Θεωρία των Αποδείξεων στην Ταξινόμηση Εικόνων Τηλεπισκόπησης

ΣΤΕΛΙΟΣ Π. ΜΕΡΤΙΚΑΣ
Καθηγητής Πολυτεχνείου Κρήτης

Περίληψη
Στο άρθρο αυτό παρουσιάζεται η «θεωρία των αποδείξεων» στην ταξινόμηση εικόνων Τηλεπισκόπησης. Κύριος στόχος είναι να εισαχθούν οι λιγότερες δημοφιλείς έννοιες των συναρτήσεων αξιοπιστίας στην ταξινόμηση. Η συνάρτηση αξιοπιστίας μπορεί να θεωρηθεί ως μία γενίκευση της κλασικής έννοιας της συνάρτησης πιθανότητας κατά Bayes που περιλαμβάνει, όμως, έναν τρόπο αξιολόγησης της ισχύος μιας στατιστικής υπόθεσης βασιζόμενης σε αποδεικτικά στοιχεία. Τέλος, παρουσιάζονται επτά παραδείγματα για την κατανόηση της «θεωρίας των αποδείξεων».

1. ΕΙΣΑΓΩΓΗ

Η ταξινόμηση είναι μέθοδος κυρίως της πολυδιάστατης Στατιστικής (Duda and Hart, 1973· Anderson, 1984· Awcock and Thomas, 1995· Castleman, 1996) που αφορά στον διαχωρισμό αντικειμένων και την καταχώρισή τους σε δύο ή περισσότερες ομάδες ή τάξεις. Πρώτος στόχος της ταξινόμησης είναι η περιγραφή της "διαφοροποίησης" των αντικειμένων. Δεύτερος στόχος είναι η επιλογή ενός "κανόνα" (κατάλληλου αλγόριθμου), ώστε να διαχωριστούν τα αντικείμενα σε δύο ή περισσότερες ομοειδείς τάξεις.

Η ταξινόμηση της ψηφιακής εικόνας (Jensen, 1995· Richards, 1993· Congalton and Green, 1998· Schowengerdt, 1983· 1997) είναι η διαδικασία αντιστοιχίας ή τοποθέτησης των τιμών φωτεινότητας των εικονοστοιχείων σε ομάδες που αντιστοιχούν σε διαφορετικά επιφανειακά υλικά ή συνθήκες, και τα οποία παρουσιάζουν την ίδια μορφή, τις ίδιες περίπου ιδιότητες. Σε μελέτη αξιολόγησης της ποιότητας του νερού, για παράδειγμα, ένα πρώτο βήμα θα ήταν να χρησιμοποιηθεί η ταξινόμηση της εικόνας, ώστε να αναγνωριστούν όλα τα εικονοστοιχεία της εικόνας που αντιστοιχούν στο νερό. Μετέπειτα, η ταξινόμηση μπορεί να επικεντρωθεί σε λεπτομερέστερη μελέτη των εικονοστοιχείων αυτών, ώστε να χαρτογραφηθεί η ποιότητα του νερού, το βάθος του πυθμένα κ.ο.κ.

Οι περισσότερες παραδοσιακές μέθοδοι της ταξινόμησης, όπως της μέγιστης πιθανοφάνειας, της ελάχιστης απόστασης κ.λπ. (Curran, 1985· Campbell, 1987· Richards, 1993), απαιτούν (1) όλα τα στοιχεία να είναι ψηφιακά, (2) βασίζονται σε παραμετρικά στατιστικά μοντέλα, όπως της κανονικής κατανομής του Gauss, και (3) είναι σχεδιασμένες να μην μπορούν να διαχειρίζονται δεδομένα από διαφορετικές πηγές ή ακρίβειες. Όμως στην πράξη, πολλά δεδομένα δεν ικανοποιούν όλες τις παραπάνω συνθήκες. Η μαθηματική «θεωρία των αποδείξεων» (Theory of Evidence) (Shafer, 1976· Shafer and Pearl, 1990· Yager et al., 1994) ή «θεωρία του Dempster-Shafer» είναι ένα πεδίο, στο οποίο οι πηγές δεδομένων (π.χ. εικόνες, χάρτες, προσωπική εκτίμηση μηχανικού κ.λπ.) - που αντιμετωπίζονται ξεχωριστά και ανεξάρτητα η μία από την άλλη - δεν απαιτούνται να είναι σε ψηφιακή μορφή. Επίσης, στη θεωρία αυτή το στατιστικό μοντέλο δεν απαιτείται να ακολουθείται αυστηρά αλλά να δίνει τη δυνατότητα να αντιμετωπίσει κανείς δεδομένα (πηγές) πολλών, διαφορετικών αλλά και μεγάλων διαστάσεων.

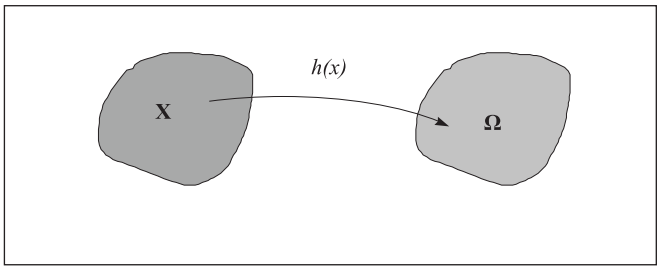
Η θεωρία των αποδείξεων χρησιμοποιείται τελευταία στην τεχνητή νοημοσύνη, όπου με τη θεωρία της μαθηματικής πιθανότητας προσπαθεί να υλοποιήσει την υποκειμενική κρίση.

2. ΑΞΙΟΠΙΣΤΙΑ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ ΥΠΟΘΕΣΗ

Η ταξινόμηση μπορεί να θεωρηθεί μια διαδικασία απεικόνισης από τον χώρο **X** των δεδομένων ή των παρατηρήσεων στον χώρο των τάξεων **Ω** (σχήμα 1). Στην Τηλεπισκόπηση ο χώρος **X** συνήθως εκφράζεται από την πολυφασματική εικόνα που είναι το σύνολο των ψηφίδων. Τα στοιχεία του χώρου **Ω** είναι όλες οι επιζητούμενες τάξεις της εδαφικής κάλυψης (π.χ. έδαφος, βλάστηση, νερά κ.λπ.) που θα προκύψουν από τη διαδικασία της ταξινόμησης.

Έστω ότι σε μία ταξινόμηση εικόνας αναμένεται να προκύψουν τέσσερις τάξεις A={δάσος}, B={καλλιέργεια}, C={νερά}, D={αστική περιοχή} (σχήμα 2). Το πεπερασμένο

Υποβλήθηκε: 17.2.1999 Έγινε δεκτή: 8.7.1999



Σχήμα 1: Ταξινόμηση ως μία διαδικασία απεικόνισης.
Figure 1: Classification as a mapping procedure.

σύνολο όλων αυτών των τάξεων συμβολίζεται με Ω={δάσος, καλλιέργεια, νερά, αστικά} και ονομάζεται *πλαίσιο διακρίσεων* (frame of discernment). Το σύνολο αυτό θεωρείται ότι περιέχει ανεξάρτητα στοιχεία και καλύπτει όλες τις δυνατές τάξεις που ενδέχονται να προκύψουν από την ταξινόμηση.

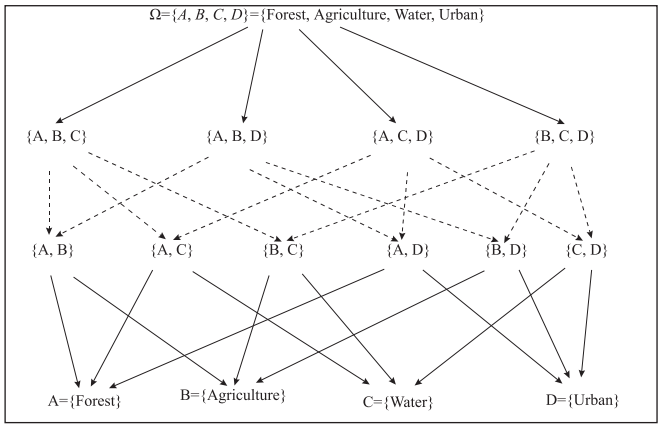
Μερικά αποδεικτικά στοιχεία (π.χ. στοιχεία εκπαίδευσης του αλγορίθμου ταξινόμησης) μπορούν να ωθήσουν τον αναλυτή να δείξει κάποιο βαθμό εμπιστοσύνης για την υπόθεση του συνόλου {*A*, *B*}= {δάσος, καλλιέργεια} του Ω. Η νέα αυτή υπόθεση αντιστοιχεί στην υπόθεση {*A* ή *B*}= {δάσος ή καλλιέργεια}. Επίσης, ένα νέο αποδεικτικό στοιχείο μπορεί να βοηθήσει τον αναλυτή να αποκλείσει την τάξη *A*={δάσος} σε κάποιο βαθμό.

