

1η Σειρά Ασκήσεων (Αξιολόγηση της Αποτελεσματικότητας της Ανάκτησης)

Ανάθεση: 4 Μαρτίου

Παράδοση: 18 Μαρτίου (2 εβδομάδες)

Άσκηση 1 (4 βαθμοί)

Θεωρείστε μια συλλογή αξιολόγησης που αποτελείται από 20 έγγραφα $\{d_1, \dots, d_{20}\}$. Η συλλογή αξιολόγησης περιλαμβάνει μια επερώτηση q για την οποία γνωρίζουμε ότι τα έγγραφα της συλλογής που είναι συναφή με αυτήν είναι 5, συγκεκριμένα τα $\{d_1, d_{11}, d_{18}, d_{19}, d_{20}\}$. Θέλουμε να αξιολογήσουμε την αποτελεσματικότητα τριών συστημάτων $S1, S2$ και $S3$. Για το λόγο αυτό υποβάλλουμε σε κάθε σύστημα την επερώτηση q και λαμβάνουμε τις εξής απαντήσεις:

$$Ans(S1, q) = \langle d_{11}, d_4, d_{18}, d_2, d_9, d_7, d_8, d_6, d_1, d_{20} \rangle$$

$$Ans(S2, q) = \langle d_9, d_7, d_5, d_6, d_{11}, d_4, d_8, d_2, d_1, d_{18}, d_{19}, d_{20} \rangle$$

$$Ans(S3, q) = \langle d_{18}, d_{11}, d_1, d_5, d_2 \rangle$$

Το αριστερότερο στοιχείο της κάθε απάντησης παριστάνει το υψηλότερα διαβαθμισμένο έγγραφο, αυτό που το σύστημα υπολόγισε ως το πιο συναφές με την επερώτηση q . Συγκρίνετε τα τρία αυτά συστήματα ως προς τα εξής μέτρα: (α) Ακρίβεια (Precision), (β) Ανάκληση (Recall), (γ) F-Measure, (δ) R-Ακρίβεια (R-Precision) και (ε) Fallout. Δώστε τα συγκεντρωτικά στοιχεία σε ένα πίνακα της μορφής:

System	Precision	Recall	F-Measure	R-Precision	Fallout
S1					
S2					
S3					
Ranking of systems					

Σχολιάστε το αποτέλεσμα της σύγκρισης.

Άσκηση 2 (4 βαθμοί) 2007

Σχεδιάστε τις καμπύλες ακρίβειας/ανάκλησης (P/R curves) των συστημάτων της προηγούμενης άσκησης. Για κάθε σύστημα δώστε 2 γραφήματα: ένα που να απεικονίζει τα P/R σημεία όπως προκύπτουν από τις απαντήσεις, και ένα χρησιμοποιώντας κανονικοποιημένα επίπεδα ανάκλησης (standard recall levels). Αν βλέπατε μόνο αυτά τα γραφήματα (και όχι τις απαντήσεις) θα μπορούσατε να επιλέξετε το καλύτερο σύστημα ;

Ένας εναλλακτικός τρόπος αξιολόγησης της αποτελεσματικότητας ενός, είναι οι καμπύλες Recall-Fallout. Ορίζονται ανάλογα με τις καμπύλες Precision-Recall, μόνο που τώρα ο άξονας Q έχει τιμές του Fallout, ενώ ο Y τιμές του Recall . Σχεδιάστε τις καμπύλες Recall-Fallout των συστημάτων της προηγούμενης άσκησης.

Άσκηση 3 (2 βαθμοί)

Θεωρείστε ότι έχουμε μια συλλογή αξιολόγησης που αποτελείται από ένα σύνολο εγγράφων V και ένα σύνολο επερωτήσεων Q . Για κάθε ζεύγος $q, v_i \in Q \times V$, έχει οριστεί το $label_q(v_i)$, το οποίο εκφράζει τη συνάφεια του v_i ως προς το q σε μια κλίμακα από το 0 έως το 5 (1,2,3,4,5) όπου 0=χειρότερο,..., 5=καλύτερο. Έστω ότι θέλουμε να αξιολογήσουμε την απόκριση ενός συστήματος σε μια επερώτηση q . Έστω ότι τα κορυφαία k στοιχεία της απόκρισης είναι το διατεταγμένο σύνολο $\langle v_1, \dots, v_k \rangle$. Για την αξιολόγηση του έχουμε στη διάθεση μας τα εξής μέτρα:

(α)

$$M_1(k) = \sum_{i=1}^k (2^{label_q(v_i)} - 1)$$

(β)

$$M_2(k) = \sum_{i=1}^k \frac{1}{\log_2(i+1)} (\text{label}_q(v_i) - 1)$$

(γ)

$$M_3(k) = \sum_{i=1}^k \frac{1}{\log_2(i+1)} (2^{\text{label}_q(v_i)} - 1)$$

Σχολιάστε σύντομα το κάθε ένα. Συγκρίνετε τα και δικαιολογήστε ποιό θα χρησιμοποιούσατε. Στηρίξτε τις απόψεις δίνοντας μικρά παραδείγματα.

ΠΡΟΑΙΡΕΤΙΚΗ (Εξάσκησης/Bonus)

Άσκηση 4 (2 βαθμοί)

Έστω ότι η συλλογή αξιολόγησης αποτελείται από 100 έγγραφα $\{d_1, \dots, d_{100}\}$ και γνωρίζουμε ότι υπάρχουν 3 έγγραφα της συλλογής, συγκεκριμένα τα $\{d_1, d_2, d_3\}$, που είναι συναφή με την επερώτηση q . Θέλουμε να αξιολογήσουμε την αποτελεσματικότητα τριών συστημάτων $S1, S2$ και $S3$ τα οποία επιστρέφουν ως απάντηση έγγραφα συνοδευμένα από ένα βαθμό συνάφειας. Υποβάλλουμε σε κάθε σύστημα την επερώτηση q και λαμβάνουμε τις εξής απαντήσεις:

$$\text{Ans}(S1, q) = \langle d_1, \{d_2, d_{20} - d_{40}\}, d_3 \rangle$$

$$\text{Ans}(S2, q) = \langle d_1, d_2, d_3 \rangle$$

$$\text{Ans}(S3, q) = \langle \{d_1, d_8\}, d_2, d_3 \rangle$$

Η απάντηση $\langle \{d_1, d_8\}, d_2, d_3 \rangle$ σημαίνει ότι τα d_1, d_8 ισοβαθμούν στην πρώτη θέση (άρα έλαβαν τον μεγαλύτερο βαθμό συνάφειας). Η απάντηση $\langle d_1, \{d_2, d_{20} - d_{40}\}, d_3 \rangle$ σημαίνει ότι το d_1 έλαβε το μεγαλύτερο βαθμό, ενώ μετά ακολουθεί μια ομάδα από 22 έγγραφα τα οποία ισοβαθμούν, και στο τέλος της κατάταξης βρίσκεται το d_3 . Για κάθε ένα από τα 3 συστήματα απαντήστε τα ακόλουθα ερωτήματα: (α) Ποια είναι η R -ακρίβεια (R-precision); (β) Ποιο είναι το αναμενόμενο μήκος αναζήτησης για να βρούμε 2 συναφή;