

Θ.Ε. ΠΛΗ22 (2020-21) – ΓΡΑΠΤΗ ΕΡΓΑΣΙΑ #4

ΑΠΑΝΤΗΣΕΙΣ

Στόχος: Βασικός στόχος της 4ης εργασίας είναι η εξοικείωση σε θέματα Θεωρίας Πληροφορίας και συγκεκριμένα με τα μέτρα ποσότητας πληροφορίας τυχαίων μεταβλητών, τις σχετικές έννοιες και τα μέτρα διακριτών πηγών χωρίς μνήμη, καθώς και με τις ιδιότητες κωδίκων και την εφαρμογή αλγορίθμων κωδικοποίησης πηγής. Επίσης περιλαμβάνονται στην ύλη και τα κανάλια επικοινωνίας, όπως οι έννοιες της αβεβαιότητας, της αμοιβαίας πληροφορίας μεταξύ εισόδου και εξόδου και της χωρητικότητας ενός διακριτού καναλιού χωρίς μνήμη, καθώς και ο υπολογισμός τους.

Περιγραφή

Η 4^η εργασία περιλαμβάνει έξι (6) θέματα που σχετίζονται με τα 3 πρώτα κεφάλαια του τόμου Γ' «Θεωρία Πληροφορίας και Κωδικοποίησης»..

ΓΕχ (Γραπτή Εργασία χ) ή ΕΞχ(Εξετάσεις έτους χ Α ή Β)/Ακαδημαϊκό Έτος/ Αριθμός θέματος

| ΘΕΜΑ 1 | | | |
|---|-------------------|--------------------------|------------------------|
| Στόχος της άσκησης είναι η εξοικείωση με βασικά γνωρίσματα των τυχαίων μεταβλητών, με τις συνδυασμένες και υπό συνθήκη πιθανότητες καθώς και με σχετικά μέτρα πληροφορίας. | | | |
| Σχετικές ασκήσεις: ΓΕ3/1617/Θ4, ΓΕ3/1516/Θ4, ΓΕ4/1213/Θ1, ΓΕ3/1718/Θ5 | | | |
| Μια έρευνα κατέγραψε τον παρακάτω πίνακα συνδυασμένων πιθανοτήτων για τους λόγους απόλυσης και το επίπεδο μόρφωσης των απολυμένων $p(\text{Επίπεδο μόρφωσης}_i, \text{Λόγος απόλυσης}_j)$: | | | |
| $p(\text{Επίπεδο μόρφωσης}_i, \text{Λόγος απόλυσης}_j)$ | Επίπεδο μόρφωσης | | |
| Λόγος απόλυσης | Βασική Εκπαίδευση | Δευτεροβάθμια Εκπαίδευση | Τριτοβάθμια Εκπαίδευση |
| Κλείσιμο Μονάδας | 0,02 | 0,03 | 0,05 |
| Μείωση Παραγωγής | 0,075 | 0,325 | 0,1 |
| Κατάργηση θέσης Εργασίας | 0,04 | 0,08 | 0,28 |

Ζητούνται τα εξής:

(α) Ο πίνακας των υπό συνθήκη πιθανοτήτων ανάμεσα στο λόγο απόλυσης και το επίπεδο μόρφωσης $p(\text{Επίπεδο μόρφωσης}_i / \text{Λόγος απόλυσης}_j)$, η συνάρτηση πιθανότητας του επιπέδου μόρφωσης $p(\text{Επίπεδο μόρφωσης}_i)$ καθώς και η συνάρτηση πιθανότητας του λόγου απόλυσης $p(\text{Λόγος απόλυσης}_j)$.

(β) Η μέση ποσότητα πληροφορίας του λόγου απόλυσης και του επιπέδου μόρφωσης.

(γ) Η συνδυασμένη εντροπία του επιπέδου μόρφωσης ως προς το λόγο απόλυσης.