Αποδεικτικά στοιχεία, που δεν επιβεβαιώνουν την τάξη *A,* ισοδυναμούν με επιβεβαίωση της υπόθεσης {όχι *A*} που αντιστοιχεί στην αποδοχή της υπόθεσης {*B* ή *C* ή *D*}. Αποδεικτικά στοιχεία, που δεν επιβεβαιώνουν την τάξη *A,* κάνουν τον αναλυτή να υιοθετήσει ένα ανάλογο ποσοστό εμπιστοσύνης στο σύνολο των τριών τάξεων {*B* ή *C* ή *D*} που απομένουν.

Όλα τα δυνατά υποσύνολα των υποθέσεων, που μπορεί να προκύψουν από το πλαίσιο διακρίσεων Ω, συμβολίζονται με 2^{Ω}. Σε κάθε υπόθεση που προκύπτει μπορεί κανείς να δώσει κάποιο ποσοστό εμπιστοσύνης (αξιοπιστίας). Χρησιμοποιούμε εδώ τον όρο "υπόθεση" με αυτή τη διευρυμένη έννοια για να δηλώσουμε οποιοδήποτε υποσύνολο των αρχικών υποθέσεων (τάξεων) στο Ω.

Σε ένα σύνολο *n* στοιχείων του πλαισίου Ω υπάρχουν 2^n υποσύνολα Ω_i. Το κενό σύνολο, ∅, είναι και αυτό ένα από τα υποσύνολα, αλλά αντιστοιχεί στην υπόθεση ότι είναι ψευδής και δεν φαίνεται, για παράδειγμα, στο σχήμα 2.

Η θεωρία των αποδείξεων χρησιμοποιεί έναν αριθμό στο διάστημα [0, 1] για να δείξει τον βαθμό εμπιστοσύνης (αξιοπιστία) μιας στατιστικής υπόθεσης, δεδομένων κάποιων αποδεικτικών στοιχείων. Ο αριθμός αυτός είναι ο βαθμός με τον οποίο τα αποδεικτικά στοιχεία επιβεβαιώνουν ή απορρίπτουν μια υπόθεση. Σημειώνεται ότι αποδεικτικά στοιχεία που δεν επιβεβαιώνουν μια υπόθεση, θεωρούνται ως αποδεικτικά στοιχεία (αποδείξεις) για την αντιφατική υπόθεση (όχι *A*).



Σχήμα 2: Ταξινόμηση με τη θεωρία των αποδείξεων.
Figure 2: Possible classification from a four-element frame of discernment.

Ο συντελεστής βάρους (βαθμός) των αποδεικτικών στοιχείων σε επιβεβαίωση μιας συγκεκριμένης υπόθεσης Ω_i αντιπροσωπεύεται από μία συνάρτηση που ονομάζεται *βασική πιθανότητα καταχώρισης* (basic probability assignment, bpa). Η βασική πιθανότητα είναι μια γενίκευση της παραδοσιακής συνάρτησης πυκνότητας πιθανότητας κατά Bayes. Συμβολίζεται με *m* και ορίζεται στο διάστημα [0, 1]. Αναφέρεται σε κάθε υποσύνολο Ω_i από τον γενικευμένο χώρο όλων των υποσυνόλων 2^{Ω}. Η βασική πιθανότητα καταχώρισης $m(\Omega_i)$ αντιστοιχεί σε κάθε στοιχείο Ω_i (singleton) του πλαισίου Ω έτσι, ώστε όλες οι τιμές του *m* να αθροίζονται στη μονάδα:

$$\sum_{i=1}^{2^n} m(\Omega_i) = 1 \qquad \forall \Omega_i \in \Omega \tag{2.1}$$

Επομένως, το *m* είναι ένας βαθμός εμπιστοσύνης (αξιοπιστίας) που αναλογεί σε κάθε σύνολο Ω_i του σχήματος 2 και όχι μόνο σε εκείνα τα σύνολα (στοιχεία) της τελευταίας γραμμής, όπως είναι η περίπτωση της συνάρτησης πυκνότητας πιθανότητας. Εξ ορισμού το *m*=0 πρέπει να αντιστοιχεί στο κενό σύνολο, ∅, επειδή το σύνολο αυτό αντιστοιχεί σε ψευδή υπόθεση.

Η ποσότητα *m*(*A*), όπου το *A* είναι ένα στοιχείο του 2^{Ω}, είναι ένα μέτρο της αξιοπιστίας που αντιστοιχεί στο *A* και μόνον στο *A*. Το *m*(*A*) δεν μπορεί να υποδιαιρεθεί περαιτέρω μεταξύ υποσυνόλων του *A*. Επομένως, δεν περιλαμβάνει τμήματα της αξιοπιστίας που προσδίδονται στα υποσύνολα του *A*. Θα ήταν επομένως χρήσιμο να οριστεί μία ποσότητα που να υπολογίζει τον ολικό βαθμό αξιοπιστίας του συνόλου *A*. Η ποσότητα αυτή θα περιλαμβάνει όχι μόνον την αξιοπιστία αποκλειστικά του *A* αλλά και όλων των υποσυνόλων του. Η συνάρτηση αυτή ονομάζεται συνάρτηση αξιοπιστίας (belief function).

Αν $m(A)$ είναι ο βαθμός αξιοπιστίας που αντιστοιχεί στο υποσύνολο A και δεν διατίθεται κάποια αξιοπιστία για τα άλλα υπόλοιπα υποσύνολα του Ω, τότε $m(\Omega)= 1 - m(A)$. Η ποσότητα $m(\Omega)$ είναι ένα μέτρο της ολικής αξιοπιστίας που εναπομένει μετά την κατανομή της στα διάφορα υποσύνολα του Ω. Δηλαδή η υπολειπόμενη αξιοπιστία, $1 - m(A)$, δεν ανατίθεται στο συμπληρωματικό σύνολο του A, το A^c, όπως απαιτείται στον συμβατικό ορισμό της πιθανότητας κατά Bayes, αλλά σε όλο το Ω.

Σημειώνεται ότι οι συναρτήσεις m κατέχουν μη μηδενικό βαθμό αξιοπιστίας (δηλαδή, $m\neq0$), μόνον όταν υπάρχουν αποδεικτικά στοιχεία, ενώ αποκτούν μηδενικό βαθμό αξιοπιστίας, $m(A)=0$, όταν δεν υπάρχουν αποδεικτικά στοιχεία (evidence).

Παράδειγμα 1

Έστω ότι υπάρχουν αποδεικτικά στοιχεία που να επιβεβαιώνουν με βαθμό αξιοπιστίας 68% το γεγονός ότι κάποια από τις επιφανειακές καλύψεις είναι είτε A={δάσος} είτε B={καλλιέργεια}, αλλά δεν επιβεβαιώνεται η επιλογή μεταξύ A και B. Ο υπολειπόμενος βαθμός αξιοπιστίας, 1 - 68%= 32%, ανατίθεται στο Ω. Άρα $m(\{\text{δάσος}\}\vee\{\text{καλλιέργεια}\})$= 0,68 και $m(\{\text{δάσος, καλλιέργεια, νερά, αστικά}\}) = m(\Omega)$ = 0,32 και η τιμή του m για κάθε άλλο υποσύνολο του Ω είναι 0.

Παράδειγμα 2

Έστω ότι αποδεικτικά στοιχεία διαψεύδουν κατά 73% το γεγονός ότι η ψηφίδα θα πρέπει να καταχωρισθεί στην τάξη A={δάσος}. Τούτο ισοδυναμεί με επιβεβαίωση της καταχώρισης της ψηφίδας σε οποιαδήποτε τάξη πλην της A με βαθμό αξιοπιστίας 73%. Άρα $m(\{\text{νερά, καλλιέργεια, αστικά}\})$= 73% και = $m(\Omega)$ = 27% , ενώ η τιμή της συνάρτησης m για κάθε άλλο υποσύνολο του Ω είναι 0.

