

Clasificador Difuso Neuronal Aplicado a Casos de Datos Sintéticos

Fuzzy Neural Classifier Applied to Cases of Synthetic Data

Jose Gerardo Chacón Rangel¹. Anderson Smith Florez Fuentes³,
Johel Enrique Rodriguez Fernandez³

RESUMEN

En este artículo se presenta el desarrollo de un sistema computacional difuso neuronal que permite clasificar casos de datos sintéticos a través patrones con solapamiento controlado. Se construyó una serie de modelos neuronales con lógica difusa y redes neurales que fueron analizados utilizando diferentes porcentajes de solapamiento. En función de los resultados obtenidos, se seleccionó el mejor modelo para clasificar los patrones de acuerdo con criterios apropiados de desempeño como error permitido y tiempo de entrenamiento. Se obtuvo un modelo capaz de identificar un tipo de clase, que tiende a minimizar los errores de clasificación. El modelo difuso neuronal de este tipo puede ayudar a especialistas de diferentes disciplinas a diagnosticar con un mínimo de error, cuando los datos presentan rasgos con patrones solapados.

Palabras Clave: Datos sintéticos, difuso neuronal, redes neurales, solapamiento controlado.

ABSTRACT

This article presents the development of a computational system that allows diffuse neuronal classify cases of synthetic data through patterns overlap with controlled. They built a series of models with neural fuzzy logic and neural networks that were analyzed using different percentages of overlap. Depending on the results obtained, was selected the best model to classify the patterns in accordance with appropriate criteria for performance as permissible and training time. We obtained a model able to identify a type of class, which tends to minimize the errors of classification. The diffuse neuronal model of this type can help specialists from different disciplines to diagnose with a minimum of error, when data are traits with overlapping patterns.

Keywords: Diffuse neuronal, controlled overlapping, synthetic data, neural networks

1. Universidad de Pamplona, Facultad de Ingenierías y Arquitecturas, Ingeniería de Sistemas (Villa del Rosario), Email: gerardo61@gmail.com

2. Universidad de Pamplona, Facultad de Ingenierías y Arquitecturas, Ingeniería de Sistemas (Villa del Rosario), Email: andersonflorezf@unipamplona.edu.co

3. Universidad de Pamplona, Facultad de Ingenierías y Arquitecturas, Ingeniería de Sistemas (Villa del Rosario), Email: joheler@gmail.com

Clasificador Difuso Neuronal Aplicado a Casos de Datos Sintéticos

INTRODUCCIÓN

Hay un interés creciente en investigar sobre el uso de la lógica difusa en el desarrollo de sistemas neuronales, tal como lo revela el considerable número de artículos sobre el tema en los números de revistas especializadas como por ejemplo los artículos de Hayashi et al (1990), Ramírez (1996) Ma et al (1999), Mitra y Hayashi (2000) y Duch et al (2004).

Los sistemas que combinan técnicas de lógica difusa y redes neuronales (sistemas neurodifusos) pueden ser utilizados en una gran variedad de áreas y aplicaciones. Por ejemplo, pueden ser usados en el diagnóstico de enfermedades, para clasificar patrones que representan las condiciones de sintomatología de un conjunto de pacientes (un patrón con múltiples atributos por paciente) a los cuales se les ha diagnosticado un tipo de enfermedad. Un caso es especialmente interesante es cuando los datos de entrada tienen áreas de solapamiento entre ellos, que dificultan la formulación de una conclusión definitiva.

El objetivo de este trabajo es la construcción de un clasificador basado en lógica difusa y redes neuronales para la clasificación de datos sintéticos con solapamiento controlado que aporte resultados que permita ser viable la aplicación este método en otras áreas.

MÉTODO

Posteriormente a la revisión bibliográfica, se realizaron: el diseño del modelo neurodifuso representativo de un sistema híbrido entre redes neuronales y lógica difusa para pruebas de patrones solapados; la generación de conjuntos de datos sintéticos correspondientes a diversos casos de prueba; la selección de los datos reales para experimentar; la simulación del sistema para patrones solapados utilizando una herramienta de software; la interpretación de los resultados de las simulaciones realizadas del modelo para determinar la influencia de un etiquetado difuso en el estudio de patrones solapados.

