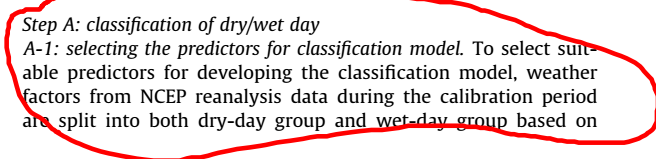
**Phương pháp phân loại ngày có mưa và ngày ko mưa (dry/wet day classification)**

Trong bài báo có đề cập:



Predictors: tức là một số yếu tố khí hậu như áp suất, hướng gió, độ ẩm…sẽ ảnh hưởng đến mưa. Một ngày có mưa hay không phụ thuộc cơ bản vào các yếu tố này. Do vậy, việc phân loại ngày có mưa hay không mưa cần input là những số liệu này.

==> Hiểu gần đúng vấn đề. Predictor là model SVM được xây dựng trên các features được đầu vào. Features trong bài toán này là các features trong table 01. với mỗi tập features sau khi train và test dựa trên accuracy ta sẽ chọn model có accuracy tốt nhất.

Ví dụ: với bộ features 01 = [Mslp, P5\_f] ta xây dựng được model SVM với accuracy là 60%. Với bộ features 02 = [Mslp, P5\_f, p\_850] ta xây dựng được model SVM với accuracy là 80%. Thì ta chọn features là 02, với model đã được xây dựng. Vì nó là tập tốt nhất để xaya dựng model. Thường thì chia tập dataset ra 70% để train, 30% để test.

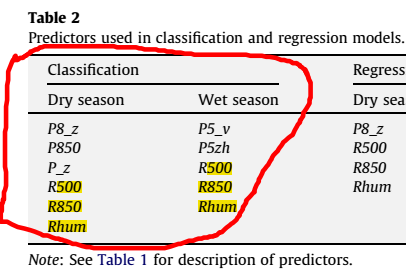
**Bài báo đã đề xuất:**

6 yếu tố ảnh hưởng đến ngày có mưa hay ngày không có mưa vào mùa khô – dry season (table 2) ==> OK. cái này là họ chạy test và chọn ra được features này nhé.

5 yếu tố ảnh hưởng đến ngày có mưa hay ngày không có mưa vào mùa mưa – wet season (table 2) ==> OK

éTức là đề xuất việc phân loại ngày có mưa hay không có mưa phụ thuộc vào các biến đó. Thế nên input của bài toán sẽ là các biến đó, và ouput của bài toán sẽ là ngày có mưa hay ngày không mưa. ==> Biến ở đây mà chú hiểu trong machine learning nó nói là FEATURE đấy nhé. Thường thì số lượng features và data có càng nhiều càng tốt. Nếu được thì khi xin data, nên xin hết features mà có trong table 01. Nếu thấy có feature nào mà mình nghi ngờ là tốt cho quá trình dự báo thì mình xin hết luôn. Thế nhé

Việc phân loại ngày có mưa hay ngày không có mưa cần tách số chuỗi số liệu ra thành 2 mùa, một chuỗi số liệu đại điện cho mùa mưa (số liệu nằm trong khoảng tháng 5-11 hằng năm) và một chuỗi số liệu đại diện cho mùa nắng (tháng 12 – 4 hàng năm). Vì đặc tính khí hậu 2 mùa là khác nhau, nên đề xuất tách ra 2 mùa để việc dự báo chính xác hơn. ==> Cái này thuộc về chuyên ngành chú nên anh ko rõ.



Sau khi input và xây dựng mô hình SVM cho bài toán, sẽ tìm được phương trình 12, 13.

Ldry season, Lwet season : nếu có giá trị dương là ngày mưa, giá trị âm là ngày không mưa. Như vậy bài toán đã được giải quyết. ==> OK. Dương là ngày mưa, âm là ngày không mưa. Cái này khi train SVM mình định nghĩa nhé. Phương trình 12,13 là input để xây dựng SVM model nhé. Không phải là output đâu.

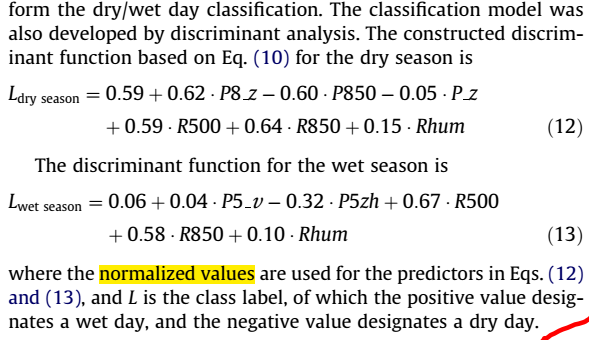
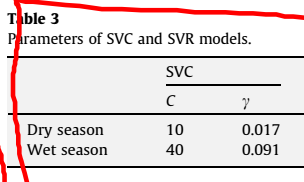


Table 3 cho thấy giá trị hiệu chỉnh của 2 tham số SVM của 2 mùa. ==> Chính xác, khi xây dựng model SVM, người ta chỉ quan tâm tới 2 tham số này.



OK. Khi nào có data thì PM nhé. Gửi data cho anh qua mail: [nvthao.math@gmail.com](mailto:nvthao.math@gmail.com). để anh chạy thử xem sao

best regards.