(δ) Η υπό συνθήκη (δεσμευμένη) εντροπία του επιπέδου μόρφωσης ως προς το λόγο απόλυσης και την υπό συνθήκη (δεσμευμένη εντροπία) του λόγου απόλυσης ως προς το επίπεδο μόρφωσης.

(ε) Την αμοιβαία πληροφορία του επιπέδου μόρφωσης και του λόγου απόλυσης.

Ενδεικτική Μεθοδολογία: Να εφαρμόσετε τις σχέσεις που συνδέουν τις συνδυασμένες με τις υπό συνθήκη πιθανότητες. Επίσης, να εφαρμόσετε τις σχέσεις υπολογισμού των διαφόρων μέτρων πληροφορίας, λαμβάνοντας υπόψη (όπου είναι δυνατό) τις σχέσεις μεταξύ των διαφόρων μέτρων για απλοποίηση των υπολογισμών.

ΘΕΜΑ 2

Στόχος της άσκησης είναι η κατανόηση εννοιών που σχετίζονται με την ποσότητα πληροφορίας, εντροπία, υπό συνθήκη εντροπία, αμοιβαία πληροφορία και την εξάρτηση μεταξύ τυχαίων μεταβλητών.

Σχετικές ασκήσεις: ΓΕ4/1920/Θ1, ΓΕ4/1819/Θ1

Έστω η τυχαία μεταβλητή (TM) X , η οποία περιγράφει την κατανομή της βαθμολογίας στις γραπτές εργασίες των φοιτητών της ΠΛΗ22. Έστω μια δεύτερη συσχετισμένη TM Y , η οποία περιγράφει το αν ένας φοιτητής πέρασε επιτυχώς ή απέτυχε στις τελικές εξετάσεις, με την παρακάτω από κοινού κατανομή:

| | | X | | | | | | | | | | | |
|-----|---------|--------|-----|-----|------|------|------|------|------|------|------|------|----|
| | | Βαθμός | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Y | Απέτυχε | 0.0 | 0.0 | 0.0 | 0.02 | 0.05 | 0.08 | 0.12 | 0.15 | 0.09 | 0.04 | 0.01 | |
| | Πέρασε | 0.0 | 0.0 | 0.0 | 0.0 | 0.01 | 0.02 | 0.05 | 0.10 | 0.10 | 0.10 | 0.06 | |

I) Να βρεθούν:

- Οι εντροπίες $H(X)$, $H(Y)$.
- Η από κοινού εντροπία $H(X, Y)$.
- Οι υπό συνθήκη εντροπίες $H(X|Y)$ και $H(Y|X)$. Πώς σχετίζονται οι $H(X)$, $H(Y)$ με τις αντίστοιχες υπό συνθήκη εντροπίες και γιατί;
- Η αμοιβαία πληροφορία $I(X; Y)$.

II) Υπολογίστε και σχολιάστε ποια ποσότητα πληροφορίας μεταφέρουν οι παρακάτω εκφράσεις για το βαθμό της γραπτής εργασίας ενός φοιτητή της ΠΛΗ22:

- Η εργασία βαθμολογήθηκε με βαθμό άνω του 4.
- Η εργασία βαθμολογήθηκε με βαθμό 10.
- Η εργασία βαθμολογήθηκε με βαθμό μικρότερο ή ίσο του 10.

Ενδεικτική Μεθοδολογία: Στο (I) α βρεθούν πρώτα οι περιθώριες συναρτήσεις πιθανότητας μάζας των τυχαίων μεταβλητών X και Y . Στο (II) να βρεθούν οι πιθανότητες εμφάνισης για καθένα από τα τρία ζητούμενα.

ΘΕΜΑ 3

Στόχος της άσκησης είναι η εξοικείωση με την έννοια της εντροπίας και τον υπολογισμό της, καθώς και εξοικείωση με τους αλγόριθμους κωδικοποίησης Huffman και Fano.

