Ελληνικό Στατιστικό Ινστιτούτο
Πρακτικά 21ου Πανελληνίου Συνεδρίου Στατιστικής (2008), σελ 249-258

ΕΝΤΟΠΙΣΜΟΣ ΑΛΛΑΓΗΣ ΤΗΣ ΔΥΝΑΜΙΚΗΣ ΚΑΤΑΣΤΑΣΗΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕ ΜΕΤΡΑ ΠΛΗΡΟΦΟΡΙΑΣ

Α. Παπάνα, Δ. Κουγιουμτζής
Γενικό Τμήμα Πολυτεχνικής σχολής Α.Π.Θ.
agpapana@gen.auth.gr, dkugiu@gen.auth.gr

ΠΕΡΙΛΗΨΗ

Ο εντοπισμός των δυναμικών αλλαγών ενός συστήματος αποτελεί ένα αναπτυσσόμενο πρόβλημα στην ανάλυση χρονοσειρών με ποικίλες εφαρμογές, όπως στην πρόβλεψη επιληπτικών κρίσεων, σεισμών ή καιρού. Στόχος της εργασίας είναι να αξιολογήσει την ικανότητα μέτρων πληροφορίας να διαχωρίζουν χρονοσειρές από διαφορετικά συστήματα, καθώς και να εντοπίζουν ομαλές μεταβολές στην δυναμική κατάσταση ενός συστήματος. Τα μέτρα που αξιολογούνται είναι γραμμικά και μη γραμμικά μέτρα συσχέτισης και εντροπίας, καθώς και ένα νέο μέτρο που ορίζεται ώστε να εκφράζει την απόκλιση μιας χρονοσειράς από την κανονικότητα. Επιπλέον, ελέγχουμε αν βελτιώνεται η διακριτική ικανότητα των μέτρων όταν υπολογίζονται από μετασχηματισμούς των χρονοσειρών. Η διακριτική ικανότητα των μέτρων αξιολογείται με Monte Carlo προσομοιώσεις από στοχαστικά και χαοτικά συστήματα και αναλύεται στατιστικά με τις καμπύλες *ROC*. Αρκετά μέτρα πληροφορίας αποδείχτηκαν χρήσιμα για τον εντοπισμό δυναμικών αλλαγών ενός συστήματος. Ως εφαρμογή, εξετάστηκαν χρονοσειρές εγκεφαλογραφήματος επιληπτικού ασθενούς. Τα μέτρα πληροφορίας εντόπισαν διαφορετική εγκεφαλική λειτουργία στον ασθενή 4 ώρες και μία ώρα πριν από την επιληπτική κρίση.

1. ΕΙΣΑΓΩΓΗ

Βασικό πρόβλημα στην ανάλυση χρονοσειρών είναι ο εντοπισμός της αλλαγής της δυναμικής κατάστασης ενός συστήματος από κάποια πραγματοποίησή του. Η δυναμική κατάσταση ενός συστήματος εκφράζει την πολυπλοκότητά του, δηλαδή τη μη-γραμμικότητα και τη στοχαστικότητά του. Ο εντοπισμός των δυναμικών αυτών αλλαγών γίνεται με παρακολούθηση κάποιων χαρακτηριστικών μέτρων που βασίζονται σε ανακατασκευή του χώρου καταστάσεων (Trulla et al., 1996, Hively et al., 2000, Tykierko, 2008), στην θεωρία πληροφορίας (Hively et al., 2000) ή συγχρονισμού (Mormann et al., 2000). Η μελέτη αυτή επικεντρώνεται σε μέτρα πληροφορίας. Η αξιολόγηση των μέτρων γίνεται με Monte Carlo προσομοιώσεις σε διάφορα συστήματα και η στατιστική αξιολόγησή τους πραγματοποιείται με καμπύλες *ROC* (Receiver Operating Characteristic curves).

2. ΜΕΤΡΑ ΠΛΗΡΟΦΟΡΙΑΣ

Για τη μέτρηση των γραμμικών συσχετίσεων μιας χρονοσειράς $\{x_t\}$, t=1,...,n (μήκους n) με μέση τιμή $\overline{x}$, χρησιμοποιείται η δειγματική συνάρτηση αυτοσυσχέτισης $r(\tau)$. Το $r(\tau)$ αποτελεί ένα μέτρο της «μνήμης» ενός συστήματος που εκτιμά τη συσχέτιση της μεταβλητής της χρονοσειράς σε διαφορετικές χρονικές υστερήσεις

$$r(\tau) = \frac{\sum_{i=1}^{n-\tau}(x_i - \overline{x})(x_{i+\tau} - \overline{x})}{\sum_{i=1}^{n-\tau}(x_i - \overline{x})^2}. \tag{1}$$

