

Νέες Τεχνικές Βελτίωσης της Αποδοτικότητας των Εξελικτικών Πιθανοτικών Νευρωνικών Δικτύων

B. Α. Γεωργίου, Φ. Α. Αλεβίζος, Μ. Ν. Βραχάτης

Εργαστήριο Υπολογιστικής Νοημοσύνης (CI Lab), Τμήμα Μαθηματικών,
University of Patras Artificial Intelligence Research Center (UPAIRC),

Πανεπιστήμιο Πατρών, 26110 Πάτρα,
{vlg, philipos, vrahatis}@math.upatras.gr

Περίληψη

Στην εργασία αυτή προτείνεται μια μεθοδολογία βελτίωσης της αποδοτικότητας των Εξελικτικών Πιθανοτικών Νευρωνικών Δικτύων (ΕΠΝΔ). Η βελτίωση αυτή επιτυγχάνεται ενσωματώνοντας διάφορες μορφές των αλγορίθμων bagging καθώς και με τη στάθμιση των εξόδων του ΕΠΝΔ με κατάλληλα βάρη. Το κλασικό πιθανοτικό νευρωνικό δίκτυο (ΠΝΔ) είναι ένα μοντέλο που χρησιμοποιείται για ταξινόμηση και αναγνώριση προτύπων σε προκαθορισμένες κλάσεις. Στην πραγματικότητα είναι μια τροποποιημένη μορφή της Διαχωριστικής Ανάλυσης με χρήση πυρήνων (Kernel Discriminant Analysis) η οποία ενσωματώνει τον κανόνα ταξινόμησης του Bayes, δηλαδή ταξινομεί τις άγνωστες παρατηρήσεις σύμφωνα με τις εκτιμήσεις των εκ των υστέρων πιθανοτήτων τους. Η επιλογή των παραμέτρων λείανσης των πυρήνων του ΠΝΔ επιτυγχάνεται από τον εξελικτικό αλγόριθμο Βελτιστοποίησης Σμήνους Σωματιδίων (ΒΣΣ).

Οι αλγόριθμοι bagging δημιουργούν μια συστάδα από μοντέλα αποφάσεων και υπολογίζουν το τελικό αποτέλεσμα βάση ενός πλειοψηφικού συστήματος. Επίσης η κατάλληλη στάθμιση των εξόδων των πυρήνων του ΕΠΝΔ με τη ΒΣΣ προσφέρει μεγαλύτερο βαθμό ελευθερίας στο μοντέλο να προσαρμοστεί καλύτερα στα δεδομένα. Το προτεινόμενο σχήμα ενσωματώνει τις παραπάνω τεχνικές σε ένα ΕΠΝΔ το οποίο εφαρμόζεται σε προβλήματα βιοϊατρικής από τον πραγματικό κόσμο (real-world applications).

1. ΕΙΣΑΓΩΓΗ

Τα τελευταία χρόνια έχει παρατηρηθεί μια ραγδαία εξέλιξη των μοντέλων τα οποία έχουν την ικανότητα ταξινόμησης και αναγνώρισης προτύπων. Τα προβλήματα της βιοϊατρικής είναι μια κατηγορία προβλημάτων στην οποία χρησιμοποιούνται νευρωνικά δίκτυα καθώς και άλλα σύγχρονα μοντέλα [Baldi and Brunak (2001)]. Ένα από αυτά τα μοντέλα είναι και το Πιθανοτικό Νευρωνικό Δίκτυο (ΠΝΔ) [Specht (1990)] το οποίο έχει εφαρμοστεί με επιτυχία σε προβλήματα βιοϊατρικής [Huang (2002)] αλλά και σε άλλα επιστημονικά πεδία πάνω σε προβλήματα ταξινόμησης [Ganchev *et al.* (2004)].

