

АНАЛІТИЧНІ ЗАЛЕЖНОСТІ, ЩО ЗВ'ЯЗУЮТЬ ФУНКЦІЇ СХОЖОСТІ ДЛЯ БІНАРНИХ ОБ'ЄКТІВ І ВІДСТАНЬ ХЕММІНГА

*д-р техн. наук, проф. В.Д. Дмитрієнко, д-р техн. наук,
проф. О.Ю. Заковоротний, д-р техн. наук, проф. С.Ю. Леонов,
Національний технічний університет "Харківський політехнічний
інститут", м. Харків*

Оскільки функції схожості для бінарних векторів (об'єктів) та для об'єктів (векторів), описуємих за допомогою біполярного алфавіту, використовуються для оцінки подібності двійкових об'єктів, то повинні бути залежності, що зв'язують функції подібності для бінарних об'єктів (векторів) і відстані Хеммінга R_X . В табл. 1 наведені вищезазначені класичні функції афінності (перший стовпець таблиці) і функції подібності, модифіковані за допомогою відстані Хеммінга (другий стовпець табл. 1).

В табл. 1 за допомогою змінних R_X , a і b підраховується число розрядів з співпадаючими одиничними або нульовими компонентами у порівнюваних векторів. За допомогою змінних f і g підраховуються одиничні компоненти, які є у одного вектора J_1 і відсутні у другого вектора J_2 .

Таблиця 1 – Взаємозв'язок між відстанню Хеммінга і функціями подібності для бінарних об'єктів

Класичні функції подібності	Функції подібності з використанням відстані Хеммінга	Відстань Хеммінга, виражене через бінарні функції подібності
1	2	3
Функція подібності Russel and Rao $S_1 = \frac{a}{a+b+f+g}$	$S_1 = \frac{a}{a+b+R_X}$	$R_X = \frac{a}{S_1} - (a+b)$
Функція подібності Sokal and Michener $S_2 = \frac{a+b}{a+b+f+g}$	$S_2 = \frac{a+b}{a+b+R_X}$	$R_X = \frac{a+b}{S_2} - (a+b)$
Функція подібності Jaccard and Needham $S_3 = \frac{a}{a+f+g}$	$S_3 = \frac{a}{a+R_X}$	$R_X = \frac{a}{S_3} - a$

1	2	3
Функція подібності Kulzinsky $S_4 = \frac{a}{f + g}$	$S_4 = \frac{a}{R_X}$	$R_X = \frac{a}{S_4}$
Функція подібності Dice $S_5 = \frac{a}{2a + f + g}$	$S_5 = \frac{a}{2a + R_X}$	$R_X = \frac{a}{S_5} - 2a$
Функція подібності Yula $S_6 = \frac{ab - fg}{ab + fg}$	Громіздкий вираз	$R_X = f + g$
Коефіцієнт кореляції (correlation) $S_7 = \frac{ab + fg}{[(a + f)(a + g)(b + g)(b + f)]^{1/2}}$	Громіздкий вираз	$R_X = f + g$

У третьому стовпці табл. 1 вперше приведені прості аналітичні залежності для визначення відстані Хеммінга через класичні функції подібності Russel and Rao, Sokal and Michener, Jaccard and Needham, Kulzinsky, Dice. Отримати прості аналітичні залежності, що зв'язують функцію подібності Yule S_6 і коефіцієнт кореляції S_7 не вдалося. Але визначити відстань Хеммінга за функціями S_6 і S_7 можна за допомогою співвідношення $R_X = f + g$.