Ορίζουμε ένα νέο μέτρο συσχέτισης, την αθροιστική συνάρτηση απολύτων αυτοσυσχετίσεων $cr(\tau_{\max})$, δηλαδή το άθροισμα των απολύτων τιμών των $r(\tau)$ για υστερήσεις τ=1,..,$\tau_{\max}$, όπου $\tau_{\max}$ είναι η υστέρηση που αντιστοιχεί στην αποσυσχέτιση της χρονοσειράς (ο υπολογισμός γίνεται με βάση την αμοιβαία πληροφορία και δίνεται παρακάτω). Ένα ακόμα μέτρο της μνήμης ενός συστήματος είναι ο χαρακτηριστικός χρόνος ή χρόνος αποσυσχέτισης που ορίζεται ως η ελάχιστη υστέρηση ώστε $r(\tau)$=1/e.

Ένα γενικό μέτρο συσχέτισης που περιλαμβάνει και μη γραμμικές συσχετίσεις είναι η αμοιβαία πληροφορία. Για δυο διακριτές τυχαίες μεταβλητές X και Y, η αμοιβαία πληροφορία εκφράζει τη μείωση της αβεβαιότητας της μιας μεταβλητής λόγω της ύπαρξης της άλλης και ορίζεται ως

$$I(X,Y) = \sum_x \sum_y p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}, \tag{2}$$

όπου $p_X(x)$, $p_Y(y)$, $p_{X,Y}(x,y)$ οι περιθώριες συναρτήσεις μάζας πιθανότητας και η από κοινού συνάρτηση μάζας πιθανότητας των X, Y και (X, Y) αντίστοιχα (Kantz & Schreiber, 1997). Για συνεχείς τυχαίες μεταβλητές, η $I(X, Y)$ ορίζεται όπως στην (2) αλλά με χρήση ολοκληρωμάτων. Για στάσιμες χρονοσειρές η αμοιβαία πληροφορία χρονικής υστέρησης τ ορίζεται ως $I(\tau)=I(X_t, X_{t-\tau})$, για $t = \tau+1,.., n$. Εδώ θεωρούμε τις εκτιμήσεις της αμοιβαίας πληροφορίας σε χρονοσειρά: με διαμέριση των τιμών κάθε μεταβλητής σε διαστήματα ίσου μήκους (ED), ίσης πιθανότητας (EP) ή κατάλληλα προσαρμοσμένη διαμέριση (AD) (Darbellay & Vajda, 1999), και με γειτονιές k-κοντινότερων σημείων (KNN) (Kraskov et al., 2004). Υπάρχουν και άλλοι εκτιμητές που δεν χρησιμοποιούνται στη παρούσα εργασία (Papana & Kugiumtzis, 2008). Επίσης, ορίζουμε την αθροιστική συνάρτηση αμοιβαίας πληροφορίας $cI(\tau_{\max})$ ως το άθροισμα των $I(\tau)$, για τ=1,…,$\tau_{\max}$. Για να ορίσουμε το $\tau_{\max}$ υπολογίζουμε τη μέση τιμή I_0 και την τυπική απόκλιση s_0 της $I(\tau)$ για ένα εύρος μεγάλων υστερήσεων όπου οι μεταβλητές X_t και $X_{t-\tau}$ αποσυσχετίζονται (για τα συστήματα που μελετήσαμε θεωρήσαμε τ=90,…,100). Ως $\tau_{\max}$ ορίζουμε την πρώτη υστέρηση για την οποία η αμοιβαία πληροφορία εισέρχεται μέσα στα όρια $I_0 \pm s_0$ και παραμένει για τουλάχιστον τρεις διαδοχικές υστερήσεις. Το $\tau_{\max}$ δεν είναι σταθερό, αλλά αλλάζει σε κάθε χρονοσειρά για αυτό και το χρησιμοποιούμε ως ακόμα ένα μέτρο που εκφράζει το χρόνο γενικής αποσυσχέτισης της χρονοσειράς (συμβολίζεται *dct)*.