Τα Πιθανοτικά Νευρωνικά Δίκτυα (ΠΝΔ) ανήκουν σε μια κατηγορία νευρωνικών δικτύων που συνδυάζουν τα χαρακτηριστικά της στατιστικής αναγνώρισης προτύπων καθώς και αυτά των νευρωνικών δικτύων εμπρόσθιας τροφοδότησης (feedforward neural networks, FNNs). Τα ΠΝΔ συνδυάζουν τη διαχωριστική ανάλυση (discriminant

analysis), χρησιμοποιώντας συναρτήσεις πυρήνων, και τον κανόνα ταξινόμησης του Bayes. Το απλό ΠΝΔ μπορεί να χαρακτηριστεί σαν μια ευφυής μνήμη διότι κάθε πρότυπο εκπαίδευσης (training pattern) αποθηκεύεται σε ένα νευρώνα του δικτύου [Berthold and Diamond (1998)].

Ένας σημαντικός παράγοντας που επηρεάζει την απόδοση ενός ΠΝΔ είναι οι παράμετροι λείανσης (smoothing parameters) των πυρήνων των ΠΝΔ. Ένας τρόπος βελτίωσης της αποδοτικότητας των ΠΝΔ είναι η ενσωμάτωση του αλγόριθμου Βελτιστοποίησης Σμήνους Σωματιδίων (ΒΣΣ) για την εκτίμηση των παραμέτρων λείανσης του ΠΝΔ [Georgiou *et al.* (2006)]. Το μοντέλα αυτά ονομάζονται Εξελικτικά Πιθανοτικά Νευρωνικά Δίκτυα (ΕΠΝΔ).

Στην παρούσα εργασία προτείνονται τρόποι βελτίωσης της αποδοτικότητας των ΕΠΝΔ. Για κάθε κλάση νευρώνων χρησιμοποιείται διαφορετικός πίνακας των παραμέτρων λείανσης, η στάθμιση των εξόδων των πυρήνων του ΠΝΔ επιτυγχάνεται με χρήση της ΒΣΣ και τέλος ενσωματώνεται και η τεχνική bagging για την περαιτέρω βελτίωση της απόδοσης των ΕΠΝΔ.

Το προτεινόμενο σχήμα έχει εφαρμοστεί σε δύο σύνολα δεδομένων από το χώρο της βιοϊατρικής, το Wisconsin Breast Cancer Database (WBCD) και το Heart Disease (HD) από τη βάση δεδομένων Proben1 [Prechelt (1994)]. Επίσης τα αποτελέσματα αυτά έχουν συγκριθεί με τα αντίστοιχα αποτελέσματα που παρήχθησαν από νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης.

2. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ

Για λόγους πληρότητας, παρουσιάζονται περιληπτικά κάποιες βασικές έννοιες. Τα ΠΝΔ ανήκουν στην κατηγορία μοντέλων με επίβλεψη (supervised) τα οποία χρησιμοποιούνται ευρέως στην ταξινόμηση και ομαδοποίηση των δεδομένων. Η τελική απόφαση ενός ΠΝΔ για την ταξινόμηση μιας άγνωστης οντότητας σε K προκαθορισμένες κλάσεις βασίζεται στον κανόνα απόφασης του Bayes και τη μη παραμετρική εκτίμηση της συνάρτησης πυκνότητας πιθανότητας του Parzen [Parzen (1962), Specht (1990)]. Ένα σημαντικό προτέρημα των ΠΝΔ, σε σύγκριση με τα FNNs, είναι ότι προσφέρουν τρόπους ερμηνείας της δομής του δικτύου [Berthold and Diamond (1998)], ενώ στα FNNs είναι δύσκολο να ερμηνευθεί τι αναπαριστά ο κάθε νευρώνας του δικτύου. Η τυπική διαδικασία εκπαίδευσης ενός ΠΝΔ απαιτεί ένα μόνο πέρασμα όλων των προτύπων του συνόλου εκπαίδευσης [Specht (1990)]. Το χαρακτηριστικό αυτό μεταφράζεται σε πιο γρήγορη εκπαίδευση του ΠΝΔ από την εκπαίδευση ενός FNN.