Para la realización de este trabajo se construyeron una serie de modelos con lógica difusa y redes neuronales a partir de uno propuesto por Pal y Mitra (1992, 1999), el Perceptrón Multicapa Difuso (Fuzzy Multi-layer Perceptron o FMLP), que incorpora rasgos difusos en la capa de salida de una red neuronal tipo perceptrón multicapa. En este modelo, el valor de pertenencia de un patrón a una clase se define como (μ) y se utiliza como etiqueta de entrenamiento durante la fase de aprendizaje. Durante el proceso de entrenamiento de la red, los pesos w_i son actualizados por retropropagación del error con respecto a estos valores de (μ) .

El modelo utilizado incorpora la borrosidad (“fuzzy-ness”) en las capas de entrada y de salida de una red neuronal. Cualquier valor de los atributos de entrada son descritos en términos de alguna combinación de valores de pertenencia en el conjunto de propiedades lingüísticas bajo (L), medio (M) y alto (H). El valor de pertenencia de un patrón a una clase se define como (), este se representa en la capa de salida. Durante el proceso de entrenamiento de la red, los pesos w_i son actualizados por retropropagación del error con respecto a valores de membresía (). Este modelo se muestra en la figura 1 y se abrevia como modelo FMLP.

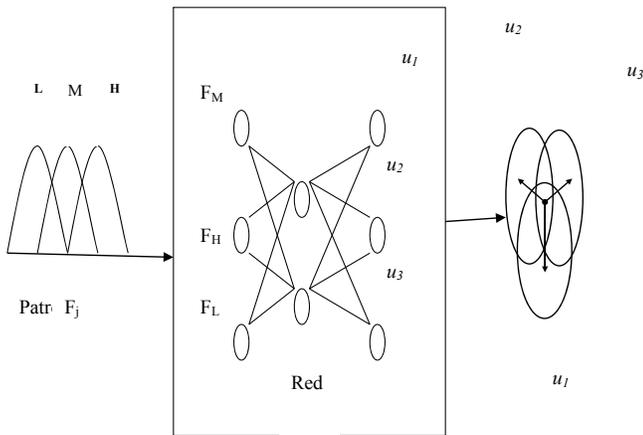


Figura 1. Modelo Razonamiento aproximado basado en Redes Neuronales con retropropagación

El procesamiento de la capa de entrada es resumido en el diagrama de bloques de la figura 2.

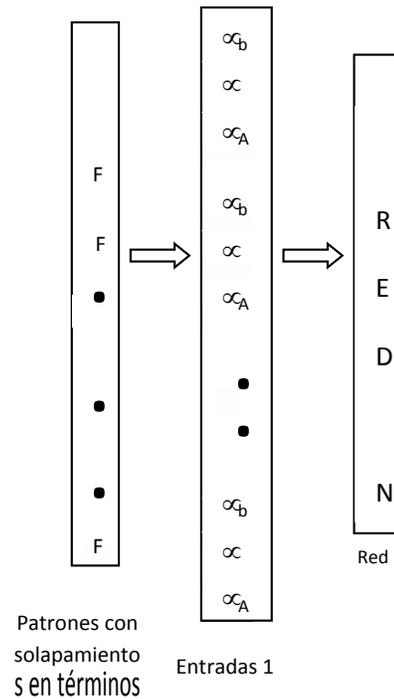


Figura 2. Diagrama de bloques del modelo FMLP

El algoritmo para este modelo fue desarrollado en Matlab bajo un ambiente gráfico y se probó su desempeño para la clasificación de patrones los cuales tendrán un solapamiento previamente definido. Se toma como datos iniciales un conjunto de patrones pertenecientes a un conjunto de datos sintéticos, diseñados de forma que el solapamiento entre los patrones esté definido anticipadamente en términos de porcentajes. El solapamiento se incrementa progresivamente en los patrones en forma controlada de manera que permita analizar y estudiar la influencia del etiquetado difuso en la red neuronal y medir el desempeño de la misma.

La generación de reglas y los aspectos asociados a la inferencia de conocimiento en modelos de reconocimiento de patrones neurodifusos es la etapa final en el proceso del FMLP. La figura 3 muestra una vista general de las etapas involucradas en el proceso de inferencia y generación de reglas.