Η θεωρητική τιμή του $I(\tau)$ είναι γνωστή μόνο για κανονικές διαδικασίες με συνάρτηση αυτοσυσχέτισης $\rho(\tau)$ (π.χ. αυτοπαλινδρομούμενο σύστημα με κανονικό θόρυβο εισόδου) $I_G(\tau)=-0.5\cdot log(1-\rho^2(\tau))$ (Kullback, 1959). Αν $I(\tau)$ η εκτιμώμενη αμοιβαία πληροφορία (με *ED*, *EP*, *AD, KNN*) και $I_G(\tau)$ η αναμενόμενη τιμή της αν η διαδικασία ήταν κανονική, τότε ορίζουμε την αθροιστική συνάρτηση απόκλισης από την κανονικότητα $cdI(\tau_{\max})$ ως το άθροισμα των διαφορών $|I(\tau)- I_G(\tau)|$, για $\tau=1,..,\tau_{\max}$.

Η εντροπία είναι μέτρο της τυχαιότητας ενός συστήματος. Η εντροπία Shannon μιας διακριτής μεταβλητής X δίνεται ως $H(X)=-\sum p(x)\cdot log\, p(x)$ (με ολοκλήρωμα για συνεχή τυχαία μεταβλητή), όπου $p(x)$ η συνάρτησης μάζας πιθανότητας της X. Ορίζουμε δύο μέτρα με βάση την εντροπία του Shannon. Το πρώτο είναι η εντροπία για τη διδιάστατη μεταβλητή $(X_t, X_{t-\tau})$ με τιμές από τη χρονοσειρά και για τον υπολογισμό της θεωρούμε διαμέριση των τιμών της X_t και της $X_{t-\tau}$ σε διαστήματα ίσου μήκους (συμβολίζουμε $ShEnt(\tau)$). Το δεύτερο μέτρο εντροπίας θεωρεί ως μεταβλητή ένα παράγωγο μέγεθος από την ανάλυση ποσοτικοποίησης επαναληψιμότητας (*Recurrence Quantification Analysis* (Eckmann et al., 1987)). Συγκεκριμένα από κάθε χρονοσειρά μήκους n, σχηματίζουμε τα διανύσματα $\mathbf{x}_i=(x_i,x_{i+\tau},\ldots,x_{i+(m-1)\tau})$, $\mathbf{x}_j=(x_j,x_{j+\tau},\ldots,x_{j+(m-1)\tau})$, $i, j=1,.., n-(m-1)\tau$, και τον τετραγωνικό πίνακα $\mathbf{R}$ με $n-(m-1)\tau$ γραμμές και στήλες, όπου στην (i, j) θέση έχει μονάδα, αν η ευκλείδια απόσταση των $\mathbf{x}_i$ και $\mathbf{x}_j$ είναι μικρότερη από κάποιο κατώφλι ε, αλλιώς έχει μηδέν (θέσαμε ως ε το 15% της τυπικής απόκλισης της χρονοσειράς). Η μεταβλητή X στο δεύτερο μέτρο εντροπίας (συμβολίζουμε $ShEntRQA(m)$) είναι το μήκος γραμμών του $\mathbf{R}$ με μονάδες που είναι παράλληλες στην κεντρική διαγώνιο (για $i=j$).

Η εντροπία Tsallis (Tsallis, 1988) είναι ένα μέτρο που εκφράζει τη μακροσκοπική μνήμη ενός συστήματος και στη διακριτή περίπτωση ορίζεται ως

$$S_q(p)=\frac{1}{q-1}\left(1-\sum_x p^q(x)\right) \tag{3}$$

για κάποιο q (σταθερά) που χαρακτηρίζει τη μνήμη του συστήματος. Το μέτρο της εντροπίας Tsallis υπολογίζεται όπως το $ShEnt(\tau)$ για τη μεταβλητή $(X_t, X_{t-\tau})$ (συμβολίζουμε $TsEnt(\tau,q)$).

Το τελευταίο μέτρο πολυπλοκότητας που εξετάσαμε είναι η δειγματική εντροπία (Richman & Moorman, 2000) που εξετάζει αν διαδοχικές τιμές μιας χρονοσειράς επαναλαμβάνονται με κάποια απόκλιση (αντιστοιχεί στο ε του πίνακα επανεμφάνισης). Αν A είναι ο μέσος όρος των εμφανίσεων όλων των δυνατών τμημάτων m διαδοχικών τιμών στη χρονοσειρά και B αντίστοιχα για $(m+1)$ διαδοχικές τιμές, τότε η δειγματική εντροπία ορίζεται ως $-\ln(A/B)$.