Η δομή ενός ΠΝΔ είναι παρόμοια με αυτή ενός FNN μόνο που η αρχιτεκτονική του ΠΝΔ περιορίζεται σε τέσσερα επίπεδα, που είναι το *επίπεδο εισόδου, προτύπων, άθροισης* και *εξόδου*. Τα δεδομένα υπό μορφή διανυσμάτων $X = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ τοποθετούνται στους p νευρώνες εισόδου και στη συνέχεια περνούν στο επίπεδο

προτύπων. Οι νευρώνες του επιπέδου προτύπων είναι χωρισμένοι σε K κλάσεις, όσες και οι κλάσεις στις οποίες είναι ταξινομημένα τα δεδομένα. Η έξοδος του i -οστού νευρώνα προτύπων της k -οστής κλάσης υπολογίζεται με τη χρήση ενός κανονικού (Gaussian) πυρήνα της μορφής:

$$f_{i,k}(X) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(X - X_{i,k})^\top \Sigma^{-1}(X - X_{i,k})\right), \quad (1)$$

όπου $X_{i,k} \in \mathbb{R}^p$ είναι το κέντρο του πυρήνα και Σ ο $p \times p$ πίνακας των παραμέτρων λείανσης. Ως $X_{i,k}$ θεωρούμε το i -οστό διάνυσμα παρατηρήσεων της k -οστής κλάσης των δεδομένων. Το επίπεδο άθροισης του δικτύου υπολογίζει τη δεσμευμένη πιθανότητα της κάθε κλάσης,

$$G_k(X) = \pi_k \cdot \sum_{i=1}^{M_k} f_{i,k}(X), \quad k \in \{1, \dots, K\}, \quad (2)$$

όπου M_k είναι ο αριθμός των νευρώνων του επιπέδου προτύπων της k -οστής κλάσης και π_k είναι η εκ των προτέρων πιθανότητά της. Ένα διάνυσμα X ταξινομείται σε εκείνη την κλάση στην οποία ο αντίστοιχος νευρώνας άθροισης αποκτά τη μεγαλύτερη τιμή, δηλαδή έχει τη μεγαλύτερη εκ των υστέρων πιθανότητα.

Μια ταχύτερη εκδοχή του ΠΝΔ μπορεί να υλοποιηθεί χρησιμοποιώντας μόνο ένα μέρος του συνόλου εκπαίδευσης και όχι ολόκληρο το σύνολο. Ένα τέτοιο σύνολο εκπαίδευσης μπορεί να κατασκευαστεί είτε με τυχαίο τρόπο είτε αναζητώντας κάποιους «αντιπροσώπους» του συνόλου εκπαίδευσης μέσω μιας τεχνικής ομαδοποίησης. Με τη χρήση του αλγόριθμου ομαδοποίησης K -means [MacQueen (1967)], αναγνωρίζεται ένας επαρκής αριθμός αντιπροσώπων (κέντρων) από κάθε κλάση του συνόλου εκπαίδευσης. Δηλαδή παράγονται κάποια κέντρα από κάθε κλάση τα οποία χρησιμοποιούνται ως κέντρα των πυρήνων του ΠΝΔ αντί να χρησιμοποιηθεί όλο το σύνολο εκπαίδευσης. Αυτό οδηγεί σε ένα αρκετά μικρότερο και ταχύτερο ΠΝΔ. Συγκεκριμένα ο αριθμός των κέντρων που έχουν εξαχθεί μέσω του αλγορίθμου K -means είναι το 10% του μεγέθους του συνόλου εκπαίδευσης. Δηλαδή, το μέγεθος του επιπέδου προτύπων του προτεινόμενου ΠΝΔ είναι 10 φορές μικρότερο από το απλό ΠΝΔ.

Όπως αναφέρθηκε και παραπάνω, η απόδοση ενός ΠΝΔ επηρεάζεται σε μεγάλο βαθμό από την επιλογή του πίνακα των παραμέτρων λείανσης. Για την εκτίμηση των τιμών των παραμέτρων λείανσης που θα δώσουν ικανοποιητικά αποτελέσματα, χρησιμοποιήθηκε ο αλγόριθμος ΒΣΣ [Georgiou *et al.* (2006)].