Παράδειγμα 3

Υποθέστε ότι δεν υπάρχουν αποδεικτικά στοιχεία που να προσδώσουν κάποια αξιοπιστία κατά την καταχώριση μιας ψηφίδας στις τάξεις {δάσος, καλλιέργεια, νερά, αστικά}. Η βασική πιθανότητα καταχώρισης αντιστοιχεί $m(\Omega)$=1 και 0 σε οποιοδήποτε υποσύνολο του Ω. Η συνάρτηση αξιοπιστίας, που προκύπτει σε αυτή την περίπτωση, ονομάζεται κενή συνάρτηση αξιοπιστίας (vacuous belief function).

Κατά τη θεωρία πιθανοτήτων του Bayes θα γινόταν προσπάθεια να αντιπροσωπευθεί αυτή η "άγνοια" με το 25% σε κάθε μονοσύνολο του Ω (π.χ. $P(\{\text{δάσος}\})$=25%), χωρίς να υπάρχει οιαδήποτε πληροφορία (αποδεικτικά στοιχεία) για το γεγονός αυτό. Άρα η ανάθεση του 25% σε κάθε μονοσύνολο του Ω θα σήμαινε επιπλέον πληροφορία που θα έπρεπε να είχαν παράσχει τα αποδεικτικά στοιχεία. Το οποίο όμως δεν είναι αληθές.

Η *συνάρτηση αξιοπιστίας* (belief function) ενός συνόλου A είναι το άθροισμα όλων των βαθμών αξιοπιστίας που προσδίδονται σε κάθε υποσύνολο A_i του A (δηλαδή $A_i \subset A$). Συμβολίζεται με Bel(A). Για το σύνολο $\{A, B, C\}$, για παράδειγμα, η συνάρτηση αξιοπιστίας είναι:

$$\begin{aligned}\text{Bel}(\{A,B,C\}) &= m(\{A,B,C\}) + \\ &+ m(\{A,B\}) + m(\{A,C\}) + m(\{B,C\}) + \\ &+ m(\{A\}) + m(\{B\}) + m(\{C\})\end{aligned} \qquad (2.2)$$

Άρα Bel(A) είναι ένα μέτρο του ολικού βαθμού αξιοπιστίας του A και όχι του βαθμού αξιοπιστίας, $m(A)$, που προσδίδουμε αποκλειστικά στο υποσύνολο A. Η συνάρτηση αξιοπιστίας για το ολικό σύνολο Ω, δηλαδή η Bel(Ω), ισούται, φυσικά, με τη μονάδα, επειδή Bel(Ω) είναι το άθροισμα των τιμών του $m(\Omega_i)$ για κάθε υποσύνολο Ω_i του Ω. Το άθροισμα αυτό θα πρέπει να ισούται με 1 (εξίσωση 1) από τον ορισμό της βασικής πιθανότητας καταχώρισης.

Επομένως, ο *βαθμός αξιοπιστίας* (support, belief measure), που δείχνουμε στην καταχώριση της ψηφίδας **x** στην τάξη A ή και σε οποιοδήποτε υποσύνολο A_i αυτής, είναι:

$$\text{Bel}(A) = \sum_{A_i \subset A} m(A_i) \qquad (2.3)$$

Αν το σύνολο A είναι το κενό σύνολο, τότε Bel(∅)= 0, ενώ για όλο το σύνολο Ω έχουμε Bel(Ω)= 1. Ορίζουμε εδώ τον βαθμό αξιοπιστίας αντί της πιθανότητας, επειδή οι βαθμοί αξιοπιστίας δεν ακολουθούν τις ιδιότητες της κλασικής θεωρίας των πιθανοτήτων.

Το πλεονέκτημα της χρήσης της θεωρίας των αποδείξεων σε σχέση με τη θεωρία των πιθανοτήτων είναι η ικανότητα να εκφράζει κανείς τον βαθμό της άγνοιας κατά την ταξινόμηση. Δηλαδή ο βαθμός εμπιστοσύνης, που δείχνουμε για την καταχώριση της ψηφίδας **x** στην τάξη A, δεν ισούται με την υπολειπόμενη της μονάδας εμπιστοσύνη που θα προσδίδαμε στο συμπληρωματικό σύνολο του A (όχι τάξη A). Ως εκ τούτου ισχύει η σχέση:

$$\text{Bel}(A) + \text{Bel}(A^c) = \text{Bel}(A) + \text{Bel}(\neg A) \le 1 \qquad (2.4)$$

Η ποσότητα Bel($\neg A$) εκφράζει το όριο μέχρι το οποίο τα αποδεικτικά στοιχεία στηρίζουν την αντίφαση του A (δηλαδή το $\neg A$). Επομένως, το 1–Bel($\neg A$) είναι το όριο μέχρι το οποίο τα αποδεικτικά στοιχεία δεν επιτρέπουν πλέον να αμφιβάλλει κανείς την καταχώριση της ψηφίδας **x** στη τάξη A. Η ποσότητα αυτή, 1– Bel($\neg A$), είναι ο *βαθμός ευλογοφάνειας* (plausibility) της προτίμησης στην καταχώριση της ψηφίδας **x** για την τάξη A και ισούται με:

$$\text{Pl}(A) = 1 - \text{Bel}(\neg A) \qquad (2.5)$$

Δηλαδή το εύλογο προκύπτει, επειδή η καταχώριση στην τάξη *A* δεν μπορεί να διαψευσθεί και επομένως είναι πιθανή και εύλογη. Αξιοπιστία της καταχώρισης θεωρείται ο *ελάχιστος* βαθμός εμπιστοσύνης (κάτω όριο της πιθανότητας) που προκύπτει από τα αποδεικτικά στοιχεία ως προς την προτίμηση της συγκεκριμένης καταχώρισης της ψηφίδας **x** στην τάξη *A*. Ενώ ευλογοφάνεια είναι ο *μέγιστος* βαθμός εμπιστοσύνης των αποδείξεων (άνω όριο της πιθανότητας) ως προς την προτίμηση της ταξινόμησης της ψηφίδας **x** στην τάξη *A*.

Η πιθανότητα κατά την οποία η ψηφίδα **x** δεν καταχωρίζεται ούτε στην τάξη *A* αλλά ούτε και στο συμπλήρωμά της A^C (δηλαδή, όχι *A*, ή $\neg A$) θεωρείται ότι είναι ο *βαθμός της άγνοιας*.

Το διάστημα μεταξύ του μέτρου της ευλογοφάνειας και αξιοπιστίας [Bel(*A*), Pl(*A*)] ονομάζεται *διάστημα αξιοπιστίας* (evidential interval, belief interval) της καταχώρισης ψηφίδας **x** στην τάξη *A*.

Παράδειγμα 4

Θεωρήστε ότι μια εικόνα αποτελείται από μία φασματική ζώνη καταγραφής. Κατά την ταξινόμησή της οι ψηφίδες πρέπει να καταχωρισθούν σε τρεις και μόνον τρεις τάξεις: *A*={δασικές εκτάσεις}, *B*={αγροτικές} και *D*={αστικές}. Θεωρήστε, επίσης, ότι είναι κατά κάποιο τρόπο γνωστό ότι οι βασικές πιθανότητες καταχώρισης των τάξεων είναι: *m*(*A*)= 33%, *m*(*B*)= 27% και *m*(*D*)=14%. Υποθέστε, όμως, ότι είμαστε αβέβαιοι για τη διαδικασία καταχώρισης των ψηφίδων στις τρεις παραπάνω τάξεις ή ακόμη και για την ποιότητα των δεδομένων έτσι, ώστε να ζητούμε να προσδώσουμε ένα επίπεδο αξιοπιστίας *m*(*A*∨*B*∨*D*)=87% κατά την ταξινόμηση μιας ψηφίδας σε οποιαδήποτε τάξη. Άρα η αβεβαιότητα καταχώρισης των ψηφίδων σε τάξεις είναι 1 - 87%=13%, παρ' όλο που είμαστε σίγουροι για τις προηγούμενες βασικές πιθανότητες καταχώρισης των τάξεων *m(A)*= 33%, *m(B)*= 27% και *m(D)*=14%.