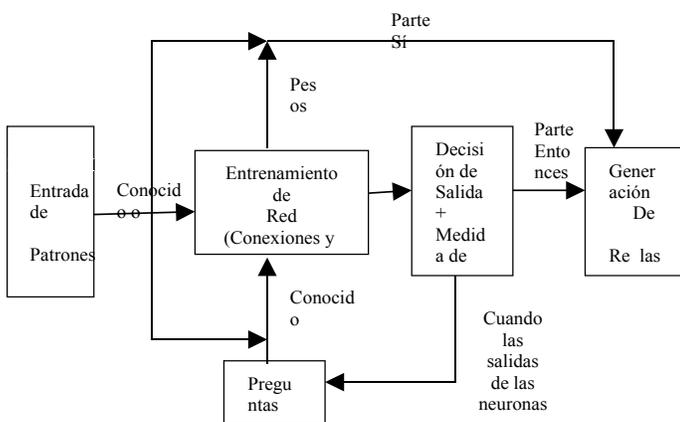


Figura 3: Proceso de clave para criptografía asimétrica.

Figura 3. Diagrama de bloques de inferencia y de la fase de generación de reglas del sistema FMLP

El valor usado para escoger formas lingüísticas tales como muy parecido, parecido, más o menos, no parecido e incapaz de reconocer, así como otros detalles constructivos como el proceso de encontrar las trayectorias de máxima ponderación en la red y generar las reglas difusas, y el algoritmo general para el etiquetamiento difuso en forma generalizada están basados en las explicaciones de Pal y Mitra (1999).

Para la fase de experimentación se diseñaron dos conjuntos de datos bidimensionales, los datos con solapa-

miento entre clases en forma circular y los datos con solapamiento entre clases en forma cuadrada, con el objeto de estudiar si la forma del solapamiento tiene efecto de alguna manera en el proceso de entrenamiento de la red. En la figura 4 se muestran los parámetros a considerar para generar los datos en forma circular.

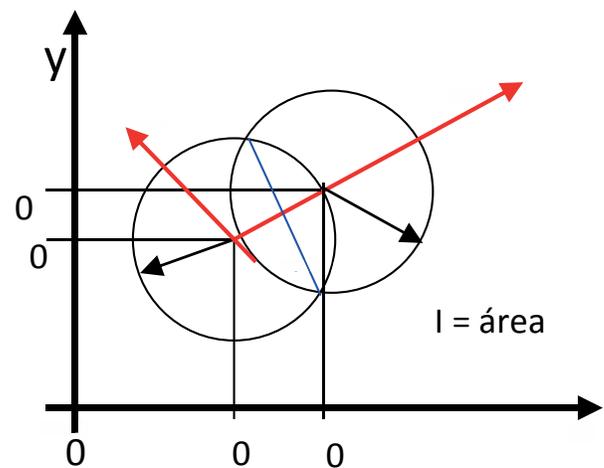


Figura 4. Parámetros en la data con solapamiento en zona circular.

Se generó un conjunto de datos con 152 patrones bidimensionales, 91 patrones para entrenar, 15 patrones para prueba y 46 para validar en forma aleatoria. Los valores de los patrones se generaron en el intervalo $[0, 1]$ de esta forma se evita la normalización.

Se generó un conjunto de datos sintéticos con solapamiento entre clases en forma cuadrada en forma análoga a la data circular con valores aleatorios en el intervalo $[0, 1]$ con el fin de evitar la normalización, este conjunto de datos está formado por 91 patrones para entrenar la red, 15 patrones para probar la red y 46 patrones para validar la red para un total de 152 patrones bidimensionales.

RESULTADOS Y DISCUSIÓN

Para el etiquetamiento difuso de los patrones de entrada para los datos sintéticos como para los datos hepáticos se desarrolló un programa bajo el ambiente gráfico de Matlab (GUI). En la figura 5 se muestra la pantalla de éste.

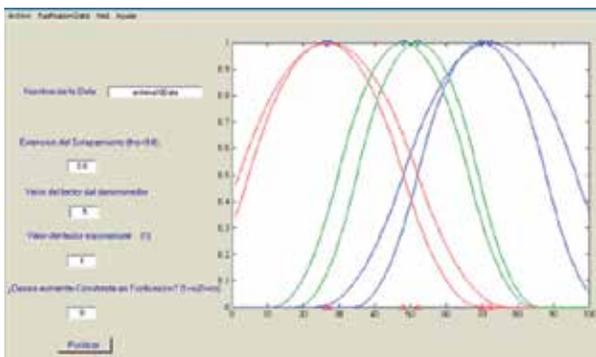


Figura 5. Pantalla de programa para difusificar los patrones de entrada.

