Multi Criteria Quality Assurance for Transit Connections

August 16, 2023

1 Introduction

This documents describes how to compare different transit connections and sets of transit connections regarding different criteria. The set of criteria used here are only used as an example and can be replaced or extended to fit a specific use case.

In general we want that the following properties apply to our rating function:

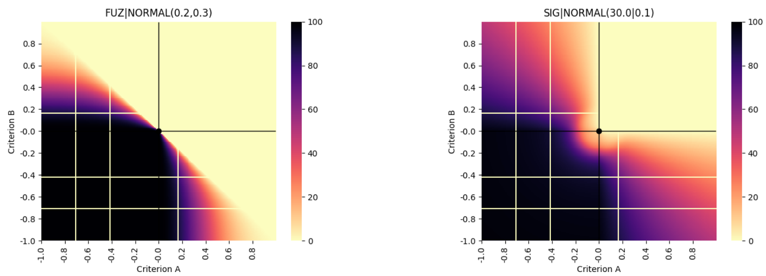
1. The rating function should be symmetric in the sense that $r(A, B) = -1 \cdot r(B, A)$ where A and B are sets of connections and r is rating the improvement of A over B.

2. In case we add duplicates of connections to either set A or B, the function value of r should stay the same. In case of adding a connection to either set A or B that is very similar to one or more of the connections already contained in those sets, the change in function value of r should be small.

Delling et al. present a measure to compare Pareto optimal connections [1]. Their measure is based on fuzzy-logic [2]. This enables them to extend Pareto-dominance from a Boolean decision where a connection either dominates another connection to a spectrum between zero and one where 0. In this work, we build on this idea of a more detailed function to compare connections, introduce an improved measure and extend its use to from connections to connection sets.

2 Multi Criteria Earliest Arrival Query

Many customers use a mix of different criteria to make a choice for a specific transit connection. The most common criteria are travel time and number of transfers. A connection can therefore be described as a vector of criteria $x = (x_1, \ldots, x_k)$. Each customer has a different weighting function for each of those criteria and selects the best connection (i.e. the minimum) regarding the weighted sum of criteria. For a customer with weights $c = (c_1, \ldots, c_k)$, this



(a) Fuzzy Measure (b) Sigmoid Measure

Figure 1: Fuzzy measure (left) and sigmoid measure (right) plot comparing connection the connection at the center (0, 0) with other connection having criteria (x, y). The color encodes the degree to which the connection at the center is dominating the connection at (x, y) which has criteria x and y.

introduces the weighting function $f_c(x) = c \cdot x$. Considering a customer with weights c would choose the connection where $f_c(x)$ produces the smallest value. Therefore, for a connection set X, this customer would select $\text{argmin}_x f_c(x), x \in X$.

Based on this approach, it is possible to compare not just two connections, but also connection sets. For a single customer with a specific criteria weighting c and two connection sets A and B, we can compare the value of $f_c(a_{\min})$ with $f_c(b_{\min})$ where $a_{\min} = \text{argmin}_x f_c(x), x \in A$ and $b_{\min} = \text{argmin}_x f_c(x), x \in B$. Since a single customer weighting c is not representative for all users, we can add more weighting functions $c_1, \dots, c_n$ and add the sums of $\sum_{c=c_1,\dots c_n} \text{argmin}_x f_c(x), x \in A$ (same for B) and compare the resulting sums to see which connection is better.

3 Multi Criteria Range Query

3.1 Quality Measures

In [1], Delling et al. a *the degree of domination* between journey J_1 and J_2 as $d(J1, J2) \in [0, 1]$ as $2n_b + n_e - M = 1 - \frac{n_w}{n_b}$ where M is the number of optimization criteria, n_e the criteria where J_1 and J_2 are equal, n_b where J_1 is better than J_2 and n_w where J_1 is worse than J_2. They use $\mu_= := \exp\left(\frac{ln(\chi)}{\epsilon^2} x^2\right)$. As we can see in Figure 1 the Fuzzy measure does not allow for distinguishing between clearly dominated journeys (upper right quadrant) and non-dominated Pareto optimal journeys in the upper right triangle of Quadrants 2 and 4.

