Ανάλυση Απόδοσης Πληροφοριακών Συστημάτων

Διαλέξεις 12-13: Βελτιστοποίηση της Εντροπίας

Καθ. Γιάννης Γαροφαλάκης
ΜΔΕ Επιστήμης και Τεχνολογίας Υπολογιστών
Τμήμα Μηχανικών Η/Υ και Πληροφορικής

Εισαγωγή

- Η εντροπία παίζει σημαντικό ρόλο στη διαμόρφωση πιθανοτικών συστημάτων σε πολλές περιοχές.

- Το ενδιαφέρον εδώ επικεντρώνεται στην εντροπία της πληροφορίας.

Βασικές Συνεισφορές

- Η μετρική Shannon για εντροπία της πληροφορίας(1948)
- Η αρχή Μέγιστης Εντροπίας του Jayne (MaxEnt-1957)
- Η αρχή Ελάχιστης Εντροπίας του Kullback (MinEnt-1959)

Αρχή Μέγιστης Εντροπίας

- Μέθοδος στατιστικής παρεμβολής, όταν η πληροφορία μας για ένα πρόβλημα υπάρχει με την μορφή μέσων τιμών κι άλλων ροπών.

- Υπολογίζει τις κατανομές, μεγιστοποιώντας τη μετρική Shannon, υπό γραμμικούς περιορισμούς ροπών.

Αρχή ελάχιστης εντροπίας του Kullback

- Μαθηματική μέθοδος που βασίζεται:
  - Στο Kullback-Leibler(K-L) μέτρο
    - Ένα μετρό για την απόσταση μεταξύ 2 πιθανοτικών κατανομών.
    - Εκφράζεται ως το άθροισμα 2 λογαριθμικών όρων, ένας από τους οποίους είναι η μετρική Shannon.
  - Στην ελαχιστοποίηση του K-L μέτρου υπό συγκεκριμένους γραμμικούς περιορισμούς
- Όταν η μία κατανομή είναι η ομοιόμορφη, τότε η αρχή ελάχιστης εντροπίας ισοδυναμεί με την αρχή μέγιστης εντροπίας.

Τι είναι η εντροπία;

- Μετασχηματισμός από τάξη σε αταξία.
- Ουσιαστικά ισοδυναμεί με την αβεβαιότητα.
- Όσο περισσότερη πληροφορία συγκεντρώνεται, τόσο αυξάνεται η εντροπία.
- Εντροπία $\equiv$ Πιθανοτική Αβεβαιότητα

Τι είναι η πιθανοτική αβεβαιότητα;

- Είναι η αβεβαιότητα που σχετίζεται με την πιθανότητα αποτελεσμάτων.

- Για κάθε πρόβλημα μπορεί να υπάρχουν *n* πιθανά αποτελέσματα με πιθανότητες $p_1, p_2, \ldots p_n$, όπου $p_1 \geq 0$, $p_2 \geq 0, \ldots,$ $p_n \geq 0$ και $\sum_{i=1}^{n} p_i = 1$
  - Η αβεβαιότητα αφορά το ποιο αποτέλεσμα θα έρθει.
  - Δεν είναι γνωστές οι τιμές $p_1, p_2, \ldots p_n$.
  - Διαφορετικές κατανομές έχουν διαφορετική αβεβαιότητα.

Τι είναι η βελτιστοποίηση;

- Περιλαμβάνει:
  - Μεγιστοποίηση
  - Ελαχιστοποίηση ή
  - ταυτόχρονη μεγιστοποίηση μιας συνάρτησης κι ελαχιστοποίηση μιας άλλης.
- Συνηθίζεται η βελτιστοποίηση υπό περιορισμούς
  - π.χ βελτιστοποίηση απόδοσης φοιτητή υπό περιορισμούς χρόνου και προετοιμασίας
  - Για προβλήματα βελτιστοποίησης εντροπίας προτιμάται η μέθοδος του Lagrange.

Γιατί βελτιστοποιούμε την εντροπία;(1)

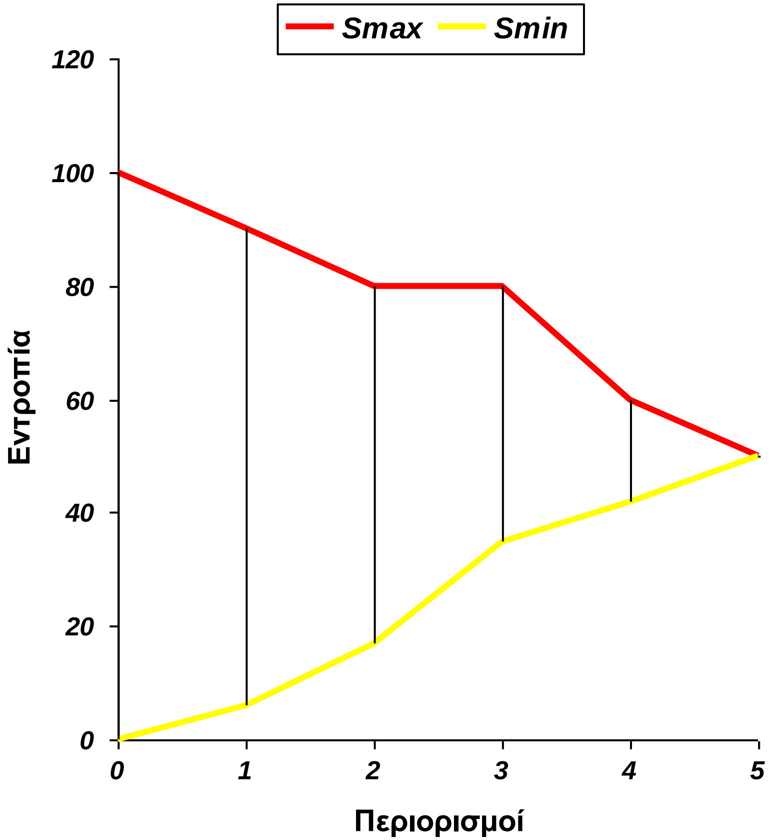
- Στόχος είναι η μείωση της αβεβαιότητας με τη συγκέντρωση περισσότερης πληροφορίας.
- Μελετώντας ένα πρόβλημα χωρίς κάποιες πληροφορίες η αβεβαιότητα για τις κατανομές είναι πολύ μεγάλη.
- Συλλέγοντας βηματικά πληροφορίες, προστίθενται περιορισμοί στο πρόβλημα, που μειώνουν την αβεβαιότητα.

