Thủ tục IniE (E; t_bound; S)

Đầu vào: chuỗi tuần tự S = ⟨S1; : : : Sn⟩ với thời gian xảy ra tương ứng của mỗi sự kiện; thời gian giới hạn t_bound, Vector danh sách sự kiện E =∅, mỗi phần tử trong E chứa 1 vector lưu giá trị thời gian xuất hiện các sự kiện $E_i = \emptyset$.

Đầu ra: Vector C chứa các mẫu tìm được với thông tin thời điểm xuất hiện.

Nội dung giải thuật:

```
1  C := ∅;
2  E := ∅;
3  foreach(s ∈ S)
4     if (s ∉ E) then
5        E := E ∪ {s}
6     Es := Es ∪ {t(s)}
7  end(foreach)
8  foreach(e ∈ E)
9     foreach(α # e and α ∈ E )
10       if ([e|α] ∉ C) then
11          C:= C ∪ {[e|α]}
12       foreach(i∈Ee, j∈Eα and t(Eα[j])-t(Ee[i])
   < t_bound)
13       C[e|α] := C[e|α] ∪ [t(Ee[i]), t(Eα[j])]
14       end (foreach)
15    end (foreach)
16 end (foreach)
17 return C;
```

Giải thuật IniE duyệt qua chuỗi dữ liệu S duy nhất 1 lần, mỗi khi sự kiện xuất hiện sẽ được lưu vào vector E cùng với thời điểm xuất hiện của sự kiện. Sau khi duyệt qua hết chuỗi dữ liệu S, vector E chứa toàn bộ các sự kiện kèm thời điểm xuất hiện của các sự kiện. Lần lượt xét các sự kiện trong E, với mỗi cặp sự kiện khác nhau có thời điểm xuất hiện thỏa mãn tham số t_bound sẽ tạo mẫu tuần tự episode tương ứng với hai sự kiện khác nhau đó, thông tin về thời điểm xuất hiện cũng sẽ được lưu vào vector C. Vòng lặp kết thúc khi không tìm được thêm các cặp sự kiện thỏa mãn yêu cầu. Kết thúc giai đoạn 1 ta thu được vector C chứa danh sách các mẫu episode và thời điểm xuất hiện của từng mẫu episode, vector E chứa các sự kiện và thời điểm xuất hiện của các sự kiện.

Định lý 1: Độ phức tạp của IniE là $O(n+ l^2m)$.

Chứng minh: Với vòng lặp đầu tiên duyệt qua chuỗi tuần tự S để lưu lại thông tin và thời điểm xảy ra sự kiện có độ phức tạp tuyến tính phụ thuộc vào độ lớn n của chuỗi sự kiện S. Trong vòng lặp, mỗi thời điểm sự kiện s xuất hiện trong S sẽ được lưu thời gian lại theo loại trong vector sự kiện (E) nên thời gian tính toán sẽ tuyến tính theo **n**. Vậy độ phức tạp từ dòng 1-7 là O (n). Vòng lặp thứ 2 từ dòng 8 đến dòng 16 lặp qua từng loại sự kiện trong vector sự kiện E. Vòng lặp này chứa vòng lặp con lặp lại qua từng sự kiện trong E và không xét 2 sự kiện trùng nhau nên độ phức tạp lúc này là l^2 (với l là số sự kiện trong E). Với mỗi mẫu tuần tự episode được tạo ra, lần lượt xét thời gian từng cặp sự kiện tạo nên mẫu với vòng lặp từ dòng 11 đến dòng 14. Vòng lặp này sẽ lặp qua **m** lần thời gian của sự kiện xuất hiện nhiều (α hoặc e). Mỗi cặp thời gian thỏa ràng buộc tổng thời gian nhỏ hơn độ lớn t_bound sẽ được thêm vào vector C tương ứng với mẫu được tạo ra bởi 2 sự kiện đó. Như vậy độ phức tạp thủ tục IniE là $O (n + l^2m)$. Trong thực tế, số lần xuất hiện **m** của các sự kiện và số sự kiện l trong E nhỏ hơn độ lớn **n** của chuỗi dữ liệu S nhiều lần nên giải thuật IniE là chấp nhận được.

3.2 Giai đoạn 2 – Tạo các mẫu Episode dạng mở rộng và hiệu chỉnh thông tin trên các mẫu

Giải thuật ExtEE đọc thông tin các mẫu episode trong vector C và thông tin các sự kiện trong vector E tạo ra ở giai đoạn 1. Đầu tiên khởi tạo vector M để chứa các mẫu mở rộng và thông tin độ đo ủng hộ thuần. Lần lượt xét các cặp mẫu trong vector C, với mỗi cặp mẫu có cùng sự kiện cuối lần lượt xét thời điểm xuất hiện của cặp mẫu. Nếu mẫu mở rộng chưa có trong vector M thì thêm vào vector M. Mỗi cặp thời điểm thỏa giá trị t_bound sẽ được lưu lại trong vector M theo tương ứng với từng mẫu trong M và cập nhật độ đo ủng hộ của các mẫu trong C. Các cặp thời điểm thỏa giá trị t_bound sẽ có 3 trường hợp xảy ra: 2 mẫu cùng xuất hiện đúng vị trí trong thời gian t_bound, mẫu thứ nhất có thời điểm bắt đầu và thời điểm kết thúc nhỏ hơn thời điểm mở đầu và kết thúc của mẫu thứ hai (hoặc ngược lại), mẫu thứ nhất được chứa trong mẫu thứ hai (hoặc ngược lại). Kết thúc giải thuật, vector M sẽ chứa các mẫu mở rộng với độ đo ủng hộ thuần. Giải thuật được mô tả chi tiết bằng ngôn ngữ giả như sau: