

There are two important characteristics, which require a discussion.

First, if we assess the first two terms, we use $cost(x, 0)$ to denote the maximal distance of inserting and deleting x . $cost(x, 0)$ can be read as cost between x and a null-vector of the same size. However, it is noteworthy to state that this interpretation does not hold for any arbitrary cost-function. For instance, the cosine-distance does not work with a null vector, as it is impossible to compute the angle between x and a null vector. Here, the maximum distance would just amount to 1. In contrast, the family of Minkowsky distance works well with this notion, because they compute a distance between two points and not two directions.

Second, to explain the intuition behind most terms, it is necessary to establish a common understanding between the relationship of an event with its event attributes. Generally, we can have two notions of this relationship.

For the first relationship, we consider the event and its attributes as separate entities. This notion is reasonable, as some attributes remain static throughout the whole process run. If we take a loan application process as an example, an applicant's ethnic background will not change regardless of the event. This characteristic can be considered a case attribute, which remains static throughout the process run. This understanding would require us to modify the cost functions, as they treat the activity independently from its attribute values. In other words, if the activities of two events are \bar{a} and \bar{b} , but their attribute values are $(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix})$ and $(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix})$, these events may be seen as more similar than two \bar{a} and \bar{a} with attribute values $(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix})$ and $(\begin{smallmatrix} 5 \\ 0 \end{smallmatrix})$.

A second notion would treat each event as an independent and atomic point in time. Hence, a and b would be considered completely different even if their event attributes are the same. This understanding is also a valid proposition, as you could argue that an event which occurs at nighttime is not the same event as an event at daytime. Here, the time domain is the main driver of distinction and the content remains a secondary actor.

All the terms described in the Semi-structured Damerau-Levenshtein distance (SSDLD) follow the second notion. There are two reasons for this decision. First, treating event activities and event attributes separately would further complicate the SSDLD, as we would have to expand the cost structure. Second, the unmodified Damerau-Levenshtein distance applies to discrete sequences, such as textual data with atomic words or characters. By treating each event as an discrete sequence element, we remain faithful to the original function.