Γιατί βελτιστοποιούμε την εντροπία;(2)

- Σε κάθε βήμα από τις κατανομές που ικανοποιούν τους μέχρι τότε περιορισμούς μία θα έχει μέγιστη αβεβαιότητα(S_{max}) και μία θα έχει ελάχιστη αβεβαιότητα(S_{min}).
- Το σύνολο των κατανομών σε κάθε βήμα είναι υποσύνολο των κατανομών του προηγούμενου βήματος,οπότε
  - το S_{max} μειώνεται ή τουλάχιστον δε μεγαλώνει
  - και το S_{min} αυξάνει ή τουλάχιστον δε μικραίνει.

Γιατί βελτιστοποιούμε την εντροπία;(3)

- Στο τελευταίο βήμα μόνο μία κατανομή έχει και ελάχιστη και μέγιστη αβεβαιότητα.
- Σε κάθε βήμα η αβεβαιότητα μιας αποδεκτής κατανομής βρίσκεται μεταξύ S_{min} και S_{max}.
- Σχηματικά, η βελτιστοποίηση φαίνεται στο διπλανό γράφημα.



Τι εννοούμε με τη μείωση αβεβαιότητας;

- Ο αριθμός των πιθανών κατανομών μειώνεται.

- Το S_{max}-S_{min} μειώνεται.

- Δεν μειώνεται η αβεβαιότητα για το αποτέλεσμα του πειράματος, μόνο η αβεβαιότητα για τις τιμές των $p_1, p_2, \ldots p_n$

Αρχές Βελτιστοποίησης(1)

- Σε κάθε βήμα μπορεί να έχουμε άπειρες κατανομές που ικανοποιούν τους περιορισμούς:

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g_{ri} = a_r, \quad r = 1,2,\ldots.m, \qquad (*)$$

$$m+1 < n, \quad p_1 \geq 0, \ldots, p_n \geq 0$$

- Από όλες αυτές τις κατανομές, η S_{max} έχει τη μέγιστη αβεβαιότητα και η αβεβαιότητα κάθε άλλης κατανομής είναι μικρότερη της S_{max}.
- Η χρήση άλλης κατανομής από την S_{max} προϋποθέτει επιπλέον πληροφορία.

Αρχές Βελτιστοποίησης(2)

- ΑΡΧΗ: Πρέπει να χρησιμοποιούμε όλη τη διαθέσιμη πληροφορία και να μη χρησιμοποιούμε πληροφορία που δε έχει δοθεί. (επιστημονική αντικειμενικότητα και τιμιότητα)

Χρησιμοποιούμε μόνο κατανομές που υπόκεινται στους αρχικούς περιορισμούς, πιθανόν από άπειρες. Θα επιλέξουμε την S_{max}, διότι δεν έχουμε επιπλέον πληροφορία ακόμη.

Αρχή Μέγιστης Εντροπίας

Περιγραφή

Ορισμός

Από όλες τις κατανομές που υπόκεινται σε ένα δεδομένο σύνολο περιορισμών, επιλέγουμε αυτή που έχει μεγαλύτερη αβεβαιότητα.

Ερμηνεία της κατανομής S_{min}

- Ξεκινάμε με S_{max} και προχωράμε βάζοντας επιπλέον περιορισμούς στην (*).
- Έτσι, μειώνεται το S_{max}.
- Δεν μπορούμε να συνεχίσουμε κάτω από S_{min}. Οι περιορισμοί αυτοί υπάρχουν, επειδή υπάρχει το S_{min}.
- Βρίσκοντας τους περιορισμούς αυτούς, θα έχουμε το μέγιστο σύνολο περιορισμών που είναι αποδεκτοί από τους αρχικούς περιορισμούς.

Παρατηρήσεις

- Το πρόβλημα της εύρεσης όλων των πιθανών περιορισμών είναι γνωστό ως Pattern Recognition.
- Η μεγιστοποίηση κι η ελαχιστοποίηση της αβεβαιότητας είναι σημαντικές για να γνωρίζουμε το εύρος αβεβαιότητας και το μέγεθος μείωσης του από δεδομένη πληροφορία.

Παράδοξο

Ενώ στόχος μας είναι η μείωση αβεβαιότητας, γιατί προσπαθούμε να τη μεγιστοποιήσουμε;

↓

Μειώνουμε την αβεβαιότητα με τη βοήθεια της πληροφορίας που μας δίνεται κι έτσι επιτυγχάνουμε μέγιστη μείωση με χρήσης <u>όλης</u> της διαθέσιμης πληροφορίας.

Όμως, πάντα μένει αβεβαιότητα, επειδή υπάρχει πληροφορία που δε μας έχει δοθεί. Δηλαδή, μένουμε όσο το δυνατό περισσότερο αβέβαιοι για την άγνωστη πληροφορία. Είναι η πιο τίμια προσέγγιση.

