International Congress on Environmental Modelling and Software

9th International Congress on Environmental Modelling and Software - Ft. Collins, Colorado, USA - June 2018

Jun 27th, 2:00 PM - 3:20 PM

Data Mining Attribute Filtering Approach for Drought Environmental Hazard Modeling

Getachew B. Demisse
University of Nebraska - Lincoln, gdemisse2@unl.edu

Data Mining Attribute Filtering Approach for Drought Environmental Hazard Modeling

Getachew Demisse, Tsegaye Tadesse, Brian Fuchs, Mark D Svoboda
National Drought Mitigation Center, School of Natural Resources, University of Nebraska–Lincoln

ABSTRACT

Attribute filtering is the process of objectively searching the best subset of attributes from available multiple attribute options for improved environmental modeling. In this research, attributes are environmental descriptor variables or parameters, which are used to model drought-related environmental hazard. Some of the attributes used to model drought include the standardized precipitation index (SPI), standardized seasonal greenness (SSG), Atlantic meridional mode (AMM), quasi-biennial oscillation (QBO), solar flux (SF), and digital elevation model (DEM). An excerpt from the sample data and list of attributes used for the experimental analysis is presented in Appendix 1. The SPI, SSG, AMM, QBO, SF, and DEM are drought model variables or parameters that can be derived from climate, satellite, ocean-atmospheric interactions, and environmental-biophysical parameters used for modeling drought in space-time dimensions. There are other potential attributes, from climate, satellite, ocean-atmospheric interactions, and environmental-biophysical parameters that can be used for modeling drought. For operational drought modeling purposes, we may not include all the potential attributes available for modeling drought phenomena using the usual literature review and experts' judgement, which is unsystematic and dominated by arbitrary trials. Subjective experts' judgement is a major cause of model uncertainties and a major challenge for converting theoretical models into practical real-world problem-solving applications. Therefore, taking this challenge into account, the focus in this research was identifying and selecting the most relevant environmental descriptor variables from available multiple options and removing irrelevant attributes objectively (not just subjectively with the experts' judgement) without loss of information. The main goal of this research was to develop an automated data-mining attribute filtering approach for objective-based feature selection in environmental modeling. Four attribute selection algorithms (correlation-based attribute selection, principal component-based attribute selection, relief attribute evaluation, and wrapper subset evaluation) were compared for their best performances. For the experimental analysis, data from climate, satellite, biophysical, oceanic and atmospheric interactions for modeling drought related environmental hazards were used. Attribute merit values for modeling the target-dependent attributes were determined and compared with possible alternative attributes for objective-based feature selections. The average merit values for the selected attributes were also ranked. The average merit values for the selected attributes ranged from 0.5 to 0.9 for the case study conducted.

This research is complementary to the extensive review and common sense in identifying relevant attributes for a given domain; and it does not mean that the researchers have not to use their common sense and check with established truth or theory bases in the domain specific research. The methodology developed here helps to avoid the uncertainty of domain experts' attribute selection challenges, which are usually unsystematic and dominated by somewhat arbitrary trials. Future research may evaluate the developed methodology using relevant classification techniques (such as classification and regression tree or random forest) and quantify the actual information gain from the developed approach.

Keywords

Data mining; Descriptor variable; Drought; Merit value; Modeling

Appendix 1: An excerpt from the data type definition for the attribute selection methods used in java api.

The ? mark is the missing data for the attributes' values. Abbreviations: am = Atlantic Meridional Mode, amo = Atlantic Multidecadal Oscillation, best = Bivariate ENSO (*El Nino*–Southern Oscillation) time series, dmi = Dipole Mode Index, dem = Digital Elevation Model, LC = Land Cover, WHC = Water Holding Capacity of the Soil, N_Precip = Normalized Precipitation, SSG = Standardized Seasonal Greenness, SM = Soil Moisture, EC = Ecoregion, MEI = Multivariate ENSO Index, ONI = Oceanic Niño Index, PDO = Pacific Decadal Oscillation, TNI = Trans-Niño Index, TNA = Tropical Northern Atlantic Index, TSA = Tropical Southern Atlantic Index, QBO = Quasi-Biennial Oscillation, SF = Solar Flux (10.7cm), Nino 4 = Central Tropical Pacific SST, Nino 3.4 = East Central Tropical Pacific SST, NAO = North Atlantic Oscillation, PNA = Pacific North American Index, SOI = Southern Oscillation Index.

```
@relation GHA_dekad_22_out01

@attribute awc numeric
@attribute dem numeric
@attribute eco
{?,10085,10088,10089,10092,10093,10094,10096,10097,10108,10109,10
112,10113,10116,10118,10121,10122,10123,10125,10128,10129,10130,1
0131,10132,10133,10135,10136,10138,10142,10147,10151,10153,10159,
10161,10162,10167,10169,10174,10175,10176,10177,10179,10184,10189
,10190,10191,10194,10673,10695,10696,10698,10700,10702,17003}
@attribute landcover
{11,14,20,30,40,50,60,90,110,120,130,140,150,160,180,190,200,?}
@attribute amm numeric
@attribute amo numeric
@attribute best numeric
@attribute dmi numeric
@attribute mei numeric
@attribute nao numeric
@attribute Nino34 numeric
@attribute Nino4 numeric
@attribute oni numeric
@attribute pdo numeric
@attribute pna numeric
@attribute qbo numeric
@attribute sflux numeric
@attribute soi numeric
@attribute tin numeric
@attribute tna numeric
@attribute tsa numeric
@attribute Zscore numeric
@attribute SM numeric
@attribute SSG_dek22 numeric
@attribute SSG_dek23 numeric

@data
?,599,10698,200,410,190,510,21652,366,-220,2688,2905,-100,-770,-
140,-2167,1631,-700,-2846,340,140,-332,?,-1610,-1486
?,599,10698,200,-1990,119,1450,-
91815,949,360,2782,2948,900,600,610,1060,1839,-1600,-3001,-
50,10,-332,?,-991,-557
?,599,10698,200,1370,428,170,-11446,267,-220,2718,2908,200,880,-
310,-2464,1221,200,-2412,590,490,-332,?,867,1300
?,599,10698,200,3700,330,900,-39820,665,-
740,2770,2937,700,850,1520,874,1110,-500,-3253,680,220,-
332,?,247,371
```