

Annexe O

Indice de Jaccard

L'indice¹ de P. Jaccard compare la similarité entre des échantillons. Il s'agit Paul Jaccard (1868-1944) du rapport entre le cardinal de l'intersection des ensembles considérés et le cardinal de l'union des ensembles [Jaccard, 1901]. Soient A et B , l'indice est :

$$J(A, B) = \frac{\text{card}(A \cap B)}{\text{card}(A \cup B)} \quad (\text{O.1})$$

L'extension à n ensembles est :

$$J(S_1, S_2, \dots, S_n) = \frac{\text{card}(S_1 \cap S_2 \cap \dots \cap S_n)}{\text{card}(S_1 \cup S_2 \cup \dots \cup S_n)} \quad (\text{O.2})$$

La **distance de Jaccard** mesure la dissimilarité entre les ensembles. Elle consiste à soustraire l'indice de Jaccard à 1.

$$d = 1 - J(A, B) = 1 - \frac{\text{card}(A \cap B)}{\text{card}(A \cup B)} = \frac{\text{card}(A \cup B) - \text{card}(A \cap B)}{\text{card}(A \cup B)} \quad (\text{O.3})$$

L'extension à n ensembles est :

$$d = 1 - J(S_1, S_2, \dots, S_n) = \frac{\text{card}(S_1 \cap S_2 \cap \dots \cap S_n) - \text{card}(S_1 \cup S_2 \cup \dots \cup S_n)}{\text{card}(S_1 \cup S_2 \cup \dots \cup S_n)} \quad (\text{O.4})$$

1. ou le coefficient de Jaccard ou le coefficient de communauté

Bibliographie

[Jaccard, 1901] JACCARD, P. (1901). Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. Bulletin de la société vaudoise des sciences naturelles, 37:241–272.