3. ΣΥΣΤΗΜΑΤΑ ΠΡΟΣΟΜΟΙΩΣΕΩΝ

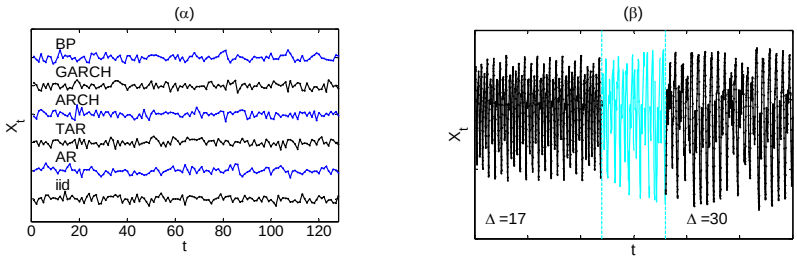
Αρχικά εξετάστηκε αν τα μέτρα διαχωρίζουν μεταξύ διαφορετικών στοχαστικών συστημάτων. Η αξιολόγησή τους έγινε με Monte Carlo προσομοιώσεις σε έξι στοχαστικά συστήματα (Πίνακας 1). Οι παράμετροι των συστημάτων επιλέχθηκαν ώστε οι χρονοσειρές που προκύπτουν να έχουν μηδενικές αυτοσυσχετίσεις (εκτός του

AR(1)) και μετασχηματίστηκαν ώστε να έχουν μέση τιμή 0 και τυπική απόκλιση 1 (Εικόνα 1α) για να αποφευχθούν πιθανές διακρίσεις τους λόγω διαφορετικών αυτοσυσχετίσεων ή της κλίμακας τιμών τους. Συγκεκριμένα, εξετάστηκε αν τα μέτρα διαχωρίζουν τον λευκό θόρυβο ή το AR(1) από τα υπόλοιπα συστήματα. Επίσης, εξετάστηκε αν τα μέτρα διαχωρίζουν μεταξύ χρονοσειρών από συστήματα διαφορετικής πολυπλοκότητας. Για το σκοπό αυτό θεωρήσαμε το χαοτικό σύστημα Mackey-Glass (Mackey & Glass, 1977) με διαφορική εξίσωση υστέρησης $dx(t)/dt=-0.1x(t)+0.2x(t-\Delta)/(1+x(t-\Delta)^{10})$ (η χρονοσειρά παράγεται για χρόνο διακριτοποίησης 0.1 και χρόνο δειγματοληψίας 17). Οι τιμές της παραμέτρου Δ στην εξίσωση του συστήματος καθορίζουν την πολυπλοκότητά του (μεγάλες τιμές του Δ αντιστοιχούν σε μεγάλη πολυπλοκότητα).

Πίνακας 1. *Τύποι – εξισώσεις στοχαστικών συστημάτων.*

Α/Α	***Στοχαστικά συστήματα***	***Τύποι - Εξισώσεις***
1	*Κανονικός λευκός θόρυβος*	$e_t \sim N(0,1)$
2	*Αυτοπαλινδρομούμενο AR*(1)	$X_t = 0.2X_{t-1}+e_t$
3	*Αυτοπαλινδρομούμενο κατωφλίου* (*TAR*)	$X_t = \begin{cases} -0.5X_{t-1}+e_t, & \text{αν } X_{t-1}<1 \\ 0.4X_{t-1}+e_t, & \text{αλλιώς} \end{cases}$
4	*Αυτοπαλινδρομούμενο υπό συνθήκη ετεροσκεδαστικό ARCH*(1)	$X_t = \sqrt{1+0.4X_{t-1}^2}+e_t$
5	*Γενικευμένο αυτοπαλινδρομούμενο υπό συνθήκη ετεροσκεδαστικό ARCH*(1,1)	$X_t = \sqrt{0.01+0.8h_{t-1}+0.15X_{t-1}^2}e_t$
6	*Διγραμμική διαδικασία* (*BP*)	$X_t = 0.6e_{t-1}X_{t-2}+e_t$

Εικόνα 1. *Μια πραγματοποίηση (α) των στοχαστικών συστημάτων με n*=128*, όπως δίνονται στην λεζάντα (μέση τιμή* 0*, τυπική απόκλιση* 1*), (β) του συστήματος Mackey-Glass, όπου το πρώτο και τελευταίο τμήμα στο γράφημα αντιστοιχούν σε* Δ=17 *και* 30 *αντίστοιχα, ενώ στο ενδιάμεσο τμήμα* Δ∈[18,29] *(σταδιακή αύξηση).*



