Supplementary Material: Running Example of Code Change Sniffer Method

Ekincan Ufuktepe, Tugkan Tuglular

May 25, 2021

1 Running Example

The probabilistic information obtained from forward slicing are encoded into Markov chain's edges along with the change information based on the type of the model, namely call graph (CG) and effect graph (EG). On the other hand, initial vector is encoded with change information, that applies for both models. Starting with encoding the edges, we construct a transition matrix, which is similar to an adjacency matrix.

Another property of Markov chain is that summation of the outgoing edge probabilities of a node should be equal to 1. Therefore, the probability summation of each row in transition matrix should be equal to 1. However, a row summation could be less than or greater than 1 depending on the probabilities obtained from forward slicing. For instance, on the left side of Fig. 1, let us assume that we encode the Markov chain model with probabilistic information with forward slicing and change information. We can see that some of the nodes' summation of outgoing edges are less than 1 or greater than 1. To satisfy the properties of Markov chain, we weight each node's outgoing edges, by dividing the summation of outgoing edges to each outgoing edge of that node. On the right-hand side of Fig. 1, we obtain the updated Markov chain after weighting the edges. Furthermore, assume that the filled methods (m_0, m_1, m_2, m_3, m_4) are the changed methods.

After the weighting process is completed, we construct the transition matrix of the Markov chain model below. According to the graph model in Fig. 1, there is no outgoing edge from methods m_2, m_3, and m_4. Therefore, in the transition matrix we would expect to have the entire row with filled with 0s. However, we have a single 1 that are placed to itself such as $m_2 \to m_2$, $m_3 \to m_3$, $m_4 \to m_4$. According to Markov chain's properties the summation of the columns for each row should be equal to 1. Thereby, for a row where the sum of column values is equal to 0, we set the $m_i \to m_i$ edge probability to 1. If the method m_i is not change, setting the probability will not affect the overall impact calculation, since that it will be multiplied with 0.

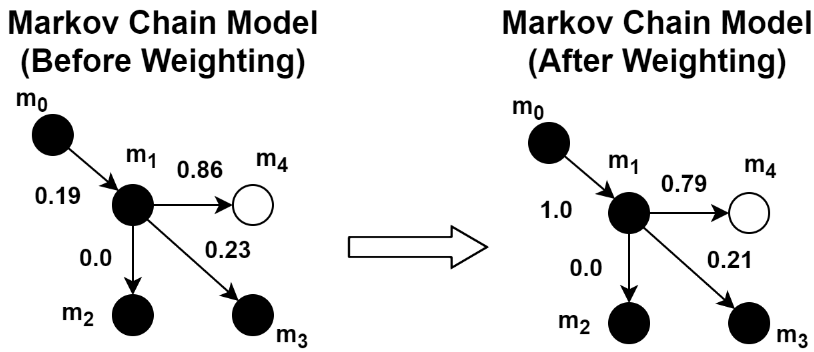

Figure 1: Markov chain model construction with weighted edges

$$T = \begin{array}{c} \\ m_0 \\ m_1 \\ m_2 \\ m_3 \\ m_4 \end{array} \begin{array}{c} \begin{array}{ccccc} m_0 & m_1 & m_2 & m_3 & m_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.21 & 0.79 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

To calculate the impact vector, in other words, the vector that contains the probabilities of predicted methods that will change, an initial vector should be multiplied with the transition matrix. We encode the initial vector with change information we have collected from diff calculation. The change information represents the likelihood of a method that could affect itself by the changes that are made for the current method. Therefore, as the amount of change increases the probability of being affected by changes will be higher. In Fig. 1, let's assume the filled nodes (methods) m_0, m_1, m_2 and m_3 in the Markov chain are the changed methods with given probabilities; $m_0 = 0.5$, $m_1 = 0.71$, $m_2 = 0.78$ and $m_3 = 0.33$. The four given change probabilities are encoded into the initial vector below. Previously, to satisfy the properties of Markov chain in the transition matrix, we weight the edges of each node's outgoing edges. Similarly, we also need to weight the initial vector values as well. According to Markov chain's properties, the summation of the probabilities in the initial vector should be equal to 1, where the sum of the probabilities in our initial vector is greater than 1.

$$I = [\ 0.5\ 0.71\ 0.78\ 0.33\ 0\ ]$$

We weight the initial vector by dividing each value in the vector by the summation of the probabilities in the vector. Thereby, we have updated our initial vector, which is given below.

$$I = [\ 0.216\ 0.306\ 0.336\ 0.142\ 0\ ]$$

Finally, we obtain the final forms of our initial vector and transition matrix, and by using the final forms of initial vector and transition matrix, we calculate the impact vector in Equation 1, which is predicted to-be changed methods. Since our initial vector and transition matrix is weighted, we expect to calculate the impact vector, where the summation of its probabilities is equal to 1.

$$I\ x\ T = [\ 0.156\ 0.181\ 0.336\ 0.181\ 0.146\ ] \tag{1}$$

Based on the Markov chain model in Fig. 1 and calculation in Equation 1, the probabilities of the methods being affected by the changes are calculated as $m_0 = 0.156$, $m_1 = 0.181$, $m_2 = 0.336$, $m_3 = 0.181$, and $m_4 = 0.146$. With respect to the probabilities, m_2 is the method that has the highest likelihood being affected by the changes.