Χρησιμοποιώντας τους συμβολισμούς της θεωρίας των αποδείξεων (Shafer, 1976), οι συναρτήσεις αξιοπιστίας, ευλογοφάνειας και τα διαστήματα αξιοπιστίας για καταχωρίσεις ψηφίδας **x** στην τάξη *A, B* και *D* είναι:

$$\begin{array}{lll} \mathrm{Bel}(A)=0{,}33 & \mathrm{Pl}(A)=1-0{,}27-0{,}14=0{,}59 & \mathrm{Pl}(A)-\mathrm{Bel}(A)=0{,}26 \\ \mathrm{Bel}(B)=0{,}27 & \mathrm{Pl}(B)=1-0{,}33-0{,}14=0{,}53 & \mathrm{Pl}(B)-\mathrm{Bel}(B)=0{,}26 \\ \mathrm{Bel}(D)=0{,}14 & \mathrm{Pl}(D)=1-0{,}33-0{,}27=0{,}40 & \mathrm{Pl}(D)-\mathrm{Bel}(D)=0{,}26 \end{array} \quad (2.6)$$

Παράδειγμα 5

Θεωρήστε ένα άλλο παράδειγμα, το οποίο αναφέρεται σε ταξινόμηση ψηφίδας σε τέσσερις τάξεις: (1) *A*(δασική), (2) *B*(αγροτική), (3) *D* (αστική) και (4) είτε στην *A*(δασική) είτε στην *B*(αγροτική). Δηλαδή στην τέταρτη τάξη έχουμε την πεποίθηση ότι η ψηφίδα ανήκει σε μία από τις δύο τάξεις, είτε στην *A*(αγροτική) είτε στην *B* (δασική), αλλά δεν γνωρίζουμε σε ποια από τις δύο. Υποθέστε ότι η αξιοπιστία κατά την καταχώριση στις τάξεις *A*={δασική}, *B*={αγροτική}, (*A* ∨ *B)*={δασική∨αγροτική} και *D*={αστική} δίδεται:

$$\begin{aligned} m(A) &= 0{,}33 \\ m(B) &= 0{,}27 \\ m(A \vee B) &= 0{,}05 \\ m(D) &= 0{,}14 \end{aligned} \quad (2.7)$$

Σύμφωνα με τα παραπάνω, ενώ είμαστε προετοιμασμένοι να αντιστοιχίσουμε ποσοστό αξιοπιστίας *m*(*A*)=33% στο γεγονός ότι η ψηφίδα ανήκει στην τάξη *A*={δασική} και ένα ποσοστό *m*(*B*)=27% στο γεγονός ότι ανήκει στην τάξη *B*={αγροτική}, είμαστε ταυτόχρονα προετοιμασμένοι να αντιστοιχίσουμε ένα επιπλέον ποσοστό *m(A∨B)*=5% στο γεγονός ότι μπορεί η ψηφίδα να ανήκει σε μία από τις δύο αυτές τάξεις *(A∨B)* - χωρίς να γνωρίζουμε ποια - και όχι στις υπόλοιπες.

Η αξιοπιστία της καταχώρισης στην τάξη *A*={δασική} είναι 33%. Ενώ η ευλογοφάνεια, σύμφωνα με την οποία η τάξη *A*={δασική} είναι η ορθώς επιλεγμένη τάξη για την ψηφίδα, ισούται με τη μονάδα μείον τη συνάρτηση αξιοπιστίας για τις αντιφατικές καταχωρίσεις (που είναι η αγροτική, η αστική και η μεικτή τάξη). Άρα, κατά την καταχώριση ψηφίδας στην τάξη *A*(αγροτική) ισχύει:

$$\begin{aligned} &\mathrm{Bel}(A) = m(A) = 33\% \\ &\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\neg A) = 1 - 27\% - 14\% - 5\% = 54\% \\ &\mathrm{Pl}(A) - \mathrm{Bel}(A) = (54 - 33)\% = 21\% \end{aligned} \quad (2.8)$$

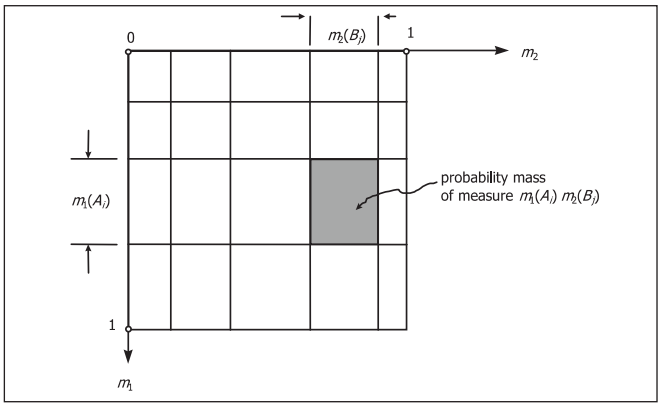
Η συνάρτηση αξιοπιστίας της καταχώρισης ψηφίδας στη μεικτή τάξη *(A∨B)* είναι:

$$\begin{aligned} &\mathrm{Bel}(A \vee B) = m(A) + m(B) + m(A \vee B) \\ &= 33\% + 27\% + 5\% = 65\% \end{aligned} \quad (2.9)$$

Δηλαδή η συνάρτηση αξιοπιστίας είναι το άθροισμα των ποσοστών των αποδείξεων που αντιστοιχούν στην τάξη αυτή και στα υποσύνολά της.

3. ΣΥΝΔΥΑΣΜΟΣ ΑΠΟΔΕΙΞΕΩΝ

Η θεωρία των αποδείξεων μάς επιτρέπει να συνδυάσουμε τα αρχικά ποσοστά αξιοπιστίας που προέρχονται από διαφορετικές πηγές αποδείξεων (π.χ. δεδομένων) και να προσδιορίσουμε την τάξη εκείνη που είναι από κοινού αποδεκτή για την καταχώριση ψηφίδας. Η διαδικασία αυτή πραγματοποιείται με τον κανόνα του ορθογώνιου αθροίσματος (orthogonal sum) του Dempster (Yager et al., 1994). Ο κανόνας αυτός αντιστοιχεί στη συνένωση των αποδεικτικών στοιχείων που προέρχονται από διαφορετικές πηγές. Για να ισχύει ο κανόνας, τα στοιχεία πρέπει να είναι μεταξύ τους στατιστικά ασυσχέτιστα.



Σχήμα 3: Υπολογισμοί του βαθμού αξιοπιστίας των ταξινομήσεων με τον κανόνα του Dempster.

Figure 3: Geometric representation of Dempster's rule of combination.

Έστω Bel_1 και Bel_2, και m_1 και m_2 δύο συναρτήσεις αξιοπιστίας και οι αντίστοιχες βασικές πιθανότητες καταχώρισης σε ένα πλαίσιο διακρίσεων Ω. Ο κανόνας του Dempster υπολογίζει μία νέα βασική πιθανότητα καταχώρισης, η οποία συμβολίζεται με $m_1 \oplus m_2$, και αντιπροσωπεύει τον συνδυασμό των επιδράσεων του m_1 και m_2. Η αντίστοιχη συνάρτηση αξιοπιστίας $\text{Bel}_1 \oplus \text{Bel}_2$ υπολογίζεται απλά από τα $m_1 \oplus m_2$ και σύμφωνα με τον ορισμό της συνάρτησης αξιοπιστίας.

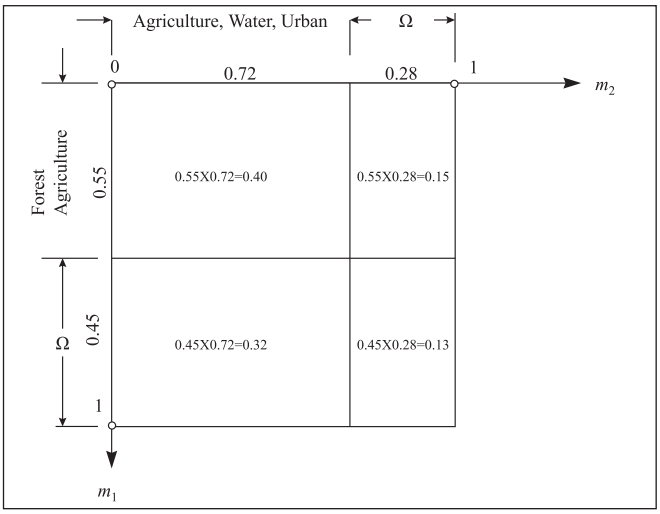
Ο κανόνας είναι πιο κατανοητός, όταν παρασταθεί γραφικώς. Αν θεωρήσουμε ότι τα σύνολα A_1, …, A_i, …, A_k, που μετρούνται από τη βασική πιθανότητα καταχώρισης $m_1(A_1)$, …, $m_1(A_k)$ βασισμένη στην πρώτη πηγή αποδεικτικών στοιχείων, τότε οι αριθμοί αυτοί αντιστοιχούν σε τμήμα μιας γραμμής, το μήκος της οποίας εκτείνεται από 0 μέχρι 1 (βλέπε σχήμα 3).