Για κάθε στοχαστικό σύστημα σχηματίζονται 100 πραγματοποιήσεις (χρονοσειρές) μήκους *n*=128, 256, 2048, στις οποίες υπολογίζονται τα μέτρα. Τα μέτρα υπολογίζονται επίσης και σε μετασχηματισμένες πραγματοποιήσεις. Θεωρούμε δύο μετασχηματισμούς: σε ομοιόμορφη περιθώρια κατανομή, $u_i=F_X(x_i)$, όπου F η εμπειρική αθροιστική συνάρτηση κατανομής και σε κανονική περιθώρια κατανομή, $z_i=\Phi^{-1}(F_X(x_i))$, όπου Φ η αθροιστική συνάρτηση πυκνότητας της τυπικής κανονικής κατανομής. Για το διαχωρισμό χρονοσειρών από το χαοτικό σύστημα Mackey-Glass θεωρούμε 5 διαφορετικές τιμές της παραμέτρου Δ=17, 30, 100, 200, 300, οι οποίες

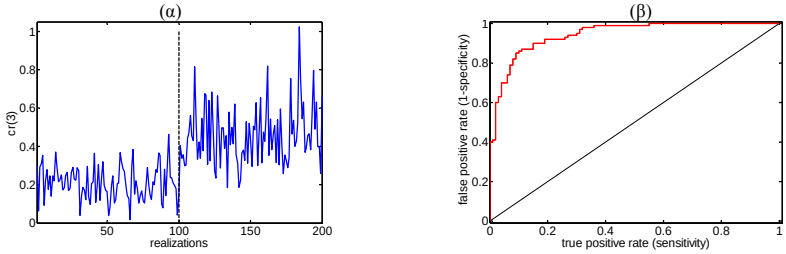
αντιστοιχούν σε 5 καταστάσεις διαφορετικής πολυπλοκότητας του συστήματος. Για κάθε μια από τις 5 περιπτώσεις σχηματίζουμε 100 πραγματοποιήσεις του συστήματος με μήκη n=2000, 4000, από τις οποίες και υπολογίζουμε τις τιμές των μέτρων.

Ο δεύτερος στόχος της εργασίας είναι να ελέγξει αν τα μέτρα εντοπίζουν ομαλές μεταβολές της δυναμικής κατάστασης ενός συστήματος και συγκεκριμένα δυναμικές αλλαγές κατά την μετάβαση του συστήματος Mackey-Glass από μια κατάστασή του (για κάποιο Δ) σε μια άλλη (για κάποιο άλλο Δ). Η μετάβαση προσομοιώνεται με σταδιακή μεταβολή (αύξηση) του Δ (Εικόνα 1β). Σχηματίζουμε μια πραγματοποίηση του συστήματος με 2.800.000 σημεία ως εξής: θεωρούμε 400.000 σημεία (100 τμήματα μήκους n=4000 ή 200 τμήματα με n=2000) για Δ=17, 30, 100, 200, 300 (5 καταστάσεις) και 200.000 σημεία (50 τμήματα με n=4000 ή 100 τμήματα με n=2000) για τα ενδιάμεσα τμήματα (4 μεταβατικές περίοδοι όπου γίνεται σταδιακή αύξηση του Δ). Τα μέτρα υπολογίζονται από μη-επικαλυπτόμενα τμήματα της χρονοσειράς μήκους n=2000 και n=4000 (για όλες τις καταστάσεις και μεταβατικές περιόδους).

4. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

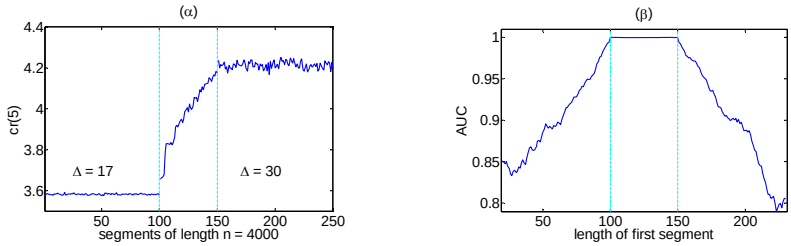
Για τη στατιστική αξιολόγηση κάθε μέτρου σχηματίζουμε ένα σύνολο τιμών του από τις 100 πραγματοποιήσεις του κάθε συστήματος. Ονομάζουμε ως δείγματα Α, Β δυο σύνολα τιμών του ίδιου μέτρου από δύο διαφορετικά συστήματα και συγκρίνουμε τις κατανομές των δειγμάτων με βάση τις καμπύλες *ROC* (Green & Swets, 1966). Οι τιμές κάθε μέτρου πρέπει να ταξινομηθούν σε δυο κατηγορίες: αν η τιμή του μέτρου ξεπερνά μια σταθερά (κατώφλι απόφασης) τότε ταξινομείται στην $1^{η}$ κατηγορία, αλλιώς ταξινομείται στην $2^{η}$. Ευαισθησία (sensitivity) είναι ο λόγος του πλήθους τιμών του Α που ταξινομούνται σωστά στο Α προς το πλήθος των τιμών που ταξινομούνται στο Α, ενώ το συμπληρωματικό της ειδικότητας (1-specificity) είναι ο λόγος του πλήθους τιμών του Β που ταξινομούνται στο Α προς το πλήθος των τιμών που ταξινομούνται στο Β. Η καμπύλη *ROC* είναι η γραφική παράσταση του (sensitivity) ως προς το (1-specificity) μεταβάλλοντας το κατώφλι απόφασης από την τιμή κατωφλιού που ταξινομεί όλες τις τιμές των μέτρων στην $1^{η}$ κατηγορία έως την τιμή κατωφλιού που ταξινομεί όλες τις τιμές των μέτρων στην $2^{η}$ κατηγορία. Το εμβαδόν της επιφάνειας κάτω από την καμπύλη (*AUC*) είναι ένας δείκτης της σωστής ταξινόμησης (για τέλεια διάκριση *AUC*=1). Ένα παράδειγμα δίνεται για τα δείγματα *cr*(3) από λευκό θόρυβο και AR(1), τα οποία διαφέρουν σημαντικά (Εικόνα 2).

