

Η χρήση της μεθόδου ΠΑ στη Βιοτεχνολογία

Μαρία Ντάγγαλου

15 Ιουνίου 2025

Περίληψη

Η Ανάλυση Κύριων Συνιστωσών (Πρινσιπαλ δμπονετ Α-ναλψις - ΠΑ) αποτελεί μια απο τις πιο ευρέως χρησιμοποιούμενες τεχνικές μείωσης διαστάσεων. Στην εργασία αυτή εξετάζεται η χρήση της ΠΑ σε εφαρμογές της βιοτεχνολογίας, με έμφαση στην ανάλυση γονιδιακής έκφρασης και δεδομένων από μεθόδους αλληλούχισης.

1 Εισαγωγή

Η μέθοδος ΠΑ είναι μία τεχνική στατιστικής ανάλυσης που χρησιμοποιείται για να μειώσει τη διάσταση δεδομένων, διατηρώντας όσο το δυνατόν περισσότερη πληροφορία. Η εφαρμογή της στη βιοτεχνολογία έχει καταστεί ιδιαίτερα σημαντική, ιδίως στην ανάλυση δεδομένων υψηλής διαστατικότητας όπως αυτά που προκύπτουν απο τεχνολογίες ΝΓΣ (Νεξτ-Γενερατιον Σεχυνςινγ) ή μικροσυστοιχίες. Στόχος της παρούσας εργασίας είναι να παρουσιαστεί συνοπτικά η μέθοδος ΠΑ, η εφαρμογή της στον βιοτεχνολογικό χώρο, καθώς και ένα θεωρητικό παράδειγμα.

2 Βιβλιογραφική Ανασκόπηση

Η ΠΑ εισήχθη από τον Καρλ Περσον το 1901 [;], και επεκτιτάθηκε από τον Χαρολδ Ηοτελλινγ τη δεκαετία του 1930 [;]. Η μέθοδος έχει χρησιμοποιηθεί σε ευρύ φάσμα επιστημών, από την ανάλυση της εικόνας έως την οικονομία.

Στη βιοτεχνολογία, χρησιμοποιείται για την οπτικοποίηση και ταξινόμηση δειγμάτων, όπως στην ανάλυση ΡΝΑ-σεχ [;], στον εντοπισμό βιοδεικτών [;], ή στην διαφοροποίηση κυτταρικών πληθυσμών σε σινγλε-σελλ δεδομένα [;].

3 Μαθηματική Περιγραφή της Μεθόδου

Έστω ότι έχουμε ένα σύνολο δεδομένων $X \in \mathbb{R}^{n \times p}$, όπου n δείγματα και p μεταβλητές. Η μέθοδος ΠΑ περιλαμβάνει:

1. Κεντράρισμα των δεδομένων

$$\tilde{X} = x - \mu, \quad \mu = \frac{1}{n} \sum_{i=1}^n X_i$$

2. Υπολογισμό του πίνακα συνδιακύμανσης:

$$S = \frac{1}{n-1} \tilde{X}^T \tilde{X}$$

3. Ιδιοδιάσπαση του S : Εύρεση ιδιοτιμών και ιδιοδιανυσμάτων

$$Sv_i = \lambda_i v_i$$

4. Επιλογή των πρώτων k ιδιοδιανυσμάτων V_K , ώστε η προβολή είναι:

$$Z = \tilde{X} V_K$$

Θεωρητικό Παράδειγμα

Έστω δεδομένα γονιδιακής έκφρασης από 100 δείγματα με 1000 γονίδια. Η εφαρμογή ΠΑ επιτρέπει την προβολή των δεδομένων σε επίπεδο 2 διαστάσεων, όπου τα δείγματα διαχωρίζονται σε π.χ. σε ασθενείς και υγιείς, αποκαλύπτοντας υποκείμενες δομές.

4 Συμπεράσματα

Η ΠΑ είναι ένα ισχυρό εργαλείο στην ανάλυση βιοτεχνολογικών δεδομένων. Επιτρέπει τη μείωση της διαστατικότητας, διατηρώντας σημαντική πληροφορία, διευκολύνοντας την ερμηνεία πολύπλοκων δεδομένων. Η ευρεία εφαρμογή της στην βιοτεχνολογία αποδεικνύει τη συνεισφορά των μαθηματικών εργαλείων σε επιστημονικά πεδία αιχμής.