Después de pasar los patrones por el proceso de difusificación entran a la red, los datos para el entrenamiento de la red se ingresan en la pantalla como se muestra en la figura 6. En esta pantalla se pide el número de capas ocultas de la red, el número de neuronas por capa oculta, la tasa de aprendizaje, el número de iteraciones y el error máximo permitido, al oprimir el botón entrenar se despliega la gráfica dinámica que se muestra en la parte derecha.

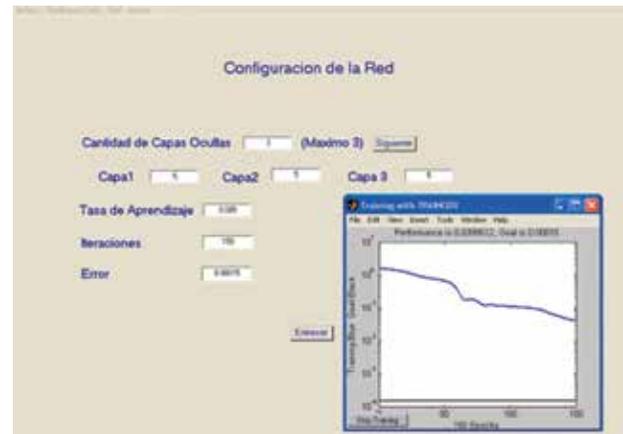


Figura 6. Pantalla que pide los datos para entrenar la red.

Al detenerse la gráfica dinámica que está indicando el desarrollo del entrenamiento (Iteraciones vs error cuadrático medio), se muestra el desempeño de la red con los datos de entrenamiento (patrones mal clasificados) como se muestra en la figura 7. El círculo verde con el centro en rojo representa un patrón que se clasificó bien, los puntos rojos solos o los círculos verdes solos que aparecen en la gráfica representan patrones mal clasificados por la red.

Para probar y validar la red se siguió un procedimiento análogo al anteriormente descrito. Para los datos sintéticos en forma circular y para los datos sintéticos en forma cuadrada se realizaron un total de 144 experimentos por tipo de datos

Como ejemplo se muestran los experimentos realizados con los datos en forma circular para un solapamiento del 40 % y los resultados obtenidos para una red con 6 neuronas en la capa de entrada, una capa oculta con 5, 10 y 15 neuronas, y una capa de salida con 2 neuronas, variando el número de iteraciones entre 1000 y 10.000 (ver la tabla 1).

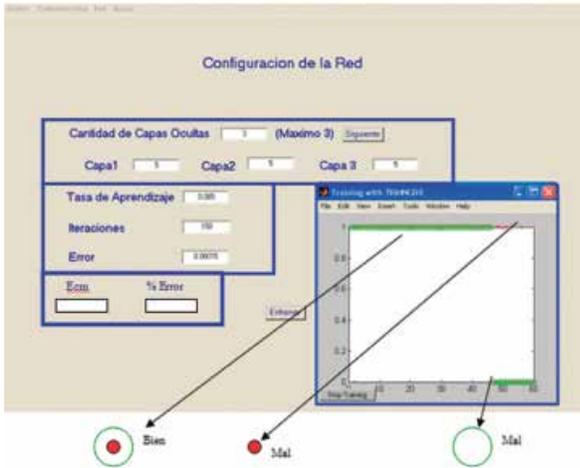


Figura 7. Pantalla que muestra los patrones clasificados

Tabla 1. Resultados obtenidos para un solapamiento del 40 % y una capa escondida.

Data Circular				
% Solapamiento	Red	ECM	% Error	Iteraciones
10	[6 15 15 15 2]	0,10870	10,86760	10000
20	[6 15 15 15 2]	0,10870	10,86760	5000
30	[6 15 15 2]	0,10870	10,86760	10000
40	[6 10 10 2]	0,13040	13,04350	10000
50	[6 15 15 15 2]	0,17390	17,39870	10000

Las redes que arrojaron mejores resultados para cada uno de los porcentajes de solapamientos de los datos sintéticos en forma circular se muestran en la tabla 2.