To overcome this shortcoming of the fuzzy measure, we introduce a new measure that is specifically tailored to our use case of rating the additional value one journey has over another journey. A simple approach would be to compare the Euclidean distance to the improvement where improvement is measured as the Euclidean distance but only in dimensions where criteria are improved. This results in $Q = I^2/d = I\sin(\alpha)$ where α is the angle between the center and x, y. Since all connections on the same angle will have the same Q value, this is not sufficient to allow more distant connections to have higher improvement values. Therefore, we introduce the following measure:

$$I(x,y,c) = \sqrt{\sum_{\substack{i=0,\dots,k\\ x_i<y_i}} (c_i \cdot x_i - c_i \cdot y_i)^2}$$

$$Q_{\mathrm{sig}}(x,y,c) = \log_2 \frac{I(x,y,c)^2}{d(x,y,c)} \cdot (\arctan(p(d(x,y,c)-q)) + \frac{\pi}{2})$$

Here, p and q are parameters similar to the fuzzy measure. The properties can be derived from Figure 1.

3.2 Range Query

For the multi criteria range query, we need a more sophisticated comparison. Consider the following example of two connection sets $A = \{(10{:}00, 11{:}00, 2)\}$ and $A = \{(10{:}00, 11{:}00, 2), (11{:}00, 12{:}00, 2)\}$ where each connection is given as a tuple of departure time, arrival time and the number of transfers. Here, the weighted sum introduced in Section 2 will rate both connection sets equally good because it does not consider departure and arrival times. However, it is obvious that B is clearly better since it gives the customer an additional option.

In case of a range query, the routing optimization has actually three Pareto citeria: it maximizes departure time (later is better), minimizes arrival time (earlier is better) and minimizes the number transfers. Therefore, a journey j can be described as three dimensional vector $x = (x_1, x_2, x_3)$ where x_1 is the negative departure time of journey $-t_{\mathrm{dep}}(j)$, x_2 is the arrival time $t_{\mathrm{arr}}(j)$ and the x_3 is the number of transfers.

We propose the following function to compute the connection-to-connection improvement ("cc") $I_{\mathrm{cc}}(x,y,c)$ of connection x over another connection y. Note that this is the basically the Euclidean distance between x and y, only taking into consideration dimensions i of x where x_i improves a criterion over y_i.

$$I_{\mathrm{cc}}(x,y,c) = Q_{\mathrm{sig}}(x,y,c)$$

Now, we can define the connection-to-set improvement ("cs") improvement I_{cs} of a connection x over a set of connections A as the minimum improvement over any connection:

$$I_{\mathrm{cs}}(x,A,c) = \operatorname*{argmin}_{a\in A} I_{\mathrm{cc}}(x,a,c)$$

Based on this definition, we can then define the set-to-set ("ss") improvement I_{ss} of a connection set A over a connection set B:

Algorithm 1: Computation of set-to-set improvement I_{ss}

Input: A, B, c // `Connection sets` A, B `and criteria weights` c

Function $\mathtt{I}_{\mathtt{ss}}(A, B, c)$**:**

    $I \longleftarrow 0$

    while $A \neq \emptyset$ **do**

        $a_{\mathtt{min}} \longleftarrow \text{argmax}_{a \in A} I_{\mathtt{cs}}(a, B, c)$

        $I \longleftarrow I + I_{\mathtt{cs}}(a_{\mathtt{min}}, B, c)$

        $A = A \setminus a_{\mathtt{min}}$

        $B = B \cup \{a_{\mathtt{min}}\}$

    return I

The algorithm matches each connection in set A with the connection in B where the improvement is maximal. Each matched connection is moved from A to B once it is matched to make sure Condition 2 is met. To also fulfill Condition 1, we use $I = I_{\text{ss}}(A, B, c) - I_{\text{ss}}(B, A, c)$ to make the measure symmetric around zero.

References

[1] Daniel Delling et al. "Computing multimodal journeys in practice". In: *International Symposium on Experimental Algorithms*. Springer. 2013, pp. 260–271.

[2] Lotfi A Zadeh. "Fuzzy logic". In: *Computer* 21.4 (1988), pp. 83–93.