Εντροπία ή Αβεβαιότητα;

Πλεονεκτήματα-Μειονεκτήματα

Αρχικός Ορισμός Εντροπίας

- Υπήρχε ενδιαφέρον από το Shannon για ένα μέτρο πληροφορίας, με στόχο να μετρήσει την πληροφορία που στέλνεται/λαμβάνεται/χάνεται σε θορυβώδη κανάλια.
- Ανέπτυξε ένα μέτρο αβεβαιότητας μιας πιθανοτικής κατανομής $\overline{p} = \left(p_1, p_2, \ldots p_n\right)$

$$S\left(\overline{p}\right) = -k \sum_{i=1}^{n} p_i \ln p_i$$

όπου *k* αυθαίρετη θετική σταθερά

- Ονομάστηκε εντροπία κι όχι μέτρο πληροφορίας, γιατί ο ορισμός είναι ίδιος με τον ορισμό της εντροπίας στη θερμοδυναμική.

Χρήση Όρου Εντροπία Αντί Για Αβεβαιότητα

Αρνητικά Σημεία	Θετικά Σημεία
Η εντροπία θεωρείται όρος μόνο της φυσικής και μηχανικής.	Ο όρος αβεβαιότητα είναι πολύ ευρύς για να αφορά μόνο τη μετρική του Shannon.
Όταν η εντροπία χρησιμοποιείται σε άλλα πεδία, λανθασμένα θεωρείται ότι επιβάλλονται νόμοι θερμοδυναμικής σε αυτά.	Δεν ενδιαφερόμαστε για αβεβαιότητα λόγω fuziness της πληροφορίας παρά μόνο για μερικές πλευρές πιθανοτικής αβεβαιότητας.
Πρεστίζ στο Shannon, γιατί η εντροπία θεωρείται «βαθιά» έννοια.	Μετρούνται εκτός από την αβεβαιότητα και: ισότητα, αταξία, διαφορετικότητα, ομοιότητα, τυχαιότητα, κ.τ.λ

Μια Άλλη Αρχή Βελτιστοποίησης Εντροπίας

Περιγραφή, Περιορισμοί

Η αρχή ανεπαρκούς εξήγησης του Laplace(1)

- Ειδική περίπτωση νόμου μέγιστης αβεβαιότητας όταν είναι γνωστοί μόνο οι περιορισμοί:

$$\sum_{i=1}^{n} p_i = 1, \quad p_1 \geq 0, \ldots, p_n \geq 0$$

- Δεν υπάρχει τότε λόγος να επιλέξουμε άλλες τιμές για τις πιθανότητες από $p_1 = p_2 = \ldots = p_n = 1/n$. Επιλέγουμε δηλαδή την ομοιόμορφη κατανομή : $U = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$
- Αυτή είναι η κατανομή με μέγιστη εντροπία από όλα τα εξαγόμενα,εφόσον κάθε άλλη κατανομή απαιτεί επιπλέον πληροφορία.

Η αρχή ανεπαρκούς εξήγησης του Laplace(2)

- Αν έχουμε επιπλέον περιορισμούς της μορφής
$$\sum_{i=1}^{n} p_i g_{ri} = a_r, \;\; r = 1,2,\ldots m, \quad m+1 < n$$
τότε το να είμαστε αβέβαιοι, μπορεί να σημαίνει να είμαστε όσο πιο κοντά στην ομοιόμορφη κατανομή, που είναι η πιο αβέβαιη.
- Χρειαζόμαστε μία μετρική απόστασης μεταξύ δύο κατανομών $\bar{p} = (p_1, p_2, \ldots, p_n)$ και $\bar{q} = (q_1, q_2, \ldots, q_n)$.

Η αρχή ανεπαρκούς εξήγησης του Laplace(3)

- Μπορούμε να χρησιμοποιήσουμε το μέτρο της Ευκλείδιας απόστασης

$$\left[\sum_{i=1}^{n}\left(p_i - q_i\right)^2\right]^{1/2} \quad ή \quad \sum_{i=1}^{n}\left|p_i - q_i\right|$$

- ή το μέτρο των Kullback-Leibler

$$\sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

- ή οποιαδήποτε άλλη μετρική μπορεί να χρησιμοποιηθεί ως μέτρο απόστασης

Επαναδιατύπωση της Αρχής Μέγιστης Αβεβαιότητας(1)

- ΑΡΧΗ 1: Από όλες τις κατανομές που ικανοποιούν δεδομένους περιορισμούς, επιλέξτε αυτή που είναι πιο κοντά στην ομοιόμορφη κατανομή *p*.
- Υπάρχουν περιπτώσεις που έχουμε λόγους να επιλέξουμε διαφορετικές p_1 , p_2 ,…,p_n . Για παράδειγμα, αν έχουμε αρχικές τιμές των πιθανοτήτων τις χρησιμοποιούμε.

Επαναδιατύπωση της Αρχής Μέγιστης Αβεβαιότητας(2)

- ΑΡΧΗ 2: Από όλες τις κατανομές που ικανοποιούν τους δοσμένους περιορισμούς, επιλέξτε αυτή που είναι πιο κοντά στην a priori κατανομή q.(Αρχή Ελάχιστης Cross-Entropy)
- Δηλαδή, έχουμε 2 τύπους διαθέσιμης πληροφορίας:
  - Με μορφή a priori κατανομής q που βασίζεται στη διαίσθηση ή στην εμπειρία σχετικά με το πρόβλημα
  - Με τη μορφή περιορισμών που βασίζεται στην παρατήρηση.