Για τη δεύτερη πηγή αποδεικτικών στοιχείων οι αριθμοί της βασικής πιθανότητας καταχώρισης $m_2(B_1)$, …, $m_2(B_j)$,…, $m_2(B_l)$ μπορούν, επίσης, να αντιστοιχούν σε τμήματα μιας άλλης γραμμής που συνδυαζόμενη με την προηγούμενη σχηματίζει ένα τετράγωνο. Θεωρήστε ότι το τετράγωνο αυτό αντιπροσωπεύει την ολική μάζα της πιθανότητας από τις δύο πηγές αποδεικτικών στοιχείων. Η τομή των δύο λωρίδων έχει μέτρο $m_1(A_i)m_2(B_j)$ και είναι η πιθανότητα που καταχωρείται στην τομή του $A_i \cap B_j$.

Αν αθροίσουμε όλα τα γινόμενα $m_1(A_i)m_2(B_j)$, όπου A_i και B_j είναι όλα τα υποσύνολα του πλαισίου Ω, το αποτέλεσμα πρέπει να είναι 1, σύμφωνα με τον ορισμό της βασικής πιθανότητας καταχώρισης:

$$\sum m_1(A_i)m_2(B_j) = \sum m_1(A_i) \sum m_2(B_j) = 1 \times 1 = 1 \qquad (3.1)$$

Η βασική πιθανότητα καταχώρισης, που αντιστοιχεί στον συνδυασμό m_1 και m_2, κατανέμει τον αριθμό 1, την ολική ποσότητα αξιοπιστίας, μεταξύ των υποσυνόλων του Ω, αναθέτοντας $m_1(A_i)m_2(B_j)$ στην τομή του A_i και B_j. Άρα, για κάθε υποσύνολο C του Ω, ο κανόνας Dempster ορίζει το $m_1 \oplus m_2(C)$ ως το άθροισμα των γινομένων του τύπου $m_1(A_i)m_2(B_j)$ για όλα τα υποσύνολα, των οποίων η τομή είναι $C = A_i \cap B_j$.



Σχήμα 4: Υπολογισμοί του βαθμού αξιοπιστίας με τον κανόνα του Dempster.

Figure 4: Computations of belief based on Dempster's rule.

Παράδειγμα 6

Έστω ότι αποδεικτικά στοιχεία στηρίζουν κατά 60% (=m_1) την επιλογή της τάξης G={δάσος, καλλιέργεια}, ενώ κάποια άλλα στοιχεία επιβεβαιώνουν κατά 70% (=m_2) την επιλογή H={καλλιέργεια, νερά, αστικά}. Ο τελικός βαθμός αξιοπιστίας, που καταλογίζεται από τις μετρήσεις αυτές, ορίζεται στο $m_1 \oplus m_2(C)$. Για υπολογιστικούς λόγους θα πρέπει να κατασκευαστεί ένας πίνακας της τομής των συνόλων με τις αντίστοιχες τιμές m_1 και m_2 στις γραμμές και στις στήλες. Μόνο οι μη μηδενικές τιμές του m_1 και m_2 εξετάζονται αφού αν $m_1(G)$ και/ή $m_2(H)$ είναι 0, τότε το γινόμενο $m_1(G)m_2(H)$ δίνει 0 στο $m_1 \oplus m_2(C)$, όπου C είναι η τομή των συνόλων G και H. Η τιμή του $m_1 \oplus m_2(C) = m_1 \oplus m_2(\{\text{δάσος, καλλιέργεια}\})$ υπολογίζεται αθροίζοντας όλα τα γινόμενα (σχήμα 4).

Στο σχήμα 4 ένα υποσύνολο εμφανίζεται μία μόνο φορά και επομένως το $m_1 \oplus m_2$ υπολογίζεται εύκολα:

$$m_1 \oplus m_2\{\text{καλλιέργεια}\} = 40\%$$
$$m_1 \oplus m_2\{\text{δάσος, καλλιέργεια}\} = 15\%$$
$$m_1 \oplus m_2\{\text{καλλιέργεια, νερά, αστικά}\} = 32\%$$
$$m_1 \oplus m_2\{\Omega\} = 13\%$$

Σημειώστε ότι η βασική πιθανότητα καταχώρισης του συνδυασμού $m_1 \oplus m_2$ ικανοποιεί τον ορισμό του, δηλαδή:

$$\sum_C m_1 \oplus m_2(C) = 1, \quad \text{με} \quad C = (A_i \cap B_j) \subset \Omega \quad \text{και} \quad \sum m_1 \oplus m_2(\varnothing) = 0$$

Επειδή το ορθογώνιο άθροισμα $\mathrm{Bel}_1 \oplus \mathrm{Bel}_2$ είναι αρκετά πολύπλοκο, δίδεται μόνο ένα παράδειγμα:

$$\mathrm{Bel}_1 \oplus \mathrm{Bel}_2\{\text{δάσος, καλλιέργεια}\} =$$
$$\mathrm{m}_1 \oplus \mathrm{m}_2\{\text{δάσος, καλλιέργεια}\} +$$
$$\mathrm{m}_1 \oplus \mathrm{m}_2\{\text{δάσος}\} + \mathrm{m}_1 \oplus \mathrm{m}_2\{\text{καλλιέργεια}\} =$$
$$= 0{,}15 + 0 + 0{,}40 = 0{,}55 = 55\%$$

Κατά συνέπεια, αν υπάρχουν δύο βασικές πιθανότητες καταχώρισης, που δημιουργούνται από δύο ανεξάρτητες ενότητες ψηφίδων ή άλλων δεδομένων, τότε η συνδυασμένη (η από κοινού) βασική πιθανότητα καταχώρισης $m_1 \oplus m_2$ για την τάξη C υπολογίζεται από τον κανόνα του Dempster (Yager et al., 1994):

$$m_1 \oplus m_2(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1-\kappa} = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j)} \qquad (3.2)$$

όπου ο τελεστής $\oplus$ ονομάζεται *ορθογώνιο άθροισμα*. Παρατηρήστε ότι στην παραπάνω σχέση έχουμε διαιρέσει με έναν παρονομαστή (1–κ) που αντιστοιχεί στο να εξαναγκαστεί το άθροισμα $m_1 \oplus m_2(\varnothing)$ να πάρει την τιμή 0 (λεπτομέρειες στον Yager et al., 1994). Χρησιμοποιώντας το ορθογώνιο άθροισμα μπορούμε να προσδιορίσουμε την αξιοπιστία και την ευλογοφάνεια με τη θεωρία των αποδείξεων. Για να εφαρμοστεί η θεωρία των αποδείξεων στην ταξινόμηση της εικόνας, θα πρέπει να ακολουθηθούν τα εξής βήματα:

1. Καθορίζεται η συνάρτηση πυκνότητας $f(\mathbf{x})$ για κάθε πηγή **X**. Κάθε φασματική ζώνη (κανάλι) της εικόνας θεωρείται ότι αποτελεί την ανεξάρτητη και ξεχωριστή πηγή αποδείξεων. Σε αυτή την περίπτωση, το φασματικό κανάλι q των 8-bit της εικόνας έχει ιστόγραμμα που δίνεται από τη συνάρτηση πυκνότητας: $f_q(x) \quad x \in \{0, 1, 2, 3, \ldots, 255\}$

2. Προσδιορίζεται η συνάρτηση του ποσοστού των αποδείξεων (αξιοπιστία) για την καταχώριση εικονοστοιχείου στην τάξη $\Omega_j \in \{A, B, C, \ldots\}$ στο κανάλι q: $m_q(\Omega_1)$, $m_q(\Omega_2)$, $m_q(\Omega_3)$,…

3. Συνδυάζονται οι βασικές πιθανότητες καταχώρισης για κάθε κανάλι q της εικόνας χρησιμοποιώντας την σχέση του Dempster (3.2). Η παραπάνω σχέση χρησιμοποιείται διαδοχικά προσθέτοντας ένα κανάλι της εικόνας σε κάποιο άλλο κάθε φορά, μέχρις ότου συνδυαστούν όλες οι βασικές πιθανότητες καταχώρισης από όλα τα κανάλια της εικόνας.