Η ΒΣΣ είναι ένας στοχαστικός αλγόριθμος βελτιστοποίησης που βασίζεται σε πληθυσμούς πιθανών λύσεων [Kennedy and Eberhart (1995)]. Η βασική ιδέα του αλγορίθμου είναι η εξελικτική διερεύνηση για «καλές» λύσεις στο χώρο αναζήτησης, ξεκινώντας από έναν πληθυσμό αρχικών λύσεων. Ο πληθυσμός καλείται *σμήνος* και τα

άτομα, δηλαδή τα σημεία αναζήτησης, καλούνται *σωματίδια*. Κάθε σωματίδιο κινείται με μεταβλητή ταχύτητα μέσα στο χώρο αναζήτησης και διατηρεί σε μια μνήμη την καλύτερη θέση που έχει περάσει. Σε κάθε επανάληψη η θέση αυτή ανακοινώνεται και στα υπόλοιπα σωματίδια του σμήνους. Εδώ χρησιμοποιήθηκε η παραλλαγή με παράγοντα περιορισμού (“constriction factor”) της ΒΣΣ [Parsopoulos and Vrahatis (2002) and (2004)].

Μια τεχνική που έχει προταθεί για τη βελτίωση της απόδοσης ενός μοντέλου ταξινόμησης είναι η τεχνική *bagging* [Breiman (1996)]. Έστω \mathcal{L} ένα σύνολο εκπαίδευσης που αποτελείται από τα $\{(\vec{x}_n, y_n), n = 1, \dots, N_{\text{train}}\}$. Σε ένα πρόβλημα ταξινόμησης όπου ένα άγνωστο διάνυσμα \vec{x} πρέπει να ταξινομηθεί σε μία από τις K προκαθορισμένες κλάσεις, τα y_i παίρνουν τιμές στο $\{1, 2, \dots, K\}$. Ανεξαρτήτως του μοντέλου που χρησιμοποιείται, τελικά κατασκευάζεται ένας ταξινομητής $\phi(\vec{x}, \mathcal{L})$. Συγκεκριμένα, υποθέτουμε ότι έχουμε μια ακολουθία από σύνολα εκπαίδευσης \mathcal{L}_m όπου το κάθε ένα αποτελείται από N_{train} ανεξάρτητες παρατηρήσεις που προέρχονται από την ίδια κατανομή με το \mathcal{L} . Ο στόχος μας είναι να κατασκευάσουμε έναν καλύτερο ταξινομητή από τον $\phi(\vec{x}, \mathcal{L})$ χρησιμοποιώντας την ακολουθία $\{\mathcal{L}_m\}$.

Συνήθως διαθέτουμε ένα μόνο σύνολο δεδομένων για την εκπαίδευση του μοντέλου, οπότε εξάγουμε m bootstrap δείγματα από το \mathcal{L} . Κάθε bootstrap δείγμα αποτελείται από N_{train} διανύσματα τα οποία εξάγονται τυχαία από το \mathcal{L} με επανατοποθέτηση. Κάποια διανύσματα από το \mathcal{L} μπορεί να μην επιλεγθούν και κάποια άλλα μπορεί να επιλεγθούν περισσότερες φορές. Κατασκευάζουμε έναν ταξινομητή $\phi(\vec{x}, \mathcal{L}_m)$ από κάθε bootstrap δείγμα \mathcal{L}_m . Για να συναθροιστούν όλοι οι ταξινομητές $\{\phi(\vec{x}, \mathcal{L}_m)\}$ χρησιμοποιείται μια πλειοψηφική διαδικασία. Έστω N_k ο αριθμός των φορών που οι ταξινομητές ψήφισαν υπέρ της κλάσης k ($N_k = \#\{m, \phi(\vec{x}, \mathcal{L}_m) = k\}$, $k = 1, \dots, K$). Η τελική ταξινόμηση επιτυγχάνεται με τον κανόνα του μέγιστου για τα $\{N_k\}$. Με άλλα λόγια, ο τελικός ταξινομητής είναι $\phi_B(\vec{x}, \mathcal{L}) = \text{argmax}_k(N_k)$. Η παραπάνω διαδικασία ονομάζεται “bootstrap aggregating” και χρησιμοποιείται το ακρωνύμιο **bagging**.