Επαναδιατύπωση της Αρχής Μέγιστης Αβεβαιότητας(3)

- Η μεγιστοποίηση αβεβαιότητας ισοδυναμεί με την ελαχιστοποίηση της <u>κατευθυνόμενης διαφοράς</u> από την ομοιόμορφη ή την a priori κατανομή.
- ΓΕΝΙΚΗ ΑΡΧΗ: Από όλες τις κατανομές που ικανοποιούν τους περιορισμούς, επιλέξτε αυτή που είναι πιο κοντά στην a priori κατανομή, κι αν αυτή δεν υπάρχει, τότε επιλέξτε αυτή που είναι πιο κοντά στην ομοιόμορφη.

Υπολογισμός της κατευθυνόμενης διαφοράς(1)

- Δεν πρόκειται για αποστάσεις στον Ευκλείδιο χώρο. Εδώ κάθε σημείο αντιπροσωπεύει μια κατανομή.
- Υπάρχουν κοινές ιδιότητες αλλά και διαφορές

Κοινές Ιδιότητες	Διαφορετικές Ιδιότητες
Θέλουμε $D(\overline{p}:\overline{q})\geq 0$ και $D(\overline{p}:\overline{q})=0$ αν $\overline{p}=\overline{q}$	Δεν απαιτούμε $D(\overline{p}:\overline{q})=D(\overline{q}:\overline{p})$
Θέλουμε η $D(\overline{p}:\overline{q})$ να είναι κυρτή για να βρίσκουμε εύκολα το ελάχιστο της	Δεν απαιτούμε $D(\overline{p}:\overline{q})+D(\overline{q}:\overline{r})\geq D(\overline{p}:\overline{r})$
Θέλουμε τα στοιχεία του $\overline{p}$ που ελαχιστοποιούν τη $D(\overline{p}:\overline{q})$ να είναι μεγαλύτερα του μηδενός	

Υπολογισμός της κατευθυνόμενης διαφοράς(2)

- Το πιο χρήσιμο και απλό μέτρο που ικανοποιεί τις απαιτήσεις μας είναι των Kullback, Leibler:

$$D(\overline{p}:\overline{q}) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

όπου επιλέγουμε *p* έτσι ώστε να ελαχιστοποιείται η μετρική υπό τους περιορισμούς. Αυτή είναι η *αρχή ελάχιστης cross entropy του Kullback .*

Υπολογισμός της κατευθυνόμενης διαφοράς(3)

- Όταν $\bar{q} = \bar{U}$

$$D(\bar{p}:\bar{u}) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} = \ln n - \left(-\sum_{i=1}^{n} p_i \ln p_i \right) = \ln n - S(\bar{p})$$

όπου $S\left(\bar{p}\right)$ η μετρική Shannon

Δηλαδή, όταν δεν έχουμε $\bar{q}$ επιλέγουμε το *p*, έτσι ώστε να μεγιστοποιείται η μετρική Shannon υπό τους περιορισμούς.

Αυτή είναι η *αρχή μέγιστης εντροπίας του Jaynes*.

Άλλες Αρχές Βελτιστοποίησης Εντροπίας

- Inverse Αρχή Βελτιστοποίησης Εντροπίας

  Μας δίνονται 2 από τα 3:

  - Σύνολο από περιορισμούς ροπών *C*
  - Μετρική εντροπίας ή cross-εντροπίας *M*
  - A priori κατανομή $\overline{q}$

  Μας ζητείται ο προσδιορισμός του τρίτου, ώστε μία δεδομένη $\overline{p}$ να έχει μέγιστη εντροπία ή ελάχιστη cross-εντροπία.

- Καθολική Αρχή Βελτιστοποίησης Εντροπίας

  Μας δίνονται τρία από τα $\overline{p}, \overline{q}, C, M$ και ζητείται να βρούμε το τέταρτο, ώστε το $D(\overline{p} : \overline{q})$ να είναι ελάχιστο.

Η μετρική Shannon

Ιδιότητες, Χαρακτηριστικά

Επιθυμητές Ιδιότητες

1. Η *S* θα πρέπει να είναι συνάρτηση όλων των $p_1, p_2, \ldots, p_n$.
2. Η $S(p_1, p_2, \ldots, p_n)$ συνεχής συνάρτηση των $p_1, p_2, \ldots, p_n$.
3. Η $S(p_1, p_2, \ldots, p_n)$ συμμετρική στα $p_1, p_2, \ldots, p_n$.
4. Η $S(1/n, 1/n, \ldots, 1/n)$ μονοτονικά αύξουσα συνάρτηση του *n*.
5. $S(p_1, p_2, \ldots, p_n) = S(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2) S\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$

$$\Downarrow$$

$$S_n(\overline{p}) = -k \sum_{i=1}^{n} p_i \ln p_i$$

Συνέπειες

- Η S είναι κοίλη συνάρτηση των $p_1, p_2, \ldots, p_n$
- Όταν μεγιστοποιείται η S υπό γραμμικούς περιορισμούς, οι πιθανότητες που τη μεγιστοποιούν είναι όλες μεγαλύτερες του μηδέν.

Ιδιότητες μετρικής Shannon (1)

1. Η $S_n(\overline{p})$ είναι συμμετρική.Δεν αλλάζει τιμή, αν αλλάξει η θέση των *p₁,p₂,…,pₙ.*
2. Οι p_i και lnp_i είναι συνεχείς συναρτήσεις του p_i, όταν $0 < p_i \leq 1$ κι έτσι η $p_i\, lnp_i$ είναι συνεχής. Όταν p_i=0, το $p_i\, lnp_i$ δεν ορίζεται, αλλά

$$\lim_{p_i \to 0} p_i \ln p_i = \lim_{p_i \to 0} \frac{\ln p_i}{1/p_i} = \lim_{p_i \to 0} \frac{1/p_i}{-1/p_i^{\,2}}$$

$$= \lim_{p_i \to 0}(-p_i) = 0$$

οπότε μπορούμε να δεχτούμε συνέχεια και για p_i*=0,* ορίζοντας $0 \ln 0 \equiv 0$

Ιδιότητες μετρικής Shannon (2)

3. Η S δεν αλλάζει αν βάλουμε απίθανο γεγονός.

