fire accident by wireless sensor network, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-295-299, 2017
 22. Pothumani, S., Anuradha, C., Monitoring android mobiles in an industry, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-537-540, 2017
 23. Pothumani, S., Anuradha, C., Decoy method on various environments - A survey, International Journal of Pure and Applied Mathematics, V-116, I-10 Special Issue, PP-197-199, 2017
 24. Pothumani, S., Anuradha, C., Priya, N., Study on apple iCloud, International Journal of Pure and Applied Mathematics, V-116, I-8 Special Issue, PP-389-391, 2017
 25. Pothumani, S., Hameed Hussain, J., A novel economic framework for cloud and grid computing, International Journal of Pure and Applied Mathematics, V-116, I-13 Special Issue, PP-5-8, 2017
 26. Pothumani, S., Hameed Hussain, J., A novel method to manage network requirements, International Journal of Pure and Applied Mathematics, V-116, I-13 Special Issue, PP-9-15, 2017
 27. Pradeep, R., Vikram, C.J., Naveenchandra, P., Experimental evaluation and finite element analysis of composite leaf spring for automotive vehicle, Middle - East Journal of Scientific Research, V-12, I-12, PP-1750-1753, 2012.
 28. Pradeep, R., Vikram, C.J., Naveenchandra, P., Experimental evaluation and finite element analysis of composite leaf spring for automotive vehicle, Middle - East Journal of Scientific Research, V-17, I-12, PP-1760-1763, 2013.
 29. Prakash, S., Jayalakshmi, V., Power quality improvement using matrix converter, International Journal of Pure and Applied Mathematics, V-116, I-19 Special Issue, PP-95-98, 2017
 30. Prakash, S., Jayalakshmi, V., Power quality analysis & power system study in high voltage systems, International Journal of Pure and Applied Mathematics, V-

- 116, I-19 Special Issue, PP-47-52, 2017
31. Prakash, S., Sherine, S., Control of BLDC motor powered electric vehicle using indirect vector control and sliding mode observer, International Journal of Pure and Applied Mathematics, V-116, I-19 Special Issue, PP-295-299, 2017
 32. Prakesh, S., Sherine, S., Forecasting methodologies of solar resource and PV power for smart grid energy management, International Journal of Pure and Applied Mathematics, V-116, I-18 Special Issue, PP-313-317, 2017
 33. Prasanna, D., Arulselvi, S., Decoupling smalltalk from rpcs in access points, International Journal of Pure and Applied Mathematics, V-116, I-16 Special Issue, PP-1-4, 2017
 34. Prasanna, D., Arulselvi, S., Exploring gigabit switches and journaling file systems, International Journal of Pure and Applied Mathematics, V-116, I-16 Special Issue, PP-13-17, 2017
 35. Prasanna, D., Arulselvi, S., Collaborative configurations for wireless sensor networks systems, International Journal of Pure and Applied Mathematics, V-116, I-15 Special Issue, PP-577-581, 2017
 36. Priya, N., Anuradha, C., Kavitha, R., Li-Fi science transmission of knowledge by way of light, International Journal of Pure and Applied Mathematics, V-116, I-9 Special Issue, PP-285-290, 2017
 37. Priya, N., Pothumani, S., Kavitha, R., Merging of e-commerce and e-market-a novel approach, International Journal of Pure and Applied Mathematics, V-116, I-9 Special Issue, PP-313-316, 2017
 38. Raj, R.M., Karthik, B., Effective demining based on statistical modeling for detecting thermal infrared, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-273-276, 2017
 39. Raj, R.M., Karthik, B., Energy sag mitigation for chopper, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-267-270, 2017
 40. Raj, R.M., Karthik, B., Efficient survey in CDMA system on the basis of error revealing, International Journal of Pure and Applied Mathematics, V-116, I-20 Special Issue, PP-279-281, 2017
 41. Rajasulochana, P., Krishnamoorthy, P., Ramesh Babu, P., Datta, R., Innovative business modeling towards sustainable E-Health applications, International Journal of Pharmacy and Technology, V-4, I-4, PP-4898-4904, 2012
 42. Rama, A., Nalini, C., Shanthi, E., An iris based authentication system by eye localization, International Journal of Pharmacy and Technology, V-8, I-4, PP-23973-23980, 2016
 43. Rama, A., Nalini, C., Shanthi, E., Effective collaborative target tracking in wireless sensor networks, International Journal of Pharmacy and Technology, V-8, I-4, PP-23981-23986, 2016

44. Ramamoorthy, R., Kanagasabai, V.,
Irshad Khan, S., Budget and
budgetary control, International
Journal of Pure and Applied
Mathematics, V-116, I-20 Special
Issue, PP-189-191, 2017
45. Ramamoorthy, R., Kanagasabai, V.,
Jivandan, S., A study on training and
development process at Vantec
Logistics India Pvt Ltd, International
Journal of Pure and Applied
Mathematics, V-116, I-14 Special
Issue, PP-201-207, 2017

