How are the dots scored and what is the codeword score?

Scoring begins by first grouping spots together that fall within a given search radius for each barcoding round at a given seed reference. Individual spots are assigned scores based on proximity to a reference spot, along with brightness and size of the dot compared to others within the search radius. A score table will be generated which has distance as the highest weight, followed by brightness then size. The range of the score table will always be the same, however, the length of the score table will be dependant on the number of dots being compared + 1 (1). The max score a dot can have will be 2.0, if the dot was determined to be the closest in distance to reference, brightest, and largest in size.

$$\begin{aligned} distance_{score} = x \in A; A = \{x \mid x \in Q, 0 \leq x \leq 1\} and\ n(A) = n(N) + 1, \\ where\ N\ is\ total\ number\ of\ dots\ being\ compared. \end{aligned} \tag{1}$$

$$\begin{aligned} brightness_{score} = x \in A; A = \{x \mid x \in Q, 0 \leq x \leq 0.75\}\ and\ n(A) = n(N) + 1, \\ where\ N\ is\ total\ number\ of\ dots\ being\ compared. \end{aligned}$$

$$\begin{aligned} size_{score} = x \in A; A = \{x \mid x \in Q, 0 \leq x \leq 0.25\}\ and\ n(A) = n(N) + 1, \\ where\ N\ is\ total\ number\ of\ dots\ being\ compared. \end{aligned}$$

Once the highest scoring dot is picked for each barcoding round at a given seed, the sum of scores for each dot will be the codeword score (2). A codeword that had no other neighbors will have a max score of 2 x number of readout sites. Additionally, the ambiguity score will also be calculated as the total number of spots being compared minus the expected number of spots with or without dropouts (3).

$$codeword\ score = \sum_{i=1}^{i=\#\ of\ dots} (distance_{score\ i} + brightness_{score\ i} + size_{score\ i}) \tag{2}$$

$$ambiguity\ score = n(dots\ compared) - n(expected\ dots) \tag{3}$$

Although the total codeword score will be dependent on the number of additional neighbors, they will be adjusted by the ambiguity score in case a specific spot for each round had the highest possible score resulting in a max codeword score even though it had many other spot choices (4).

$$ambiguity\ adjusted\ codeword\ score = \frac{codeword\ score}{ambiguity\ score} \tag{4}$$

After spot picking, the probabilities for each spot being an On barcode genereated by the SVM model is obtained. The sum of ln(probabilities) for dots picked will be the overall ln(probability) of codeword (5). The ln(probability) of codeword will then be adjusted by total codeword score where if the total codeword score is high, then the ln(probability) of codeword score will be closer to 0 (6).

$$\sum_{i=1}^{i=\#\ of\ dots} ln(spot\ probabilities_i) = \prod_{i=1}^{i=\#\ of\ dots} (spot\ probabilities_i) \tag{5}$$

$$ln(final\ score) = \frac{\sum ln(spot\ probabilities)}{ambiguity\ adjusted\ codeword\ score} \tag{6}$$

The ln(final score) can then be exponentiated to obtain a score between 0-1 (7). The closer the value is to 1, the higher confidence we will have in the decoded codeword. If probabilities cannot be generated due to low training data, then the overall adjusted codeword score will be scaled from 0-1 by dividing the value by the max achievable codeword score (7). The final score will essentially combine dot features, crowdedness, and distances.

$$final\ score = e^{ln(final\ score)}\ or\ final\ score = \frac{ambiguity\ adjusted\ codeword\ score}{max\ dot\ score\ *\ \#\ of\ readout\ sites} \tag{7}$$