   Δηλαδή, $S_{n+1}(p_1, p_2, \ldots, p_n, 0) = -\sum_{i=1}^{n} p_i \ln p_i - 0 \ln 0 =$

   $$= -\sum_{i=1}^{n} p_i \ln p_i = S_n(\overline{p})$$

4. Όταν $0 < p_i < 1$, τότε $\ln p_i < 0$ και $-\ln p_i > 0$.
   Όταν $p_i = 1$ ή 0 , τότε $p_i \ln p_i = 0$. Συνεπώς, $-p_i \ln p_i \geq 0$ και $S_n(\overline{p}) \geq 0$.

Ιδιότητες μετρικής Shannon (3)

5. Οι κατανομές
   Δ_1=(1,0,….,0)

   Δ_2=(0,1,….,0)

   …

   Δ_n=(0,0,….,1)

   δίνουν πάντα $S_n(\overline{p}) = 0$.
6. Για όλες τις υπόλοιπες κατανομές $S_n(\overline{p}) > 0$

Ιδιότητες μετρικής Shannon(4)

7. Αν ορίσουμε $\varphi(p) = -p \ln p$ τότε $\varphi'(p) = -(1 + \ln p)$ και $\varphi''(p) = -{}^{1}\!/\!_{p}$
Δηλαδή, η $\varphi(p)$ είναι κοίλη συνάρτηση του *p*.

8. Η μέγιστη τιμή της $S_n(\overline{p})$ δίνεται όταν $\overline{p} = \overline{u} = \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)$

9. Η μέγιστη τιμή είναι: $S_n(\overline{u}) = -\sum_{i=1}^{n} \frac{1}{n} \ln \frac{1}{n} = \ln n$

Το ln*n* είναι αύξουσα συνάρτηση , οπότε και το $S_n(\overline{u})$ που είναι επιθυμητό γιατί όσο περισσότερα πιθανά εξαγόμενα, τόσο μεγαλύτερη αβεβαιότητα.

Ιδιότητες μετρικής Shannon(5)

10. Έστω $\overline{p}=(p_1, p_2,..., p_3)$, $\overline{q}=(q_1, q_2,..., q_3)$ ανεξάρτητες κατανομές των Τ.Μ. *X,Y* αντίστοιχα:

P(X=x_i)=p_i, P(Y=y_j)=q_j και *P(X=x_i, Y=y_j)= $p_i\, q_j$.*

Για την κοινή κατανομή των *x,y* υπάρχουν *mn* πιθανά εξαγόμενα με πιθανότητα $p_i\, q_j$.

Η εντροπία της κοινής κατανομής είναι:

$$S_{mn}(\overline{p}\cdot\overline{q}) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_j \ln(p_i q_j)$$

$\xrightarrow{\Sigma\upsilon\nu\acute{\varepsilon}\chi\varepsilon\iota\alpha}$

Ιδιότητες μετρικής Shannon(6)

$$S_{mn}(\overline{p}\cdot\overline{q})=-\sum_{j=1}^{m}\sum_{i=1}^{n}p_i q_j \ln p_i-\sum_{j=1}^{m}\sum_{i=1}^{n}p_i q_j \ln q_j=$$

$$S_{mn}(\overline{p}\cdot\overline{q})=-\sum_{j=1}^{m}\sum_{i=1}^{n}p_i q_j \ln p_i-\sum_{j=1}^{m}\sum_{i=1}^{n}p_i q_j \ln q_j=$$

$$=-\sum_{i=1}^{n}p_i \ln p_i-\sum_{j=1}^{m}q_j \ln q_j \Rightarrow$$

$$S_{mn}(\overline{p}\cdot\overline{q})=S_n(\overline{p})+S_m(\overline{q})$$

Ιδιότητες μετρικής Shannon(7)

11. Αν $\overline{p}$ και $\overline{q}$ δεν είναι απαραίτητα ανεξάρτητες κατανομές, ορίζουμε:

$P(X=x_i)=p_i$, $P(Y=y_j / X=x_i)=q_{ij}$ και $P(X=x_i, Y=y_j)= p_i\, q_{ij}$.

Τότε $S_{mn}(\overline{p}\cdot\overline{q}) = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_i q_{ij} \ln p_i q_{ij} =$

$$= -\sum_{i=1}^{n}\left(p_i \ln p_i \sum_{j=1}^{m} q_{ij} \right) - \sum_{i=1}^{n}\left(p_i \sum_{j=1}^{m} q_{ij} \ln q_{ij} \right)$$

Όμως, $\sum_{j=1}^{m} q_{ij} = \sum_{j=1}^{m} P\left(Y = y_j / X = x_i\right) = 1$

$\xrightarrow{\text{Συνέχεια}}$

Ιδιότητες μετρικής Shannon(8)

Δηλαδή, $\bar{q} = \left(q_{i1}, q_{i2}, \ldots, q_{im}\right)$ είναι η υπό συνθήκη κατανομή των εξαγόμενων του 2ου πειράματος, όταν το πρώτο έχει αποτελέσματα $X = X_i$.