Tabla 2. Mejores resultados obtenidos para cada uno de los porcentajes de solapamiento.

Red [6 5 2] 40 % solapamiento (91 E , 15 P , 46 V)						
Iteraciones	Entrenamiento		Prueba		Validación	
	MSE	% Error	MSE	% Error	MSE	% Error
1000	0,022	2,1978	0,0667	6,6667	0,1522	15,2174
3000	0,044	4,3956	0,2	20	0,2391	23,913
5000	0	0	0,2667	26,6667	0,1957	19,5652
10000	0	0	0,2	20	0,2391	23,913
Red [6 10 2] 40 % solapamiento (91 E , 15 P , 46 V)						
Iteraciones	Entrenamiento		Prueba		Validación	
	MSE	% Error	MSE	% Error	MSE	% Error
1000	0,033	3,29	0,2	20	0,1087	10,8696
3000	0,033	3,2967	0,1333	13,3333	0,1522	15,2174
5000	0,022	2,1978	0,1333	13,3333	0,2174	21,7391
10000	0	0	0,2	20	0,1739	17,3913
Red [6 15 2] 40 % solapamiento (91 E , 15 P , 46 V)						
Iteraciones	Entrenamiento		Prueba		Validación	
	MSE	% Error	MSE	% Error	MSE	% Error
1000	0,022	2,1978	0,2	20	0,1522	15,2174
3000	0,022	2,1978	0,2	20	0,1957	19,5652
5000	0,011	1,0989	0,1333	13,3333	0,2174	21,7391
10000	0,011	1,0989	0,1333	13,3333	0,1304	13,0435

La generación de las reglas para los resultados sobresalientes de los datos circulares se hace a través del proceso de justificación; se toma la red que presentó el mejor rendimiento por cada porcentaje de solapamiento, se selecciona el mayor de los pesos desde la última capa hasta la primera y se genera la regla de acuerdo a los valores de los pesos. Como ejemplo se toma el caso para el solapamiento del 40 %.

Las reglas para un patrón específico cualquiera en cada configuración mostrada en la tabla 2 se muestran en la Tabla 3.

% SIP	REGLAS
10	REGLA: SI X ES MUY PARECIDO A low y Y ES MUY PARECIDO A médium ALTAMENTE PROBABLE CLASE 1
20	REGLA: SI X ES MUY PARECIDO A high y Y ES MUY PARECIDO A médium ALTAMENTE PROBABLE CLASE 2
30	REGLA: SI X ES MUY PARECIDO A high y Y ES MUY PARECIDO A low ALTAMENTE PROBABLE CLASE 1
40	REGLA: SI X ES MUY PARECIDO A high y Y ES MUY PARECIDO A low ALTAMENTE PROBABLE CLASE 1
50	REGLA: SI X ES MUY PARECIDO A médium y Y ES MUY PARECIDO A médium ALTAMENTE PROBABLE CLASE 22

reglas para un patrón en las mejores redes

El conjunto de reglas que represente el total de patrones de entrenamiento en la mejor red será el universo de las reglas y se generan por los pesos que representan la base de conocimiento dentro de la red.

En cuanto a la experimentación con los datos sintéticos se puede concluir lo siguiente:

Los resultados obtenidos al realizar los experimentos con los datos sintéticos en forma circular y los resultados obtenidos con la data sintética en forma cuadrada son muy similares, lo cual sugiere que la forma geométrica en que se encuentran agrupados los datos y la forma geométrica del solapamiento entre los datos no son relevantes para los resultados del comportamiento del sistema neurodifuso FMLP.

El porcentaje de aciertos de las mejores redes obtenidas con los datos sintéticos disminuyen notablemente a partir del 50% de solapamiento entre clases. Para solapamientos entre 10 % y 40 % se alcanza un buen desempeño de la red con relativa facilidad. El solapamiento entre clases no afecta el entrenamiento de la red, esta se comporta en forma homogénea.