Σχετικές ασκήσεις: ΓΕ3/2017-18/Θ7, ΓΕ3/2015-16/Θ5, ΕΞ2018Β/Θ3, ΓΕ4/2018-19/Θ4, ΓΕ4/2019-20/Θ3

Θεωρήστε διακριτή πηγή που εκπέμπει τα παρακάτω σύμβολα ASCII: A, B, C, D, E, F, G. Η πηγή μεταδίδει το μήνυμα ACGABCFADAABBCE.

(α) Ποιο είναι το εκτεμπόμενο μήνυμα εφόσον η πηγή δεν χρησιμοποιεί κωδικοποίηση; Χρησιμοποιήστε τον ελάχιστο αριθμό bit ανά σύμβολο.

(β) Έστω η πιθανότητα χρήσης των συμβόλων είναι: $A=0,3$, $B=0,2625$, $C=0,2$, $D=0,15$, $E=0,05375$, $F=0,03125$. Υπολογίστε το πληροφοριακό περιεχόμενο του κάθε συμβόλου. Ποιο σύμβολο έχει το μεγαλύτερο πληροφοριακό περιεχόμενο και γιατί;

(γ) Να υπολογίσετε την εντροπία της παραπάνω πηγής.

(δ) Να σχηματίσετε τον κώδικα Huffman και να υπολογίσετε το βαθμό συμπίεσης για το παραπάνω μεταδιδόμενο μήνυμα.

(ε) Να σχηματίσετε τον κώδικα Fano και να υπολογίσετε το βαθμό συμπίεσης για το παραπάνω μεταδιδόμενο μήνυμα.

(στ) Ποια είναι η επίδοση για τους δυο παραπάνω αλγόριθμους (Huffman και Fano); Ποιος αλγόριθμος έχει καλύτερη επίδοση και γιατί;

Ενδεικτική Μεθοδολογία: Από τις πιθανότητες που σας δίνονται υπολογίστε το πληροφοριακό περιεχόμενο του κάθε συμβόλου. Ο βαθμός συμπίεσης αντιστοιχεί στο ποσοστό των δυαδικών bits που εξοικονομούνται στην κάθε μια από τις κωδικοποιήσεις των ερωτημάτων σε σχέση με τον αριθμό των δυαδικών bits που χρησιμοποιούνται στο ερώτημα (α) για τη μετάδοση του συγκεκριμένου μηνύματος. Η επίδοση ενός αλγόριθμου κωδικοποίησης αντιστοιχεί στο πηλίκο της εντροπίας προς το μέσο μήκος των κωδικών λέξεων.

ΘΕΜΑ 4

Στόχος της άσκησης είναι η κατανόηση των αλγορίθμων κωδικοποίησης πηγών χωρίς μνήμη και η μεταξύ τους σύγκριση σε ότι αφορά στο μέσο μήκος κώδικα και στην επίδοση που επιτυγχάνουν.

Σχετικές ασκήσεις: ΓΕ3/2019-20/Θ3 και Θ4, ΓΕ3/2018-19/Θ4, ΓΕ3/2017-18/Θ7, ΓΕ3/2016-17/Θ6, ΓΕ3/2015-16/Θ5

Μια διακριτή πηγή πληροφορίας χωρίς μνήμη παράγει 6 σύμβολα A, B, Γ, Δ, E, Z με τις αντίστοιχες πιθανότητες:

$$\mathbf{p} = \left\{ \frac{1}{21}, \frac{5}{21}, \frac{3}{21}, \frac{2}{21}, \frac{6}{21}, \frac{4}{21} \right\}$$

(α) Να κατασκευάσετε δυαδικό κώδικα με τον αλγόριθμο Huffman και να υπολογίσετε το μέσο μήκος των κωδικών λέξεων.

(β) Να υπολογίσετε την επίδοση του κώδικα.

(γ) Μπορείτε να κατασκευάσετε ένα άλλο άμεσο κώδικα για την παραπάνω πηγή όπου η κάθε κωδική λέξη να έχει μήκος 2 ή 3; Είναι ένας τέτοιος κώδικας βέλτιστος;

(δ) Κατασκευάστε μία κατανομή πιθανότητας \mathbf{p}' για έξι σύμβολα για την οποία μπορεί να κατασκευαστεί κώδικας με μέσο μήκος (ως προς την \mathbf{p}') ίσο με την εντροπία $H(\mathbf{p}')$.