Άρα, $S_{mn}(\bar{p} \cdot \bar{q}) = S_n(\bar{p}) + \sum_{i=1}^{n} p_i S_m(\bar{q}_i)$

που σημαίνει ότι η εντροπία της κοινής κατανομής είναι ίση με την εντροπία του πρώτου πειράματος συν την αναμενόμενη τιμή της υπό συνθήκη εντροπίας του δεύτερου. Μπορεί να γραφτεί και ως:

$$S_{mn}(\bar{p} \cdot \bar{q}) = S_m(\bar{q}) + \sum_{j=1}^{m} q_j S_n(\bar{p}_j)$$

Ιδιότητες μετρικής Shannon(9)

12. Για κάθε κυρτή συνάρτηση f(x), έχουμε εξ΄ορισμού:

$$f\left(p_1x_1+p_2x_2+\ldots+p_nx_n\right)\leq p_1f\left(x_1\right)+p_2f\left(x_2\right)+\ldots+p_nf\left(x_n\right)$$

Αν θέσουμε $f(x)=x\ln x,\ x_i=q_{ij}$ θα έχουμε

$$\left(\sum_{i=1}^{n}p_iq_{ij}\right)\ln\left(\sum_{i=1}^{n}p_iq_{ij}\right)\leq\sum_{i=1}^{n}p_iq_{ij}\ln q_{ij}$$

Είναι $\displaystyle\sum_{i=1}^{n}p_iq_{ij}=\sum_{i=1}^{n}P\left(X=x_i\right)P\left(Y=y_i/X=x_i\right)=q_j$

Οπότε $\displaystyle q_j\ln q_j\leq\sum_{i=1}^{n}p_iq_{ij}\ln q_{ij}$

Συνέχεια $\rightarrow$

Ιδιότητες μετρικής Shannon(10)

Αθροίζοντας

$$\sum_{j=1}^{n} q_j \ln q_j \le \sum_{i=1}^{n} p_i \sum_{j=1}^{m} p_i q_{ij} \ln q_{ij} \Rightarrow$$

$$-\sum_{i=1}^{n} p_i \sum_{j=1}^{m} p_i q_{ij} \ln q_{ij} \le -\sum_{j=1}^{n} q_j \ln q_j \Rightarrow$$

$$S_m(\overline{q}) \ge \sum_{i=1}^{n} p_i S_m(\overline{q}_i)$$

Χρησιμοποιώντας την προηγούμενη ιδιότητα

$$S_{mn}(\overline{p} \cdot \overline{q}) \le S_n(\overline{p}) + S_m(\overline{q})$$

Ιδιότητες μετρικής Shannon(11)

13. $S_{n-1}(p_1+p_2, p_3, \ldots, p_n) = -(p_1+p_2)\ln(p_1+p_2) - \sum_{i=3}^{n} p_i \ln p_i$

$= -(p_1+p_2)\ln(p_1+p_2) + p_1 \ln p_1 + p_2 \ln p_2 \Rightarrow \sum_{i=1}^{n} p_i \ln p_i$

$\Rightarrow$

$S_n(p_1, p_2, \ldots, p_n) =$

$$= S_{n-1}(p_1+p_2, p_3, \ldots p_n) + (p_1+p_2)\left(-\frac{p_1}{p_1+p_2}\ln\frac{p_1}{p_1+p_2} - \frac{p_2}{p_1+p_2}\ln\frac{p_2}{p_1+p_2}\right)$$

$$= S_{n-1}(p_1+p_2, p_3, \ldots p_n) + (p_1+p_2) S_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$$

Άρα $S_{n-1}(p_1+p_2, p_3, \ldots, p_n) \le S_n(p_1, p_2, \ldots, p_n)$

Δηλαδή, αν συνδυαστούν δύο εξαγόμενα, μειώνεται η εντροπία.

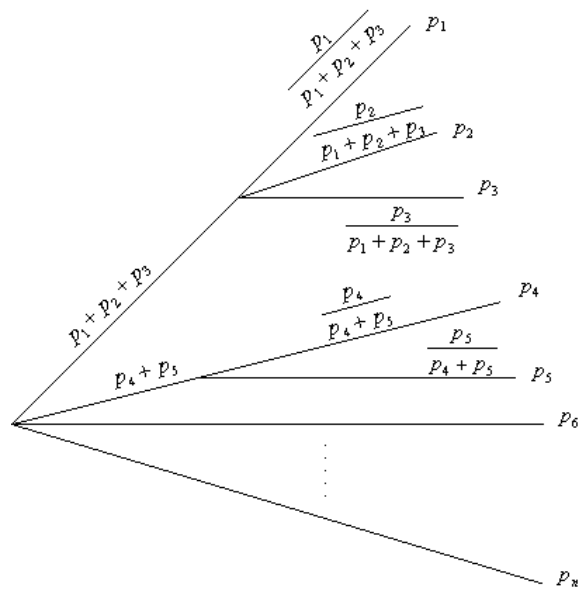
Ιδιότητες μετρικής Shannon (12)

14. Ο υπολογισμός του $S_n(p_1,p_2,\ldots p_n)$ γίνεται σε δύο στάδια:
    - Αρχικά, βρίσκουμε το $S_{n-1}(p_1+p_2,p_3,\ldots p_n)$.
    - Στη συνέχεια, έχουμε το p_1+p_2 διαιρεμένο σε 2 τμήματα p_1 και p_2, ώστε έχουμε την κατανομή $\left[\dfrac{p_1}{p_1+p_2},\dfrac{p_2}{p_1+p_2}\right]$ με εντροπία $S_2\left[\dfrac{p_1}{p_1+p_2},\dfrac{p_2}{p_1+p_2}\right]$

      και βάρος p_1+p_2.

Ιδιότητες μετρικής Shannon (13)

Αυτή είναι η αρχή διακλάδωσης και κάθε πιθανότητα υπολογίζεται όπως φαίνεται στο σχέδιο.