4. Προσδιορίζεται το τελικό διάστημα αξιοπιστίας της καταχώρισης εικονοστοιχείου στην κάθε τάξη $\Omega_j = \{\{A\}, \{B\}, \{C\}, \ldots\}$.

Παράδειγμα 7

Το παράδειγμα που ακολουθεί χρησιμοποιείται για να δείξει τον τρόπο με τον οποίο υπολογίζονται τα βήματα 3 και 4 παραπάνω. Θεωρήστε μια εικόνα που διαθέτει δύο κανάλια καταγραφής (q =1, 2), τα οποία θεωρούνται ότι είναι ανεξάρτητα το ένα από το άλλο. Στην πραγματικότητα, η συνθήκη αυτή της στατιστικής ανεξαρτησίας δεν ισχύει σχεδόν πουθενά στις ψηφιακές εικόνες. Εξετάζεται όμως παρακάτω. Οι βασικές πιθανότητες καταχώρισης m (συναρτήσεις πιθανότητας) προσδιορίστηκαν από τη στατιστική ανάλυση των εικονοστοιχείων της εικόνας ως:

Πίνακας 1: Οι πιθανότητες καταχώρισης τάξεων σε μία εικόνα.
Table 1: The classification probability in an image.

	Πιθανότητα εμφάνισης τάξεων στην εικόνα			
	Δασική (Forest)	Αγροτική (Agriculture)	Νερά (Water)	Αστική (Urban)
1	0,2	0,3	0,3	0,1
2	0,1	0,3	0,4	0,0

Από τον πίνακα 1 προκύπτει ότι τα ποσοστά των αποδείξεων m της καταχώρισης είναι ενδεικτικά:

$$\begin{aligned} &m_1(A) = 0{,}2 \\ &m_1(B) = 0{,}3 \\ &\vdots \\ &m_2(C) = 0{,}4 \\ &m_2(D) = 0{,}0 \end{aligned} \qquad (3.3)$$

όπου οι δείκτες αναφέρονται στα φασματικά κανάλια της εικόνας. Για τον υπολογισμό του ορθογώνιου αθροίσματος $m_1 \oplus m_2(\Omega_i)$, όπου Ω_i={δασική, αγροτική, νερά, αστική}={F, A, W, U}, χρησιμοποιείται ο πίνακας 2.

Αν υπήρχε και τρίτο φασματικό κανάλι ("αποδεικτική" πηγή δεδομένων), τότε θεωρούμε το παραπάνω ορθογώνιο άθροισμα $m_1 \oplus m_2(\Omega_i)$, όπου Ω_i={δασική, αγροτική, νερά, αστική}={F, A, W, U} ως ένα νέο m_1 ή m_2 και εφαρμόζουμε το m_3 ακριβώς με τον ίδιο τρόπο όπως και έγινε παραπάνω. Το ορθογώνιο άθροισμα μπορεί, επίσης, να εκφραστεί αλγεβρικώς, ώστε να μπορεί να ενσωματωθεί σε προγράμματα λογισμικού για την εφαρμογή των μεθόδων της θεωρίας των αποδείξεων και ανάλυσης πολλαπλών πηγών δεδομένων (Tessem, 1993).

Αν διατίθενται περισσότερες από δύο πηγές δεδομένων εικόνας, τότε η σειρά εφαρμογής του ορθογωνίου αθροίσματος δεν επηρεάζει το τελικό αποτέλεσμα. Δηλαδή για περισσότερες από δύο πηγές δεδομένων το ορθογώνιο άθροισμα (εξίσωση 11) μπορεί να εφαρμοστεί διαδοχικά, επειδή η σχέση αυτή είναι και *προσεταιριστική* (associative) (μπορεί να εφαρμοστεί για οποιοδήποτε ζευγάρι πηγών και κατόπιν σε μια τρίτη πηγή ή ισοδύναμα να εφαρμοστεί σε διαφορετικό ζευγάρι και σε μια τρίτη πηγή), καθώς και *αντιμεταθετική* (commutative) (η διάταξη, κατά την οποία διερευνώνται (εξετάζονται) οι πηγές των δεδομένων, δεν αλλάζει το αποτέλεσμα).

Πίνακας 2: Υπολογισμοί του βαθμού αξιοπιστίας και ευλογοφάνειας των ταξινομήσεων των στοιχείων μιας εικόνας με δύο κανάλια με τη μέθοδο των αποδείξεων.
Table 2: Computations of belief and plausibility.

	Κανάλι 1				
Κανάλι 2	**F (=0,2)**	**A (=0,3)**	**W (=0,3)**	**U (=0,1)**	**Ω(=0,1)**
F (=0,1)	F (=0,02)	∅ (=0,03)	∅ (=0,03)	∅ (=0,01)	F (=0,01)
A (=0,3)	∅ (=0,06)	A (=0,09)	∅ (=0,09)	∅ (=0,03)	A (=0,03)
W (=0,4)	∅ (=0,08)	∅ (=0,12)	W (=0,12)	∅ (=0,04)	W (=0,04)
U (=0,0)	∅ (=0,00)	∅ (=0,00)	∅ (= 0,00)	U(= 0,00)	U (=0,00)
Ù(=0,2)	F (=0,04)	A (=0,06)	W (=0,06)	U (=0,02)	Ω (=0,02)
	$\sum_{G\cap H=A} m_1(G)\, m_2(H$ =0,07	$\sum_{G\cap H=B} m_1(G)\, m_2(H)$ =0,18	$\sum_{G\cap H=C} m_1(G)\, m_2(H)$ =0,22	$\sum_{G\cap H=D} m_1(G)\, m_2(H)$ =0,02	$\sum_{G\cap H=\Omega} m_1(G)\, m_2(H)$ =0,02
	$1-\sum_{G\cap H=\varnothing} m_1(G)\, m_2(H)$ =1–0,49=0,51				
$m_1 \oplus m_2$	0,07/0,51= =0,14	0,18/0,51= =0,35	0,22/0,51= =0,43	0,02/0,51= =0,04	0,02/0,51= =0,04
$Bel_{m_1\oplus m_2}$	0,14	0,35	0,43	0,04	1
$Pl_{m_1\oplus m_2}$	0,18	0,39	0,47	0,08	1

Η χρήση της θεωρίας των αποδείξεων απαιτεί οι πηγές των δεδομένων να είναι μεταξύ τους στατιστικά ανεξάρτητες. Η συνθήκη αυτή ικανοποιείται με το να χρησιμοποιήσουμε τεχνικές *αποσυσχέτισης* (decorrelation) των δεδομένων, όπως είναι η ανάλυση κυρίων συνιστωσών (Richards, 1993).

4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Αφού εφαρμοστεί το ορθογώνιο άθροισμα, ο αναλυτής μπορεί να υπολογίσει τον βαθμό αξιοπιστίας (beleif) και ευλογοφάνειας (plausibility) για κάθε ψηφίδα σε κάθε τάξη εδαφικής κάλυψης. Για την τελική καταχώριση των εικονοστοιχείων της εικόνας σε τάξεις της εδαφικής κάλυψης η απόφαση μπορεί να ληφθεί, αφού χρησιμοποιηθεί το κριτήριο της μέγιστης αξιοπιστίας (Lee, Richards and Swain, 1987) ή της μέγιστης ευλογοφάνειας (Srinivasan and Richards, 1990) ή και τα δύο (Peddle, 1995). Η απόφαση μπορεί να θεωρηθεί ότι εμπεριέχει κάποιο βαθμό διακινδύνευσης, πάντα.