(ε) Να κατασκευάσετε δυαδικό κώδικα με τον αλγόριθμο Shannon, να υπολογίσετε το μέσο μήκος των κωδικών λέξεων και την επίδοση του κώδικα. Συγκρίνετε ως προς το μέσο μήκος και την επίδοση του κώδικα αυτού με τον κώδικα που κατασκευάσατε στο ερώτημα α).

Ενδεικτική μεθοδολογία: Στο ερώτημα (γ) θα πρέπει να χρησιμοποιήσετε την ανισότητα Kraft για να δείξετε ότι τα δύο σύνολα μηκών κωδικών λέξεων την ικανοποιούν και στη συνέχεια να συγκρίνετε το μέσο μήκος που θα βρείτε με αυτό του ερωτήματος (α). Στο ερώτημα (δ) ανατρέξτε στη σχέση που συνδέει την εντροπία με τον μήκος κωδικών λέξεων.

ΘΕΜΑ 5

Στόχος της άσκησης είναι να κατανοήσετε πως επηρεάζεται η χωρητικότητα του καναλιού καθώς και άλλα μεγέθη εντροπίας με βάση τις πιθανότητες μετάβασης του ίδιου τύπου καναλιού. Επίσης, η άσκηση χρησιμεύει να εξασκηθείτε στον υπολογισμό των βασικών μεγεθών των καναλιών όπως οι πιθανότητες εξόδου, αμοιβαία πληροφορία, χωρητικότητα κλπ.

Σχετικές ασκήσεις: ΓΕ4/1920/Θ6

Δίνονται οι παρακάτω πίνακες μετάβασης των καναλιών, A, B, Γ, με κατανομή των συμβόλων εισόδου $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$:

$$A: \quad P_{Y/X} = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 \\ 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0 & 0.9 \end{bmatrix}$$

$$B: \quad P_{Y/X} = \begin{bmatrix} 0.7 & 0.3 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0 \\ 0 & 0 & 0.7 & 0.3 \\ 0.3 & 0 & 0 & 0.7 \end{bmatrix}$$

$$\Gamma: \quad P_{Y/X} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

Για καθένα από τα κανάλια να βρείτε:

(α) Την $H(Y)$ και αμοιβαία πληροφορία. Τι παρατηρήσεις μπορείτε να κάνετε σχετικά με τις τιμές της $H(Y)$ και της αμοιβαίας πληροφορίας σε σχέση με τους πίνακες μετάβασης.

(β) Ποιο από τα κανάλια έχει την μεγαλύτερη χωρητικότητα; Τι παρατηρείτε;

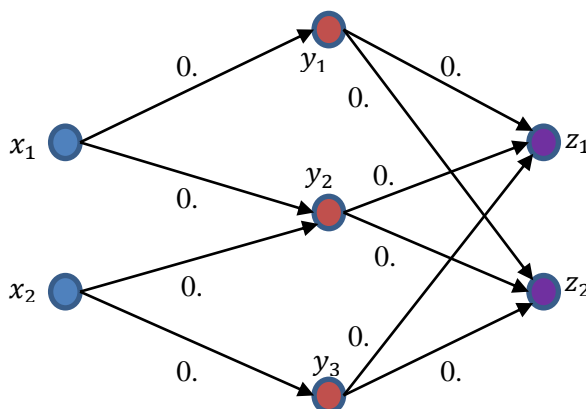
Ενδεικτική Μεθοδολογία: Για τα ερωτήματα (α) και (β) εφαρμόζουμε τους τύπους βάσει των οποίων υπολογίζουμε την $H(Y)$, $H(Y/X)$ και $I(X;Y)$ για κάθε κανάλι ξεχωριστά. Για το ερώτημα (β) θα βοηθήσει να βρείτε το είδος του πίνακα μετάβασης έτσι ώστε να υπολογίσετε χωρίς πράξεις την κατανομή των συμβόλων εξόδου και αντίστοιχα την εντροπία τους. Για να απαντήσετε το ερώτημα «τί παρατηρείτε» δείτε ποια είναι η τάση αυξομειώσης της $H(Y)$ και $I(X;Y)$ και χωρητικότητα, σε συνάρτηση με τους πίνακες μετάβασης.

