Haonan Hu

3/25/2021

2863545

If we label our sets of features:

A = {z_1, SL, SW, PL, PW}

B = {z_1, z_2, SW, PL, PW}

C = {z_1, z_2, z_3, SW, PL}

D = {z_1, z_2, z_3, z_4, SW}

E = {z_1, z_2, z_3, z_4,SL}

S = { z_1, z_2, z_3, z_4, SL, SW, PL, PW}

1. What is A ∪ E?

   {z_1, SL, SW, PL, PW, z_2, z_3, z_4}

2. What is D ∩ E?

   {z_1, z_2, z_3, z_4}

3. Create a mutation of C by adding another feature.

   C = {z_1, z_2, z_3, SW, PL}, s = z_4, C' = C, s = {z_1, z_2, z_3, z_4, SW, PL}

4. Create a mutation of C by deleting one feature.

   C = {z_1, z_2, z_3, SW, PL}, C' = C – {PL} = {z_1, z_2, z_3, SW}

5. Create a mutation of C by replacing one feature with another.

   $C = \{z_1, z_2, z_3, SW, PL\}$, $s = \{z_4\}$, $C' = C - \{PL\} \cup s = \{ z_1, z_2, z_3, SW, z_4\}$

6. Assuming each k-fold cross-validation takes 0.1 sec and the rest of the GA processing time is negligible, how long will it take to evaluate one generation?

   5 selected sets for new generation, 20 cross over sets from 5 new generation sets, 5 mutations of the ones selected for new generation, 20 mutations of the new generation cross over sets, with total of 50 sets to eval under k-fold, time = 50 * 0.1 = 5s

7. With the same assumption from 6, but if we start with 6 sets of features and select the 6 best for each new generation, how long will it take to evaluate one generation?

   6 selected sets for new generation, 30 cross over sets from 6 new generation sets, 6 mutations of the ones selected for new generation, 30 mutations of the new generation cross over sets, with total of 72 sets to eval under k-fold, time = 72 * 0.1 = 7.2s