Το πλεονέκτημα της θεωρίας των αποδείξεων σε σχέση με τη συμβατική ταξινόμηση της εικόνας είναι ότι επιτρέπει τη χρήση δεδομένων από διαφορετικές πηγές και διαφορετικού τύπου και ακρίβειας με μεγάλες διαστάσεις, ενώ ταυτόχρονα αποφεύγονται αυστηροί ορισμοί πιθανοτήτων. Για παράδειγμα, σε εφαρμογές όπου όχι μόνο απαιτούνται πολυφασματικές εικόνες αλλά και θεματικοί συμβατικοί χάρτες, ή προσωπικές εκτιμήσεις από έναν εδαφολογικό χάρτη, τότε η θεωρία των αποδείξεων μπορεί να χρησιμοποιηθεί κάλλιστα, με την προϋπόθεση ότι μπορούν να υπολογιστούν οι συναρτήσεις κατανομών των τάξεων της εδαφικής κάλυψης.

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ ιδιαίτερα τους δύο κριτές για τα εύστοχα σχόλια και τις παρατηρήσεις τους. Ο αγγλικός όρος pixel (picture element) αποδόθηκε στα ελληνικά ως "εικονοστοιχείο" αλλά και ως "ψηφίδα", σύμφωνα με τις υποδείξεις των κριτών. Οι ελληνικοί αυτοί όροι χρησιμοποιούνται εναλλάξ, συμβαδίζοντας με τις απόψεις και των δύο κριτών.

ΒΙΒΛΙΟΓΡΑΦΙΑ ΚΑΙ ΑΝΑΦΟΡΕΣ

1. Anderson, T. W. (1984). **An introduction to Multivariate Statistical Analysis,** John Wiley & Sons Inc., New York.

2. Awcock, G. L. and R. Thomas (1995). **Applied Image Analysis,** MacMillan New Electronics, Introduction to Advanced Topics, London.

3. Ball, G. H. and D. J. Hall (1965). **A Novel Method of Data Analysis and Pattern Classification,** Stanford Research Institute, Menlo Park, California.

4. Brooks, S. R. (editor, 1989). **Mathematics in Remote Sensing,** Clarendon Press, Oxford.

5. Campbell, B. J. (1987). **Introduction to Remote Sensing,** Guilford Press, New York.

6. Curran, P. J. (1985). **Principles of Remote Sensing,** Longman, New York.

7. Congalton, R. G. and K. Green (1998). **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices,** Lewis Publishers, Boca Raton, Florida, USA.

8. Castelman, K. R. (1996). **Digital Image Processing,** Prentice Hall, Englewood Cliffs, New Jersey.

9. Duda, R. O. and P. E. Hart (1973). **Pattern classification and scene analysis,** John Wiley and Sons, New York.

10. Jensen, R. J. (1995). **Introductory Digital Image Processing: A Remote Sensing Perspective,** Prentice Hall in Geographic Information Science, New Jersey, Second Edition.

11. Lee, T., J. A. Richrads and P. H. Swain (1987). Probabilistic and evidential approaches for multi-source data analysis, **IEEE Transactions on Geoscience and Remote Sensing,** Vol. **25,** No 3, pp. 283–292.

12. Pao, Y. H. (1989). **Adaptive Pattern Recognition and Neural Networks,** Addison-Wesley, Reading.

13. Peddle, D. R. (1995). Mercury⊕: An evidential Reasoning Image Classifier, **Computers & Geosciences,** Vol. **21,** No 10, pp. 1163–1176.

14. Richards, J. A. (1993). **Remote Sensing Digital Image Analysis, An Introduction,** Springer-Verlag, Berlin, Second Edition.

15. Srinivasan, A., and J. A. Richards (1990). Knowledge-based techniques for multi-source classification, **International Journal of Remote Sensing,** Vol. **11**, No. **3**, pp. 505–525.

16. Schowengerdt, R. A.(1983). **Techniques for Image Processing and Classification,** Academic Press, San Diego, California.

17. Schowengerdt, R. A. (1997). **Remote Sensing: Models and Methods for Image Processing,** Academic Press, San Diego, California.

18. Shafer, G. (1976). **Mathematical Theory of Evidence,** Princeton University Press, New Jersey.

19. Shafer, G. and J. Pearl (1990). **Readings in Uncertainty Reasoning,** Morgan Kaufmann Series in Representation and Reasoning, Morgan Kaufmann Publishers, Inc., San Mateo, California.

20. Shafer, G. (1996). **Probabilistic Expert Systems,** Cbms-NSF Regional Conference Series in Applied Mathematics, No 67, Society for Industrial & Applied Mathematics (SIAM).

21. Tessem, B. (1993). "Approximations for efficient computation in the theory of evidence", **Artificial Intelligence,** Vol. **61**, pp. 315–329.

22. Yager, R. R. J. Kacprzyk and M. Fedrizzi (1994). **Advances in the Dempster-Shafer Theory of Evidence,** John Wiley and Sons, New York.

Στέλιος Π. Μερτίκας,
Διπλ. ΑΤΜ (ΕΜΠ), M. Sc. E. (UNB), Ph. D., καθηγητής Πολυτεχνείου Κρήτης, Τμήμα Μηχανικών Ορυκτών Πόρων, Τομέας Ανίχνευσης & Εντοπισμού, Εργαστήριο Γεωδαισίας & Πληροφορικής των Γεωεπιστημών, Πολυτεχνειούπολη, 731 00 Χανιά, Κρήτη.

Extended summary

The Theory of Evidence in Remote Sensing Image Classification

STELIOS P. MERTIKAS
Professor, Technical University of Crete

Abstract
This article introduces the mathematical Theory of Evidence in classifying Remote Sensing images. Its main intent is to introduce the less familiar concepts of belief functions in image classification. The belief function can be considered as a generalisation of the classical Bayes probability function that includes, however, a way to assess the strength of evidence. To illustrate the Theory of Evidence seven examples are given.

1. INTRODUCTION

Classification is, in general, an area of multivariate statistics (Duda and Hart, 1973; Anderson, 1984; Awcock and Thomas, 1995; Castleman, 1996) that deals with grouping a set of items by assigning them to several similar classes. The first goal of a classification procedure is to describe the degree of similarity or disjointness of the items under consideration. The second goal is to select a proper rule (algorithm) so that the items are separated into groups of similar properties.

Classification of a digital image (Jensen, 1995; Richards, 1993; Congalton and Green, 1998; Schowengerdt, 1983; 1997) is a procedure of converting image pixels with similar properties, structure etc., into similar descriptive labels that correspond to different surface materials or conditions. In the evaluation study of water quality of a remote-sensing image, for example, the first step of image classification would include identification of all pixels that correspond to water. Then further steps of classification would include a detailed description of the water pixels and a pixel labeling into other subgroups designating water quality, depth, amount of suspended particles, etc.

Most traditional methods of image classification, such as minimum distance, maximum likelihood, maximum *a posteriori*, etc. (Curran, 1985; Campbell, 1987; Richards, 1993), require that the data are digital and assume parametric statistical models, such as the Gaussian distribution. These methods are not designed to handle data from different sources or of varying accuracy and they cannot cope with non-numerical data. In practice, the data usually do not obey the conditions imposed by traditional methods that classify pixels via crisp rules. The mathematical Theory of Evidence by Dempster and Shafer (Shafer, 1976; 1996; Shafer and Pearl, 1990; Yager et al., 1994) is a scientific field that can deal with statistically independent sources, which may not give digital data (e.g., photographs, geological maps, engineer's judgement, etc.). It does not require the specification of a complete probabilistic model that relates the set of class hypotheses and can handle data of high dimension. In the following, we briefly discuss the problems encountered in remote sensing image analysis and review methods available for classification.

Submitted: Feb. 17, 1999 Accepted: July 8, 1999

The Problems

In remote sensing applications, there often exist multiple data that can be used, including multispectral data from different sources, thematic maps, spatial databases, elevation data, etc. These data may be in point or area specific form and can be numerical (pixel values) or nominal (maps of labels) in nature. With the advent of sensor and computer technology, there exists an increasing demand for combining information from heterogeneous sources in order to exploit knowledge as much as possible and derive accurate analysis systems for remote sensing data (Lee et al., 1987). The combined analysis of data sources referring to different spatial representations (pixel, line, or area) becomes difficult, requiring human expert interaction for some form of registration of the various spatial systems. In this paper we are concerned with information sources defined on the same coordinate system deriving data in pixel-specific form. Therefore, the major issue under consideration becomes the joint analysis and classification of pixels from heterogeneous sources of information, with particular emphasis on handling uncertainty in the nature of data (numerical or nominal) and in the accuracy of data (noise, blurring, etc.).