3. Η ΠΡΟΤΕΙΝΟΜΕΝΗ ΠΡΟΣΕΓΓΙΣΗ ΚΑΙ ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Η επιλογή κατάλληλων τιμών των παραμέτρων λείανσης των πυρήνων του ΠΝΔ συνήθως επιτυγχάνεται μέσω μιας διαδικασίας δοκιμής και διόρθωσης. Επίσης έχουν προταθεί διάφορες εναλλακτικές μέθοδοι [Georgiou *et al.* (2006), Gorunescu *et al.* (2005)]. Το Εξελικτικό Πιθανοτικό Νευρωνικό Δίκτυο (ΕΠΝΔ), που προτάθηκε στο [Georgiou *et al.* (2006)], ενσωματώνει τον αλγόριθμο ΒΣΣ για την επιλογή των κατάλληλων παραμέτρων λείανσης. Η επιλογή του Σ , πίνακα των παραμέτρων λείανσης, επιτυγχάνεται με κριτήριο την ελαχιστοποίηση του ποσοστού εσφαλμένης ταξινόμησης του συνόλου εκπαίδευσης ή επαλήθευσης. Ο πίνακας Σ είναι ένας διαγώνιος

πίνακας του οποίου τα στοιχεία μπορεί να είναι ίσα ή να διαφέρουν.

Στην παρούσα εργασία έχουμε επεκτείνει την παραπάνω ιδέα έτσι ώστε κάθε ομάδα νευρώνων στο επίπεδο προτύπων του ΠΝΔ να έχει το δικό της πίνακα Σ_k , $k = 1, \dots, K$. Αυτό μπορεί να οδηγήσει σε καλύτερη προσαρμογή των συναρτήσεων πυρήνων στα δεδομένα, καθώς η κατανομή των δεδομένων μιας κλάσης μπορεί να διαφέρει από την κατανομή κάποιας άλλης κλάσης. Παρακάτω δίνονται οι τέσσερις παραλλαγές που χρησιμοποιήσαμε για τον πίνακα των παραμέτρων λείανσης με την αντίστοιχη διάσταση d της ελαχιστοποίησης της ΒΣΣ ($k = 1, \dots, K$):

$$\begin{array}{ll} 1. \Sigma_k = \sigma^2 \cdot I_p, & d = 1 \\ 2. \Sigma_k = \sigma_k^2 \cdot I_p, & d = k \\ 3. \Sigma_k = \text{diag}(\sigma_1^2, \dots, \sigma_p^2), & d = p \\ 4. \Sigma_k = \text{diag}(\sigma_{1k}^2, \dots, \sigma_{pk}^2), & d = k \cdot p \end{array}$$

Εκτός από τη βελτιστοποίηση των παραμέτρων λείανσης του ΠΝΔ, η ΒΣΣ χρησιμοποιήθηκε και για την καλύτερη εκτίμηση των εκ των προτέρων πιθανοτήτων π_k των κλάσεων. Οι εκ των προτέρων πιθανότητες των κλάσεων δεν υπολογίζονται αυθαίρετα και δεν εκτιμώνται από το σύνολο εκπαίδευσης αλλά βελτιστοποιούνται με τη ΒΣΣ μαζί με τις παραμέτρους λείανσης. Αυτό επιτρέπει στο ΠΝΔ να έχει μια μεγαλύτερη ευελιξία και να προσαρμόζεται καλύτερα στα δεδομένα, δηλαδή δίνει ένα μεγαλύτερο βαθμό ελευθερίας στην προσαρμογή των δεδομένων. Προφανώς τα π_k είναι περιορισμένα να αθροίζονται στη μονάδα. Στην πράξη τα π_k λειτουργούν ως ένας διαφορετικός τρόπος στάθμισης της εξόδου κάθε πυρήνα του ΠΝΔ.