Εικόνα 2. *(α) Αθροιστικές αυτοσυσχετίσεις cr*(3) *από λευκό θόρυβο και AR*(1) *από* 100 *πραγματοποιήσεις με n*=128*, (β) καμπύλη ROC από την διάκρισή τους (AUC=0.944).*



Για τον εντοπισμό δυναμικών αλλαγών στο σύστημα Mackey-Glass θεωρούμε τη σειρά των τιμών κάθε μέτρου που σχηματίζεται από τις τιμές του μέτρου από δύο διαφορετικές καταστάσεις του συστήματος και την ενδιάμεση μεταβατική περίοδο. Σε αυτή τη σειρά των τιμών θεωρούμε ένα σημείο διαχωρισμού K που ξεκινά από τη θέση 20 και μετακινείται με βήμα 1 έως και 20 θέσεις πριν το τέλος της, το οποίο χωρίζει τη σειρά σε δύο δείγματα A, B. Το *AUC* υπολογίζεται για όλα τα πιθανά δείγματα που προκύπτουν μεταβάλλοντας το K και ελέγχουμε σε ποιο K έχουμε τη βέλτιστη διάκριση (μέγιστο *AUC*) (Εικόνα 3).

Εικόνα 3*.(α) Αθροιστικές αυτοσυσχετίσεις cr*(5) *από το σύστημα Mackey-Glass για* Δ=17, $\Delta\in$[18,29] *(μεταβατική περίοδος), και* Δ=30 *(n*=4000). *(β) AUC για όλα τα δυνατά δείγματα από το διαχωρισμό της χρονοσειράς του cr*(5).



5. ΑΠΟΤΕΛΕΣΜΑΤΑ

Ο διαχωρισμός του λευκού θορύβου από τα υπόλοιπα στοχαστικά συστήματα επιτυγχάνεται με αρκετά μέτρα με βέλτιστο την εντροπία Shannon (*RQA*). Για τον διαχωρισμό του AR(1) από τα υπόλοιπα συστήματα παρατηρείται και πάλι πολύ καλή διάκριση με αναμενόμενο βέλτιστο μέτρο το $cr(\tau_{max})$ (καθώς μόνο το AR(1) έχει σημαντικές αυτοσυσχετίσεις). Ο μετασχηματισμός των χρονοσειρών σε κανονική περιθώρια κατανομή βελτιώνει τη διακριτική ικανότητα των μέτρων, ενώ σε ομοιόμορφη έχει το αντίθετο αποτέλεσμα. Συγκεντρωτικά από τις τιμές του *AUC* από όλες τις διακρίσεις μεταξύ των στοχαστικών συστημάτων παρατηρούμε ότι υπάρχει μια συνέπεια ως προς την κατάταξη των μέτρων για όλα τα n. Η εντροπία Shannon (με *RQA*) έχει τη βέλτιστη διακριτική ικανότητα (Πίνακας 2).