Comparación de los resultados de esta investigación y otros trabajos

Los experimentos sobre los datos sintéticos donde el solapamiento entre clase fue progresivo y porcentualmente creciente se considera pionero; todos los trabajos que se consultaron se han realizado sobre datos con solapamiento aleatorio, por lo tanto los resultados no se pudieron comparar con trabajos anteriores. Para el caso de los datos hepato, en el trabajo (Duch et al, 2004) se presenta los resultados obtenidos utilizando diferentes técnicas para clasificación de patrones. La comparación de varias técnicas en función de la precisión en los datos de entrenamiento y la precisión en los datos de validación se muestran en la tabla 8.

Tabla 8. Resultados obtenidos por diferentes métodos para la clasificación de datos solapados, incluyendo el trabajo realizado.

Método	Precisión datos Entrenamiento	Precisión datos Validación
IB2-IB4	81.2-85.5	43.6-44.6
Clasificador Ba esiano "Naive"	-	46.6
FOIL (lógico inductivo)	99	60.1
LDA (estadístico)	68.4	65.0
Reglas de decisión C4.5	64.5	66.3
MLP con RPROP	-	68
Árbol de decisión C4.5	94.4	75.5
"Feature Space Mapping" con funciones Gaussianas	93	75.6
"Feature Space Mapping" con 60 funciones triangulares	93	75.8
Metodo usado en este trabajo	98.9	75.5

CONCLUSIONES

El sistema difuso neuronal (FMLP) utilizado entrenado apropiadamente tiende a dar respuestas razonables ante vectores de entrada que no fueron utilizados en el entrenamiento, aproximando sus propiedades a los vectores de entrada similares a los empleados en el entrenamiento, para producir una salida correcta. Esta generalización hace posible entrenar una red con un conjunto representativo de pares de entrada / salida y obtener un resultado correcto sin necesidad de entrenar la red con todos los pares de valores posibles de entrada salida. El sistema propuesto cumple a cabalidad con esta propiedad.

El prototipo desarrollado de sistema difuso neuronal para el reconocimiento de patrones con solapamiento aleatorio y controlado entre clases ha sido plenamente probado a través de los experimentos realizados.

Este clasificador difuso se puede utilizar en múltiples disciplinas para diagnosticar enfermedades donde ocurre solapamiento con los patrones reales de sintomatología de una enfermedad específica. El trabajo produce un modelo eficiente de un sistema híbrido difuso neuronal para patrones con comportamiento solapado, con miras a establecer mejoras en cuanto a la rapidez en la determinación de diagnósticos, automatización en el entrenamiento del sistema, y facilidad de interpretación de las salidas del sistema por el usuario.

BIBLIOGRAFÍA

BISHOP, C. (1994) Neural Networks and their applications. Rev.Sci.Instrum., 65, 6, pp.1803-1832.

DUCH, W.; ADARNZAK, R. y GRABCZEWSKI, K. (2001) A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. IEEE trans. Neural Networks, vol 12, PP 277 – 306.

DUCH, W.; SETIONO, R. y ZURADA, J. (2004) Computational Methods for Rule-Based Data Understanding. Proceedings of the IEEE, Vol. 92, No. 5, pp. 771-805.

HAYASHI, Y.; IMURA, A. y Yoshida, K. (1990) “Fuzzy neural expert system and its application to medical diagnosis”, in: 8th International Congress on Cybernetics and Systems, New York City, pp. 54-61.

HAYASHI, Y. (1994) Neural expert system using fuzzy teaching input and its application to medical diagnosis , Information sciences applications, vol 1 pp 47-58

MA, L.; CHEN, C.; TIAN, Z. y HE, W. (1999) Monitoring the particle in CFB Using Fuzzy Neural Network . IEEE Engineering in medicine and biology”, Neural Networks, Vol. 7, No. 2.

MITRA, S. y HAYASHI, Y. (2000) Neuro-fuzzy rule generation: Survey in soft computing framework?, IEEE Transactions on Neural Networks, Vol. 11, pp. 748-768.

PAL, S. y MITRA, S. (1992) Multilayer perceptron, fuzzy sets, and classification. IEEE Transactions on Neural Networks, Vol. 3, No. 5.

PAL, S. y MITRA, S. (1999) Neuro-Fuzzy Pattern Methods in soft Computing, Indian Statistical Institute Calcuta.. Editorial Wiley and Sons, INC.

RAMIREZ, C., y VLADIMIROVA T. A (1995) fuzzy neural network for fuzzy classification En IEEE International Conference on Systems, Man and Cybernetics. Vancouver, Canada, pp. 322-327.