Όπως προαναφέρθηκε, ένας τρόπος βελτίωσης της απόδοσης ενός μοντέλου ταξινόμησης είναι να δημιουργήσουμε μια συστάδα ταξινομητών και η τελική απόφαση ταξινόμησης είναι η πλειοψηφούσα πρότασή τους. Η τεχνική που χρησιμοποιήθηκε στην εργασία μας είναι η τεχνική bagging. Παράγουμε μια ακολουθία από bootstrap δείγματα από το σύνολο εκπαίδευσης και από κάθε ένα από τα δείγματα κατασκευάζουμε ένα ΕΠΝΔ. Ένα άγνωστο διάνυσμα \vec{x} ταξινομείται στην κλάση που θα συλλέξει τις περισσότερες προτάσεις των ΕΠΝΔ. Αυτή η τεχνική επιτυγχάνει να κατασκευάσει έναν πιο ανθεκτικό και σταθερό ταξινομητή.

Η παραπάνω προσέγγιση εφαρμόστηκε σε δύο γνωστά σύνολα δεδομένων από το χώρο της βιοϊατρικής, συγκεκριμένα τα HD και WBCD. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από τη βάση δεδομένων Proben1 και έχουμε ακολουθήσει τις προτεινόμενες οδηγίες. Επίσης η αποδοτικότητα του προτεινόμενου ΕΠΝΔ συγκρίθηκε με αυτή των FNNs. Όλα τα αποτελέσματα των FNNs που παραθέτονται είναι τα καλύτερα αποτελέσματα που έχουν επιτευχθεί από το Proben1.

Παρακάτω παρουσιάζεται μια σύντομη περιγραφή των δεδομένων. Το πρώτο σύνολο δεδομένων είναι το HD και στόχος μας είναι να προβλέψουμε αν σε τουλάχιστον μία από τις τέσσερις κύριες βαλβίδες της καρδιάς έχει μειωθεί η διάμετρος κατά 50% ή παραπάνω. Η απόφαση βασίζεται σε προσωπικά δεδομένα όπως φύλο, ηλικία, κά-

Πίνακας 1. Ποσοστό επιτυχούς ταξινόμησης των μοντέλων

Συν. Δεδ.	Μοντέλο	Μέση Τιμή	Τυπ. Απόκλιση	Min.	Max.
HD	ΕΠΝΔ	80.46	1.10	78.26	83.04
	ΕΠΝΔ bagging	81.82	0.45	81.30	82.17
	FNN	80.11	2.27	-	-
WBCD	ΕΠΝΔ	97.67	1.60	97.13	98.85
	ΕΠΝΔ 4η Παραλ.	98.81	0.58	97.13	99.43
	FNN	98.53	0.60	-	-

πνισμα και αποτελέσματα από διάφορες ιατρικές εξετάσεις όπως πίεση αίματος και ηλεκτροκαρδιογράφημα. Έχουμε 920 άτομα και 35 μεταβλητές (στην υλοποίηση χρησιμοποιήθηκαν μόνο οι 32 μεταβλητές διότι οι υπόλοιπες ήταν σταθερές). Το δεύτερο σύνολο δεδομένων είναι το WBCD και ο στόχος είναι να προβλεφθεί αν ένας όγκος στο μαστό είναι καλοήθης ή όχι. Υπάρχουν 9 μεταβλητές (κυτταρικές μετρήσεις) και το δείγμα αποτελείται από 699 ασθενείς.