ΘΕΜΑ 6

Στόχος της άσκησης είναι η κατανόηση της έννοιας της χωρητικότητας καναλιού και του 2^{ου} θεωρήματος κωδικοποίησης του Shannon με εφαρμογή στα δυαδικά συμμετρικά κανάλια.

Σχετικές ασκήσεις: ΓΕ4/1718/Θ5, ΓΕ4/1920/Θ5, ΓΕ4/1819/Θ5

Δίδεται το παρακάτω γράφημα με τις υπό συνθήκη πιθανότητες μετάβασης για τα δύο διαδοχικά κανάλια $X \rightarrow Y \rightarrow Z$.



- Αφού σχηματιστούν οι επί μέρους πίνακες μετάβασης $X \rightarrow Y$ και $Y \rightarrow Z$, να βρεθεί ο πίνακας μετάβασης $X \rightarrow Z$ και να υπολογιστεί η χωρητικότητα του καναλιού $X \rightarrow Z$ (από άκρο-σε-άκρο).
- Να βρεθεί η χωρητικότητα καναλιού για δύο διαδοχικά κανάλια, το καθένα με πίνακα μετάβασης όπως αυτόν του καναλιού $X \rightarrow Z$ και να σχολιάσετε το αποτέλεσμα ως προς το (α).
- Να βρεθεί η χωρητικότητα καναλιού του για 100 διαδοχικά κανάλια, το καθένα με πίνακα μετάβασης όπως αυτόν του καναλιού $X \rightarrow Z$ και να σχολιάσετε το αποτέλεσμα ως προς το (α).
- Αν οι πιθανότητες εμφάνισης των συμβόλων A, B μιας δυαδικής πηγής είναι $p_A = 0.1$, $p_B = 0.9$ με τα σύμβολα να παράγονται με ρυθμό $r_s = 8$ symbols/sec, να βρεθεί αν υπάρχει δυνατότητα αξιόπιστης μετάδοσης διαμέσω του καναλιού που βρήκατε στο ερώτημα (α), αν γνωρίζουμε ότι ο μέγιστος ρυθμός μετάδοσης πάνω από το κανάλι είναι $r_c = 12$ symbols/sec.

Ενδεικτική μεθοδολογία: Στο (α) οι από άκρο-σε-άκρο πιθανότητες μετάβασης προκύπτουν ακολουθώντας τις διαφορετικές διαδρομές στο παραπάνω σχήμα. Πιο γρήγορος τρόπος επίλυσης των (α), (β) και (γ) μπορεί να προκύψει αν βρεθεί ο από άκρο-σε-άκρο πίνακας μετάβασης ως ένα γινόμενο των διαδοχικών πινάκων μετάβασης των καναλιών από τα οποία απαρτίζεται. Ειδικά για το (γ) χρησιμοποιείστε το Octave για να υπολογίσετε την 100στη δύναμη του πίνακα μετάβασης του καναλιού $X \rightarrow Z$.

Στο (δ) να βρεθεί αρχικά η εντροπία και ο ρυθμός πληροφορίας της δυαδικής πηγής σε bits/sec. Στη συνέχεια θα πρέπει να βρείτε το μέγιστο ρυθμό μετάδοσης πάλι σε bits/sec λαμβάνοντας υπόψη το ρυθμό μετάδοσης του καναλιού r_c . Συγκρίνοντας τα δύο μεγέθη, ρυθμός πληροφορίας και ρυθμός μετάδοσης, θα πρέπει να αποφανθείτε για τη δυνατότητα αξιόπιστης μετάδοσης του καναλιού (μετάδοση με οσοδήποτε μικρή πιθανότητα λάθους) σύμφωνα με το 2^ο θεμελιώδες θεώρημα του Shannon. Επειδή στην πράξη κωδικοποιούμε τα σύμβολα της πηγής πριν τα μεταδώσουμε -κατά προτίμηση με βέλτιστο τρόπο-, στην πράξη ο ρυθμός πληροφορίας της πηγής εξαρτάται από το μέσο μήκος της κωδικής λέξης L_s , που ενδέχεται να υπερβαίνει το ρυθμό μετάδοσης. Για να βρείτε αν υπάρχει δυνατότητα αξιόπιστης μετάδοσης θα πρέπει να κωδικοποιήσετε με τέτοιο τρόπο τα σύμβολα της πηγής έτσι ώστε να πλησιάσετε το όριο της εντροπίας της πηγής ανά σύμβολο. Που σημαίνει ότι θα πρέπει να κωδικοποιήσετε ανά 2 ή 3 σύμβολα μαζί και να δείτε κάθε φορά αν το μέσο μήκος/σύμβολο που επιτυγχάνετε είναι τέτοιο που δημιουργεί ρυθμό πληροφορίας μικρότερο από τον ρυθμό μετάδοσης.

Τρόπος – Ημερομηνία Παράδοσης

1. Την εργασία σας θα πρέπει να την ανεβάσετε στο σύνδεσμο <http://study.eap.gr> έως την **Τετάρτη, 14 Απριλίου 2021, 23:59**. Δεν θα γίνονται αποδεκτές και δεν θα αξιολογούνται εργασίες που αποστέλλονται στον Καθηγητή-Σύμβουλο μέσω email.
2. Όλοι οι φοιτητές θα πρέπει να χρησιμοποιήσουν το template που υπάρχει στο σύνδεσμο <http://study.eap.gr>
3. Στις 19 Απριλίου 2019 θα δημοσιευθούν οι ενδεικτικές απαντήσεις των θεμάτων της εργασίας στο <http://study.eap.gr>

Βαθμολογία ανά Θέμα

| | |
|----------------|-----------|
| ΘΕΜΑ 1 | 15 |
| Ερώτημα (α) | 3 |
| Ερώτημα (β) | 3 |
| Ερώτημα (γ) | 3 |
| Ερώτημα (δ) | 3 |
| Ερώτημα (ε) | 3 |
| ΘΕΜΑ 2 | 17 |
| Ερώτημα (I.α) | 2 |
| Ερώτημα (I.β) | 2 |
| Ερώτημα (I.γ) | 2 |
| Ερώτημα (I.δ) | 2 |
| Ερώτημα (II.α) | 3 |
| Ερώτημα (II.β) | 3 |
| Ερώτημα (II.γ) | 3 |
| ΘΕΜΑ 3 | 17 |
| Ερώτημα (α) | 2 |
| Ερώτημα (β) | 2 |
| Ερώτημα (γ) | 2 |
| Ερώτημα (δ) | 2 |
| Ερώτημα (ε) | 3 |
| Ερώτημα (στ) | 3 |
| ΘΕΜΑ 4 | 17 |
| Ερώτημα (α) | 3 |
| Ερώτημα (β) | 3 |
| Ερώτημα (γ) | 4 |
| Ερώτημα (δ) | 3 |
| Ερώτημα (ε) | 4 |
| ΘΕΜΑ 5 | 16 |
| Ερώτημα (α) | 8 |
| Ερώτημα (β) | 8 |
| ΘΕΜΑ 6 | 18 |
| Ερώτημα (α) | 4 |
| Ερώτημα (β) | 4 |
| Ερώτημα (γ) | 4 |

| | |
|---------------|------------|
| Ερώτημα (δ) | 6 |
| ΣΥΝΟΛΟ | 100 |

Οι συνολικές μονάδες θα διαιρεθούν με το 10, ώστε να προκύψει ο τελικός βαθμός της εργασίας.

Καλή Επιτυχία!!!