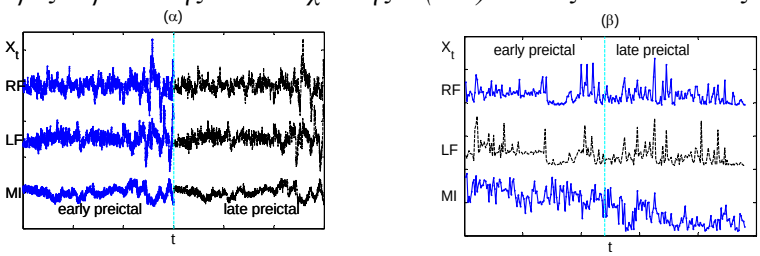
Πίνακας 2. *Τα τρία βέλτιστα μέτρα από τη διάκριση του λευκού θορύβου και του AR(*1*) από τα υπόλοιπα στοχαστικά συστήματα για όλα τα μήκη χρονοσειρών n, από τις αρχικές και μετασχηματισμένες χρονοσειρές. mAUC είναι ο μέσος όρος των AUC από όλες τις διακρίσεις και όλα τα n.*

Αρχική χρονοσειρά		***Χρονοσειρά με κανονική περιθώρια κατανομή***		***Χρονοσειρά με ομοιόμορφη περιθώρια κατανομή***	
Μέτρα	***mAUC***	***Μέτρα***	***mAUC***	***Μέτρα***	***mAUC***
ShEntRQA(2)	0.865	ShEntRQA(2)	0.867	ShEntRQA(2)	0.827
cr(3)	0.830	cdlED(3)	0.824	TsEnt(1,1.5)	0.813
cr(6)	0.819	cr(3)	0.819	ShEnt(1)	0.811

Η διάκριση των καταστάσεων του συστήματος Mackey-Glass μεταξύ Δ=17 και 30 και μεταξύ Δ=30 και 100 ήταν τέλεια με όλα σχεδόν τα μέτρα (*AUC*=1). Η διάκριση μεταξύ Δ=100 και 200 ήταν τέλεια μόνο για τα αθροιστικά μέτρα συσχέτισης (τ_{max}=100) και το χρόνο αποσυσχετισμού (με *EP*, *AD*), καθώς και για την εντροπία Shannon (*RQA*, *m*=10, 15, 20). Καθώς το σύστημα γίνεται πιο πολύπλοκο παρατηρούμε ότι η επιλογή των παραμέτρων των μέτρων είναι σημαντική. Τέλος, ικανοποιητική διάκριση καταστάσεων μεταξύ Δ=200 και 300 γίνεται μόνο με την εντροπία Shannon (*RQA*, *m*=15, 20). Τα μέτρα πληροφορίας εντόπισαν και ομαλές δυναμικές αλλαγές στο σύστημα Mackey-Glass. Η μετάβαση του συστήματος από Δ=17 σε 30 και από Δ=30 σε 100 εντοπίστηκε τέλεια (*AUC*=1) σχεδόν με όλα τα μέτρα και πάντα μέσα στη μεταβατική περίοδο. Η μετάβαση από Δ=100 σε 200 εντοπίστηκε τέλεια με τα αθροιστικά μέτρα συσχέτισης (τ_{max}=100). Κατά την μετάβαση από Δ=200 σε 300 μόνο η εντροπία Shannon (*RQA*, *m*=15,20) και ο χρόνος αποσυσχετισμού (με *AD*, *KNN*) εντόπισαν ικανοποιητικά την αλλαγή.

Ως εφαρμογή εξετάζουμε τη δυνατότητα διάκρισης του δυναμικού στον εγκέφαλο επιληπτικού ασθενή την τελευταία μία ώρα πριν την κρίση (*late preictal*) και από 4 έως 3 ώρες πριν την κρίση (*early preictal*). Η διάκριση των καταστάσεων γίνεται σε έξω-κρανιακές καταγραφές ηλεκτροεγκεφαλογραφημάτων (EEG) από τρία κανάλια στο κεντρικό (MI), αριστερό (LF) και δεξιό μπροστινό (RF) μέρος του εγκεφάλου. Για τις καταγραφές EEG από κάθε κατάσταση και κανάλι δημιουργούμε διαδοχικές μη επικαλυπτόμενες χρονοσειρές διάρκειας 30sec και υπολογίζουμε τα μέτρα. Η επιλογή του καναλιού είναι καθοριστική, καθώς μόνο για το κανάλι MI σχεδόν όλα τα μέτρα διακρίνουν ικανοποιητικά μεταξύ των καταστάσεων (Εικόνα 4, Πίνακας 3).