Εφαρμόσαμε τα προτεινόμενα μοντέλα πρόβλεψης στα δύο σύνολα δεδομένων HD και WBCD και συγκρίναμε τα αποτελέσματα μας με τα αντίστοιχα των FNN του Proben1. Για το σύνολο δεδομένων της καρδιάς η δεύτερη παραλλαγή του πίνακα Σ έδωσε τα καλύτερα αποτελέσματα, δηλαδή χρησιμοποιήσαμε διαφορετικό πίνακα Σ_k για κάθε κλάση αλλά με κοινά στοιχεία στη διαγώνιο. Από το σύνολο εκπαίδευσης πήραμε 10 bootstrap δείγματα και βασισμένοι σε αυτά κατασκευάσαμε μια συστάδα από 10 ΕΠΝΔ. Στον Πίνακα 1 παρουσιάζεται ο μέσος όρος, η τυπική απόκλιση καθώς και η ελάχιστη και μέγιστη τιμή των ποσοστών της επιτυχούς ταξινόμησης για τα σύνολα ελέγχου των δεδομένων HD και WBCD, τόσο για το απλό ΕΠΝΔ, όσο και για το προτεινόμενο ΕΠΝΔ καθώς και το FNN. Το προτεινόμενο ΕΠΝΔ πέτυχε την καλύτερη μέση επίδοση σε σχέση με τα άλλα μοντέλα. Στα δεδομένα του HD, πραγματοποιώντας ένα διορθωμένο *t*-test [Bouckaert and Frank (2004)] για τη σύγκριση των μέσων ποσοστών επιτυχίας του προτεινόμενου ΕΠΝΔ και του FNN, συμπεραίνουμε ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ τους (p -value=0.0316). Επιπλέον, η τυπική απόκλιση του ποσοστού επιτυχούς ταξινόμησης του προτεινόμενου ΕΠΝΔ ήταν 5 φορές μικρότερη από αυτή του FNN και 2 φορές μικρότερη από αυτή του απλού ΕΠΝΔ.

Στο σύνολο δεδομένων WBCD, η τεχνική bagging δεν έδωσε καλύτερα αποτελέσματα αλλά η τέταρτη παραλλαγή του πίνακα Σ και ο νέος τρόπος εκτίμησης των εκ των προτέρων πιθανοτήτων με τη ΒΣΣ βελτίωσαν την απόδοση του ΕΠΝΔ. Με άλλα λόγια, υπήρχε ένας διαγώνιος πίνακας παραμέτρων λείανσης για κάθε κλάση νευρώνων του επιπέδου προτύπων ΕΠΝΔ του οποίου τα στοιχεία δεν ήταν απαραίτητα ίσα. Επίσης, η ΒΣΣ έδωσε καλύτερες εκτιμήσεις για τις εκ των προτέρων πιθανότητες που

οδήγησαν σε μεγαλύτερα ποσοστά επιτυχίας. Συγκρίνοντας τα μέσα ποσοστά επιτυχίας του προτεινόμενου ΕΠΝΔ και του FNN, συμπεραίνουμε ότι δεν υπάρχει στατιστικά σημαντική διαφορά (p -value=0.553), δηλαδή το προτεινόμενο ΕΠΝΔ έδωσε παρόμοιο ποσοστό επιτυχίας με το FNN.

4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία προτάθηκαν διάφοροι τρόποι για τη βελτίωση της αποδοτικότητας των ΕΠΝΔ. Προτείνεται μια επέκταση της μορφής του πίνακα των παραμέτρων λείανσης του ΕΠΝΔ, έτσι ώστε κάθε κλάση νευρώνων του ΕΠΝΔ να έχει τον δικό της διαφορετικό πίνακα. Αυτό επιτρέπει στο μοντέλο να προσαρμόζεται καλύτερα στα δεδομένα. Επίσης, προτάθηκε ένας νέος τρόπος, για τη στάθμιση των εξόδων των πυρήνων του ΕΠΝΔ με τη χρήση της ΒΣΣ, που είναι διαφορετικός του συνηθισμένου τρόπου των εκτιμήσεων των εκ των προτέρων πιθανοτήτων από το δείγμα. Ένας επιπλέον τρόπος βελτίωσης της απόδοσης των ΕΠΝΔ ήταν και η ενσωμάτωση της τεχνικής bagging. Από τα πειραματικά αποτελέσματα, είναι προφανές ότι το προτεινόμενο ΕΠΝΔ πέτυχε παρόμοια ή και ανώτερη επίδοση από τα FNN στα δύο προβλήματα που μελετήθηκαν από το χώρο της βιοϊατρικής. Επιπλέον, η διασπορά του ποσοστού επιτυχούς ταξινόμησης του ΕΠΝΔ έχει μειωθεί αρκετά χρησιμοποιώντας την τεχνική bagging, δηλαδή το νέο ΕΠΝΔ είναι πολύ πιο ανθεκτικό.

5. ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστούμε το Επιχειρησιακό Πρόγραμμα Εκπαίδευσης και Αρχικής Επαγγελματικής Κατάρτισης II (ΕΠΕΑΕΚ II) και συγκεκριμένα το πρόγραμμα ΗΡΑΚΛΕΙΤΟΣ για τη χρηματοδότηση της παραπάνω εργασίας.

ABSTRACT

In this contribution novel approaches for the improvement of the performance of the Evolutionary Probabilistic Neural Networks (EPNN) are proposed. The improvement is accomplished by employing the bagging technique as well as by the proper weighting of the outputs of the EPNN's kernels. The typical Probabilistic Neural Network (PNN) is a model used for classification tasks into predefined classes. In other words, the PNN is a modification of Kernel Discriminant Analysis that employs the Bayes decision rule. It classifies an unknown vector according to the estimations of the posterior probabilities of each class. The selection of the kernels' spread parameters is accomplished by the Particle Swarm Optimization (PSO) algorithm.

The bagging technique creates an ensemble of classification models and classifies an unknown pattern using a voting system. Moreover, the weighting of the kernels' outputs by PSO offers a greater degree of freedom in order for the model to fit properly to the data. The proposed approaches are incorporated to a EPNN that is applied to several real-world classification tasks.

ΑΝΑΦΟΡΕΣ

- Baldi, P. and Brunak, S. (2001). *Bioinformatics: The Machine Learning Approach*. MIT Press.
- Berthold, M. and Diamond, J. (1998). Constructive training of probabilistic neural networks. *Neurocomputing*, pp. 167–183.
- Bouckaert, R. R. and Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. In *Proc Pacific-Asia Conference on Knowledge Discovery and Data Mining*, LNAI 3056, pp. 3–12, Sydney.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, **24**,2, pp. 123–140.
- Ganchev, T., Tasoulis, D. K., Vrahatis, M. N., and Fakotakis, N. (2004). Locally recurrent probabilistic neural networks with application to speaker verification. *GESTS International Transaction on Speech Science and Engineering*, **1**(2), pp. 1–13.
- Georgiou, V. L., Pavlidis, N. G., Parsopoulos, K. E., Alevizos, Ph. D., Vrahatis, M. N. (2006) New self-adaptive probabilistic neural networks in bioinformatic and medical tasks. *Int. Journal on Artificial Intelligence Tools*, **15**, 3, pp. 371–396.
- Gorunescu, M., Gorunescu, F., Ene, M., El-Darzi, E. (2005). A heuristic approach in hepatic cancer diagnosis using a probabilistic neural network-based model. In *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis*, pp. 1016–1024, Brest, France.
- Huang, C. J. (2002). A performance analysis of cancer classification using feature extraction and probabilistic neural networks. In *Proceedings of the 7th Conference on Artificial Intelligence and Applications*, pp. 374–378.
- Kennedy, J. and Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings IEEE International Conference on Neural Networks*, **4**, pp. 1942–1948.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, Berkeley.
- Parsopoulos, K. E. and Vrahatis, M. N. (2002). Recent approaches to global optimization problems through particle swarm optimization. *Natural Computing*, **1**(2–3), pp. 235–306.
- Parsopoulos, K. E. and Vrahatis, M. N. (2004). On the computation of all global minimizers through particle swarm optimization. *IEEE Transactions on Evolutionary Computation*, **8**, 3, pp. 211–224.
- Parzen, E. (1962). On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **3**, pp. 1065–1076.
- Prechelt, L. (1994). Proben1: A set of neural network benchmark problems and benchmarking rules. Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe.
- Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks*, **1**, 3, pp. 109–118.