Η Αρχή Μέγιστης Εντροπίας του Jayne-MaxEnt

Ορισμός, εξήγηση

Υπολογισμός μικρο-καταστάσεων με μακροσκοπικές μετρήσεις(1)

- Έστω μία Τ.Μ. X παίρνει τις τιμές $x_1,x_2,\ldots,x_n$ με πιθανότητες $p_1,p_2,\ldots,p_n$. Οι τελευταίες μπορούν χρησιμοποιηθούν για τον υπολογισμό της μέσης τιμής, της διακύμανσης κι άλλων ροπών της X, όπως το $\sum_{i=1}^{n} p_i x_i$
- Γενικά μπορούμε να έχουμε μέσες τιμές των συναρτήσεων $g_1(x),g_2(x),\ldots,g_n(x)$ για να πάρουμε

$$\sum_{i=1}^{n} p_i g_r(x_i) = a^r, \qquad r = 1,2,\ldots,m \quad (1)$$

μαζί με $$\sum_{i=1}^{n} p_i = 1 \quad (2)$$

Υπολογισμός μικρο-καταστάσεων με μακροσκοπικές μετρήσεις(2)

- Συνεπώς, μπορούμε να έχουμε $m+1$ σχέσεις των $p_1,p_2,\ldots,p_n$.
- Αν $m+1<n$, τότε οι (1) και (2) δεν μπορούν να προσδιορίσουν μοναδικά τις $p_1,p_2,\ldots,p_n$.
- Μπορούμε να δώσουμε αυθαίρετες τιμές στα $n-m-1$ p_i και να λύσουμε τις (1) και (2) για τα υπόλοιπα. Γενικά, υπάρχουν άπειρες λύσεις, αλλά δεχόμαστε μόνο αυτές με $p_i \geq 0$.

Ορισμός της αρχής μέγιστης εντροπίας του Jayne

Από όλες τις κατανομές που ικανοποιούν τους περιορισμούς μέσης τιμής (1) και (2)

επέλεξε αυτή που μεγιστοποιεί τη μετρική Shannon.

Δηλαδή:

- Την κατανομή με μέγιστη αβεβαιότητα
- Την κατανομή με ελάχιστη δέσμευση στη μη-γνωστή πληροφορία
- Την πιο τυχαία κατανομή
- Την λιγότερο επηρεασμένη κατανομή

Μαθηματική Εφαρμογή της Αρχής(1)

Μεγιστοποίηση της μετρικής Shannon $-\sum_{i=1}^{n} p_i \ln p_i$

υπό τους περιορισμούς (1),(2)

- Εφαρμόζουμε μέθοδο Lagrange

$$L = -\sum_{i=1}^{n} p_i \ln p_i - (\lambda_0 - 1)\left(\sum_{i=1}^{n} p_i - 1\right) - \sum_{i=1}^{m} \lambda_r \left(\sum_{i=1}^{n} p_i q_{ri} - a_r\right)$$

όπου $\lambda_0, \lambda_1, \ldots, \lambda_m$ οι $m+1$ πολλαπλασιαστές Lagrange

- $$\frac{\partial L}{\partial p_i} = 0 \Rightarrow -\ln p_i - \lambda_0 - \sum_{r=1}^{m} \lambda_r g_{ri} = 0 \Rightarrow$$

$$p_i = e^{\left(-\lambda_0 - \lambda_2 g_{1i} - \ldots - \lambda_m g_{mi}\right)} \qquad i = 1, 2, \ldots, n$$

Μαθηματική Εφαρμογή της Αρχής(2)

- Τα $\lambda_0, \lambda_1, \ldots \lambda_m$ υπολογίζονται με αντικατάσταση του p_i της σχέσης (3) στους περιορισμούς (1) και (2)

$$\sum_{i=1}^{n} e^{\left(-\lambda_0 - \sum_{j=1}^{m} \lambda_j g_{ji}\right)} = 1$$

$$\sum_{i=1}^{n} g_{ri} e^{\left(-\lambda_0 - \sum_{j=1}^{m} \lambda_j g_{ji}\right)} = a_r$$

Μαθηματική Εφαρμογή της Αρχής(2)

- Η μέγιστη εντροπία είναι

$$S_{\max} = -\sum_{i=1}^{n} p_i \ln p_i = -\sum_{i=1}^{n} p_i \left(-\lambda_0 - \lambda_1 g_{1i} - \lambda_2 g_{2i} - \ldots - \lambda_m g_{mi}\right)$$

$$\Rightarrow S_{\max} = \lambda_0 + \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m$$

Εφαρμογή της MaxEnt στην Queueing Theory

Πρόβλημα, Εφαρμογές

Ορισμός προβλήματος

Δίνονται:

- Κατανομή των χρόνων μεταξύ αφίξεων
- Κατανομή των χρόνων εξυπηρέτησης
- Τρόπος πρόσβασης(FCFS, Random Service, LCFS,προτεραιότητες)
- Άλλα δεδομένα(π.χ περιορισμένη ουρά)

Και ζητούνται οι κατανομές των μέτρων απόδοσης:

- Κατανομή μεγέθους ουράς
- Κατανομή χρόνου αναμονής
- Κατανομή busy περιόδου

Πότε εφαρμόζεται η MaxEnt;

- Όταν μας δίνεται λίγη πληροφορία για τις κατανομές(π.χ μερικές ροπές του χρόνου μεταξύ αφίξεων, του χρόνου εξυπηρέτηση, του μεγέθους ουράς, του χρόνου αναμονής)
- Όταν οι κατανομές που προκύπτουν με γνωστούς τρόπους είναι πολύπλοκες, η MaxEnt χρησιμοποιείται για να τις προσεγγίσει με απλούστερες.

<u>ΣΗΜΕΙΩΣΗ:</u> Η γνώση για τις ροπές μπορεί να βρεθεί με απευθείας παρατήρηση της ουράς ή με χρήση αποτελεσμάτων της Queueing Theory.

Αποτελέσματα από MaxEnt(1)

1. Όταν έχουμε άπειρο μήκος ουράς κι η μόνη πληροφορία που έχουμε αφορά το μέσο αριθμό πελατών.