Εικόνα 4. *Για τα 3 κανάλια: (α) τμήμα καταγραφής EEG 30sec από τις 2 καταστάσεις, (β) χρονοσειρές αθροιστικής αυτοσυσχέτισης cr(*100*) από τις 2 καταστάσεις.*



Πίνακας 3*. Τα 3 βέλτιστα μέτρα της διάκρισης late-early preictal καταστάσεων.*

Κεντρικό κανάλι		***Αριστερό Κανάλι***		***Κανάλι 21***	
Μέτρο	***mAUC***	***Μέτρο***	***mAUC***	***Μέτρο***	***mAUC***
cr(100)	0.947	cr(5)	0.711	SamEn(2)	0.676
cr(50)	0.940	cr(10)	0.708	cIED(5)	0.661
dctEP	0.935	SamEn(2)	0.708	cr(5)	0.655

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Αρχικά αναφέρουμε ότι το μήκος των χρονοσειρών δεν επηρεάζει την κατάταξη των μέτρων ως προς την διακριτική τους ικανότητα. Τα μέτρα πληροφορίας διαχωρίζουν ικανοποιητικά συστήματα διαφορετικής πολυπλοκότητας και εντοπίζουν ομαλές δυναμικές μεταβολές. Ο μετασχηματισμός της χρονοσειράς με περιθώρια κανονική κατανομή βελτιώνει τη διακριτική ικανότητα των μέτρων, ειδικά στην περίπτωση της διάκρισης του λευκού θορύβου. Τα μέτρα πληροφορίας φάνηκαν χρήσιμα και για τον διαχωρισμό καταστάσεων στην επιληψία. Σημαντική είναι η επιλογή του καναλιού και των παραμέτρων των μέτρων, η οποία πρέπει να διερευνηθεί περαιτέρω.

Το έργο (ΠΕΝΕΔ) συγχρηματοδοτείται κατά 90% κοινά από την Ε.Ε.: Ευρωπ. Κοινωνικό Ταμείο (75%), Υπ. Ανάπτυξης, Γ.Γ.Ε.Τ. (25%) και Rikshospitalet, Νορβηγίας (10%), στο πλαίσιο του Μέτρου 8.3 του Ε.Π. Ανταγωνιστικότητα – Γ΄ Κ.Π.Σ..

ABSTRACT

Detection of dynamical changes of a system is a developing area of time series analysis with many applications, e.g. prediction of epileptic seizures, earthquakes and weather. Our goal is to evaluate the discriminating ability of information measures in detecting different dynamical states of systems. We assess measures of linear and nonlinear correlations, and entropy measures. Moreover, we investigate whether the discriminating power of the measures is improved by estimating them from transformations of the time series. We define a new measure that expresses the deviation of a time series from normality. The discriminating ability of the measures is assessed by ROC curves on samples of measure values calculated on Monte Carlo realizations from stochastic and chaotic systems. As an application we investigate the detection of dynamical changes in the brain activity of an epileptic patient. Information measures turn out to discriminate well different dynamical states and are able to detect smooth dynamical changes, also in real applications.

ΑΝΑΦΟΡΕΣ

Darbellay G. and Vajda I. (1999). Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, **45** (4), 1315-1321.

Eckmann J.P., Kamphorst S.O. and Ruelle D. (1987). Recurrence plots of dynamical systems. *Europhysics Letter,* **4** (9), 973-977.

Green D.M. and Swets J.A. (1966). *Signal detection theory and psychophysics*. John Wiley and Sons, New York.

Hively L.M., Protopopescu V.A. and Gailey P.C. (2000). Timely detection of dynamical change in scalp EEG signals. *Chaos,* **10** (4), 864-875.

Kantz H. and Schreiber T. (1997). *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge.

Kraskov A., Stogbauer H. and Grassberger P. (2004). Estimating mutual information. *Physical Review E,* **69**, 066138.

Kullback S. (1959). *Information theory and statistics.* John Wiley and Sons, NY.

Mackey M.C. and Glass L. (1977). Oscillation and chaos in physiological control systems. *Science,* **197** (4300), 287-289.

Mormann F., Lehnertz K., David P. and Elger C.E. (2000). Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D,* **144**, 358-369.

Papana A. and Kugiumtzis D. (2008). Evaluation of mutual information estimators on nonlinear dynamic systems. *Nonlinear Phenomena in Complex Systems,* **11** (2), 225-232.

Richman J.S. and Moorman J.R. (2000). Physiological time series analysis using approximate entropy and sample entropy. *American Journal of Physiology, Heart and Circulatory Physiology,* **278** (6)**,** H2039-49**.**

Shannon C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal,* **27**, 379–432 and 623-656.

Trulla L., Giuliani A., Zbilut J.P. and Webber C.L. (1996). Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A,* **223**, 255-260.

Tsallis C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics,* **52**, 479-487.

Tykierko M. (2008). Using invariants to change detection in dynamical system with chaos. *Physica D,* **237**, 6–13.