Multichannel data encode information from different frequency bands at each spatial location (pixel). These data can be presented in a pixel-specific vector form and analyzed jointly via multidimensional statistical methods. Such methods rely upon multivariate distribution functions that can be inferred from the data. For non-numerical data, however, such a consideration through classical approaches is not suitable.

Initial approaches to the study of heterogeneous pieces of information considered sources independently and focused on the combination of results. In Strahler and Bryant (1978), the elevation data are used to define elevation regions of interest and, subsequently, multispectral data are classified within each elevation range separately. Hutchinson (1982) uses information from other sources only to resolve ambiguities created in the classification of spectral data.

These approaches do not exploit to the maximum degree the amount of correlation existent in the multi-sensor data. To handle such information accurately, we need rigorous approaches for the joint consideration of evidence from multiple sources, for each single pixel. Towards this direction, the notion of the global membership function as an extension to the joint probability function, its relation to evidential calculus and their applications in remote sensing are presented in Lee et al. (1987). We continue with an overview of the theory of evidence and examine its application in the classification of remote sensing data.

Classification Approaches

The information employed in classification regards aspects of pixel or region similarity. Deterministic techniques rely on spatial similarity expressed by distances of measurement vectors. The classical Bayesian theory deals with the stochastic nature of data/measurements based on specific models of distribution and on the Bayesian logic for hypothesis testing. Modern approaches alleviate assumptions regarding precise distribution models, which are only approximations of reality. Neural approaches cope with unknown nonlinear models underlying the distribution of the data, but they still implement the Bayesian logic in classification. The Theory of Evidence expresses the degree of ignorance in the classification problem, based on approximate reasoning. It considers crisp sets, but introduces uncertainty in the decision making process (logic) by considering probability intervals with lower and upper bounds rather than explicit probability values. The Fuzzy Set theory expresses uncertainty in the definition of the sets or classes and, thus, it necessarily handles uncertainty in its reasoning logic.

The Theory of Evidence provides the means of dealing with data uncertainty in complex mixed data sets (numerical or nominal information) with reasonable cost, thus providing the ability to analyze data sets of high dimension. The Theory of Evidence has recently been used in artificial intelligence to provide a way of deriving subjective judgment in a more natural way than the probability theory (Shafer, 1996). In essence, the probability assignment describes the frequency of observation of a pattern in a set, whereas the belief function in the theory of evidence reflects the degree to which a pattern resembles the characteristics of this set.

In this paper we present the fundamental issues of the theory of evidence and trace its application in the classification of remote sensing images. In Section 2 we discuss the concept of the belief function as opposed to the probability function in the context of statistical hypothesis testing. The advantages of the belief function over the probability function are elucidated through several examples. Section 3 provides the basis for combining evidence from different sources through Dempster's rule of combination. The decision-making logic according to the theory of evidence is presented and discussed through classification examples. Section 4 concludes the paper by summarizing the potential and the advantages of the theory of evidence in remote sensing data analysis.

2. BELIEF FUNCTIONS IN STATISTICAL HYPOTHESES

Classification could be considered as a mapping from the space of data X onto the space of (feature) classes Ω (fig. 1). In Remote Sensing the space of data X is commonly expressed by the multispectral image which is the set of all vector-valued image pixels. A feature vector h(x) is assigned to each element x of X. The space Ω spans the entire feature space and is composed of a set of mutually exclusive and exhaustive (i.e. cover the entire space) classes of interest. In our consideration, the elements of Ω represent the land cover material (i.e., soil, vegetation, surface water etc.). The classification process assigns (labels) each image pixel to one class, by testing the corresponding hypotheses about these classes based on the evidence provided by the data. In most realistic cases, however, this evidence is not enough to distinguish among the different classes, implying uncertainty in the hypothesis testing and imposing the need for testing combinations of hypotheses.

The set of all possible subsets of the frame of discernment **Ω** is denoted by 2^{Ω} and defines the complete set of hypotheses about the classification problem. Here we use the term "hypothesis" in that generalized context to declare any subset of the initial hypotheses in Ω.

Consider, for example, a classification scheme of four classes, i.e., A={Forest}, B={Agriculture}, C={Water},

D={Urban}, as in fig. 2. The finite set of all possible classes is denoted by Ω ={Forest, Agriculture, Water, Urban} and is commonly called the *frame of discernment*. Figure 2 presents the frame of discernment Ω and its subsets. For a frame of *n* elements, there exist 2^n subsets Ω_i. The empty set, ∅, is also one of the subsets, but corresponds to a hypothesis that is false and is not shown in figure 2.

The theory of evidence attempts to quantify the natural process of subjective judgement. It is important, therefore, to characterize the positive or negative influence that evidence creates towards one hypothesis. Some evidence (e.g., data) may support a certain hypothesis and convince the analyst to place some degree of belief in the choice of , say, a two-element subset of Ω, such as the class {*A*, *B*}= {Forest, Agriculture}. Alternatively, some new evidence may lead to exclude the class *A*={Forest}, to some degree, from further consideration.

Evidence that does not support class *A* is equivalent to the negation of the hypothesis for the class *A* ={Forest}. In this case, the support leans on the choice of {NOT *A* }, which obviously corresponds to the choice of the hypothesis {*B* OR *C* OR D}. Evidence that does not approve class *A,* convince the analyst to adopt a degree of belief in the subset of the remaining classes {*B* OR *C* OR *D*}.

The mathematical Theory of Evidence does not assume hard definitions of probability because the stochastic distributions of classes and the precise probability assignments in Ω are unknown. It rather assigns lower and upper bounds to the distributions and, in addition, expresses the degree of ignorance in the classification. The Theory of Evidence uses a number in the interval [0, 1] to indicate the strength (or degree) of belief assigned to a hypothesis given a piece of evidence. This number is the degree to which the evidence supports the specific hypothesis. Evidence against a hypothesis *A* is regarded as evidence for the negation (NOT *A*) of the hypothesis.

3. DISCUSSION AND CONCLUDING REMARKS

Instead of producing a thematic map of classification showing the various land cover types, the Theory of Evidence allows the construction of a map showing each class and at the same time depicting the distribution of belief in classification (or plausibility). This is always useful in practice, where there are doubts in assigning a certain pixel to a class.

The main advantage of the Theory of Evidence versus conventional image classification approaches is that it allows the use of data, or pieces of evidence, from different heterogeneous sources, in diverse forms (numerical and/or nominal). Compared with the classical Bayesian theory, the theory of evidence is distinguished by its ability to express the degree of ignorance in a classification process. Moreover, the theory of evidence handles problems under uncertainty and provides a means of deriving, as well as exploiting subjective judgement. As opposed to the probability theory, it does not assume specific definitions of probability. Thus, the theory of evidence alleviates restrictive assumptions regarding specific distribution models and precise probability assignments, which are unknown in practice. Therefore, the theory of evidence renders the reasoning process robust to model uncertainties, as well as noise disturbances. As opposed to the fuzzy set theory, the theory of evidence considers crisp sets, but incorporates approximate reasoning in its logic to express uncertainty and possibility in the classification of a measurement to a set. Thus, the theory of evidence deals with uncertainty in the classification process itself rather than in the definition of classes. It provides a powerful means of dealing with data uncertainty in complex mixed data sets at a reasonable cost, thus providing the ability to analyze data sets of high dimension.

In summary, the Theory of Evidence does not examine the probability of belonging in terms of how often (in the universe of discourse of images) the pixel **x** truly belongs to the class *A*, but the possibility of belonging in terms of how similar the characteristics of the pixel **x** are to those of the class *A*. It provides a natural way of examining similarity and uses approximate reasoning to model subjective judgement in grouping pixels together. The major characteristics of the theory of evidence that make it useful in the classification of remote sensing data are the following:

- Enables subjective judgement that relates to the natural way of operation of the human expert.
- Can deal with data from different sources and can handle different forms of data (numerical or nominal) with different natural properties.
- Does not require fitting of specific distribution models to the data and does not force the data to obey strict assumptions.
- Does not require precise specification of probability assignments for each measurement value.
- Provides robustness in the classification process and can handle noise corrupt or blurred images.

Stelios P. Mertikas,
Dipl. Eng., (NTUA), M. Sc.E. (UNB), Ph. D. (UNB), Professor, Technical University of Crete, Mineral Resources Engineeering Department, Division of Exploration & Positioning, Laboratory of Geodesy & Geomatics Engineering, Chania, Crete, GR-731 00.