---

p_i: η πιθανότητα να είναι *i* πελάτες στο σύστημα, *i*=0,1,2,…..

m: μέσος αριθμός πελατών

Θέλουμε μεγιστοποίηση της $S = -\sum_{i=0}^{\infty} p_i \ln p_i$ υπό τους περιορισμούς

$$\sum_{i=0}^{\infty} p_i = 1, \quad \sum_{i=0}^{\infty} i p_i = m$$

Εφαρμόζοντας μέθοδο Lagrange καταλήγουμε σε

$p_i = (1-b)\cdot b^i$,όπου $\quad b = \dfrac{m}{1+m}$

Αποτελέσματα από MaxEnt(2)

2. Περιορισμένο μήκος ουράς, μέσος αριθμός πελατών

---

p_i: η πιθανότητα να είναι i πελάτες στο σύστημα, i=0,1,2,…,N

m: μέσος αριθμός πελατών

Η μεγιστοποίηση εντροπίας δίνει

όπου $p_i = ab^i \qquad i = 0,1,2,\ldots,N$

$$a\sum_{i=0}^{N} b^i = 1 \text{ και } a\sum_{i=0}^{N} ib^i = m$$

έτσι ώστε το b δίνεται από τη σχέση

$$\frac{b + 2b^2 + \ldots + Nb^N}{1 + b + b^2 + \ldots + b^N} = m$$

Αποτελέσματα από MaxEnt(3)

3. Άπειρο μήκος ουράς, μέσος αριθμός πελατών, πιθανότητα p_0 άδειο σύστημα

---

p_i: η πιθανότητα να είναι *i* πελάτες στο σύστημα, *i*=1,2,…..

m: μέσος αριθμός πελατών

Θέλουμε μεγιστοποίηση της $S=-\sum_{i=1}^{n} p_i \ln p_i$ υπό τους περιορισμούς

$\sum_{i=1}^{\infty} p_i = 1 - p_0$ και $\sum_{i=1}^{\infty} i p_i = m$

Βρίσκουμε $p_i = ab^i$

όπου $\frac{ab}{1-b} = 1 - p_0, \qquad \frac{ab}{(1-b)^2} = m$

Συνέχεια

$\rightarrow$

Αποτελέσματα από MaxEnt(4)

Τελικά είναι

$$p_i = \frac{(1-p_0)^2}{m-1+p_0} - \left(\frac{m-1+p_0}{m}\right)^i, \qquad i = 1,2,\ldots$$

$$\text{Αν} \quad p_0 = 1-\rho \quad , \rho = \frac{m}{1+m}$$

$$p_i = \rho(1-\rho)^i$$

Αποτελέσματα από MaxEnt(5)

4. Κλειστό δίκτυο ουρών, γνωρίζουμε μέσο αριθμό πελατών σε κάθε ουρά

m: αριθμός ουρών

N: συνολικός αριθμός πελατών

Q_i: μέσος αριθμός πελατών στις ουρές, i=1,2,…,m

$p(n_1,n_2,\ldots,n_m)$: η πιθανότητα να είναι $n_1,n_2,\ldots,n_m$ πελάτες στις m ουρές
Θέλουμε μεγιστοποίηση του

$$S = -\sum_{n_1+n_2+\ldots+n_m=N} p(n_1,n_2,\ldots,n_m)\ln p(n_1,n_2,\ldots,n_m)$$

υπό τους περιορισμούς

$$\sum_{n_1+n_2+\ldots+n_m=N} p(n_1,n_2,\ldots,n_m)=1 \text{ και } \sum_{n_1+n_2+\ldots+n_m=N} n_i\, p(n_1,n_2,\ldots,n_m)=Q_i$$

Συνέχεια →

Αποτελέσματα από MaxEnt(6)

- Με τη μέθοδο Lagrange παίρνουμε

$$p(n_1, n_2, \ldots n_m) = \frac{x_1^{\,n_1} x_2^{\,n_2} \cdots x_m^{\,n_m}}{\sum x_1^{\,n_1} x_2^{\,n_2} \cdots x_m^{\,n_m}}$$

όπου $x_1, x_2, \ldots x_m$ είναι οι πολλαπλασιαστές Lagrange και υπολογίζονται από τους περιορισμούς

Η κατανομή αυτή είναι product-form πιθανοτική κατανομή. Η οριακή κατανομή για την κατανομή του μεγέθους ουράς της πρώτης ουράς είναι:

$$p(n_1) = x_1^{\,n_1} \frac{\sum\limits_{n_2+n_3+\ldots+n_m=N-n_1} x_2^{\,n_2} x_3^{\,n_3} \cdots x_m^{\,n_m}}{\sum\limits_{n_2+n_3+\ldots+n_m=N} x_2^{\,n_2} x_3^{\,n_3} \cdots x_m^{\,n_m}}, \qquad n_1 = 0,1,2,\ldots,N$$

Extended MaxEnt

Εφαρμογές

Επιπλέον εφαρμογές MaxEnt

- Η αρχή MaxEnt εφαρμόζεται και σε αναλογίες
  - Π.χ, αν $x_1,x_2,\ldots x_n$ μη-αρνητικοί αριθμοί, τότε με $p_i = \dfrac{x_i}{x_1 + x_2 + \cdots + x_n}$ και $\sum_{i=1}^{n} p_i = 1$ και $p_i \geq 0$ το $S = -\sum_{i=1}^{n} p_i \ln p_i$ αποτελεί ένα μέτρο ισότητας των x_i.
- Πεδία εφαρμογής
  - Επιχειρήσεις
  - Οικονομία
  - Αστικός Σχεδιασμός
  - Ανάλυση γλωσσών

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης