Using Bayesian Networks to Estimating Rainfall Distribution Given Polarimetric Radar Data

Jose Picado
Vahid Ghadakchi
Zahra Iman

Probabilistic Graphical Models
Winter 2015

Introduction

- Rainfall amount prediction problem
  - Using gauges
  - Remote sensing
- Using Bayes Nets
  - Designing the structure
  - Learning the structure

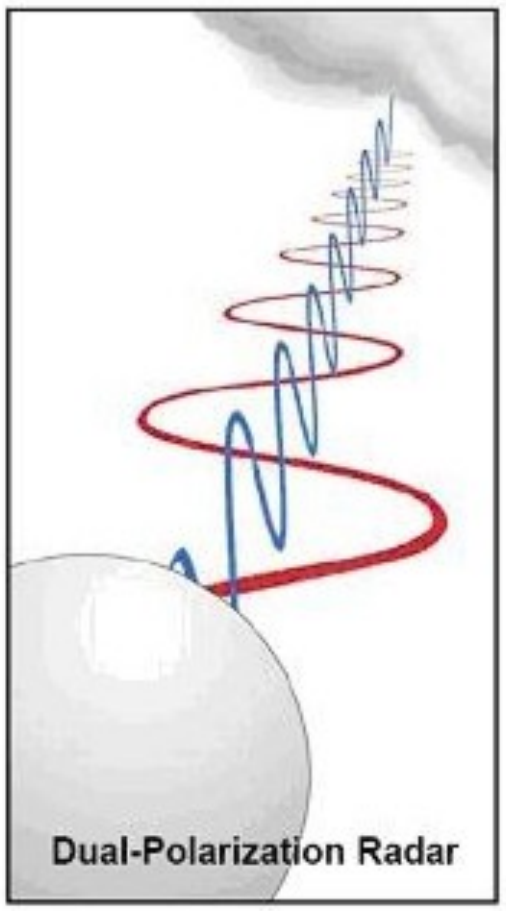


Image courtesy of NOAA

Features

TimeToEnd	RR1	RhoHV
DistanceToRadar	RR2	Velocity
Composite	RR3	Zdr
HybridScan	RadarQualityIndex	LogWaterVolume
HydropmeteorType	Reflectivity	MassWeightedMean
Kdp	ReflectivityQC	MassWeightedSD

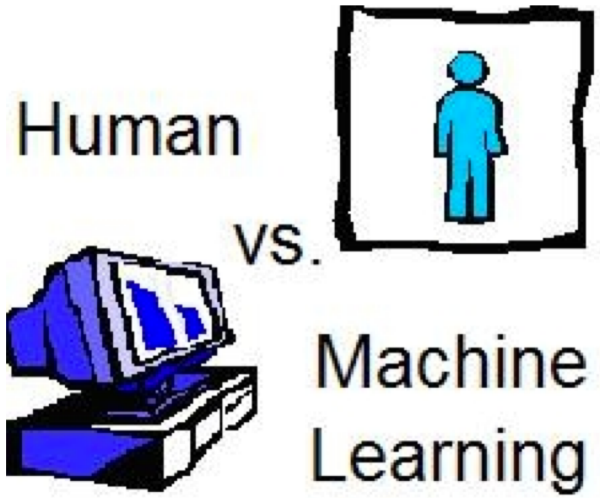
- Features with composite values aggregated into a single value
- Introduced feature: Measurements
- Label: Expected – actual amount of rain reported by rain gauge for that hour

Dependencies

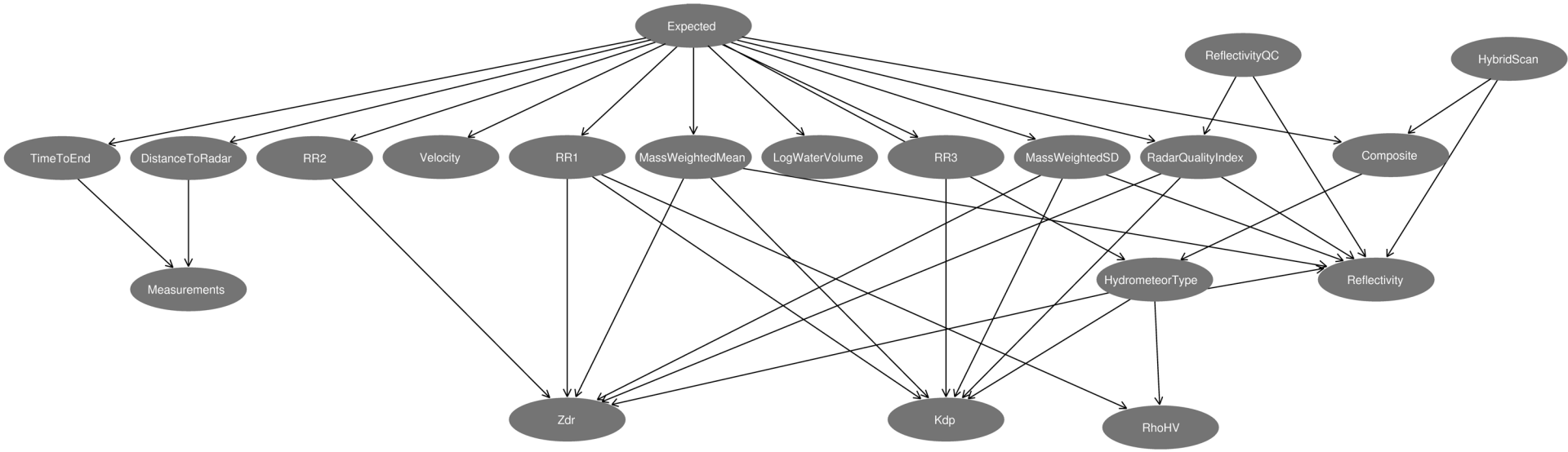
- RR1: Rain rate from the HCI-based algorithm
  - HydrometeorType, Zdr, RhoHV, Kdp
- RR2: Rain rate from the Zdr-based algorithm
  - Zdr
- HydrometeorType: rain, snow, big drops, birds,…
  - Based on HCA algorithm with inputs of Zdr, Kdp, RhoHV, Reflectivity
- RadarQualityIndex: value from 0 to 1 (good data)
  - Based on an algorithm proposed in a paper with inputs of Reflectivity, Zdr, Kdp, RhoHV

Approaches

- Bayes network with manually designed structure
- Bayes network with learned structure



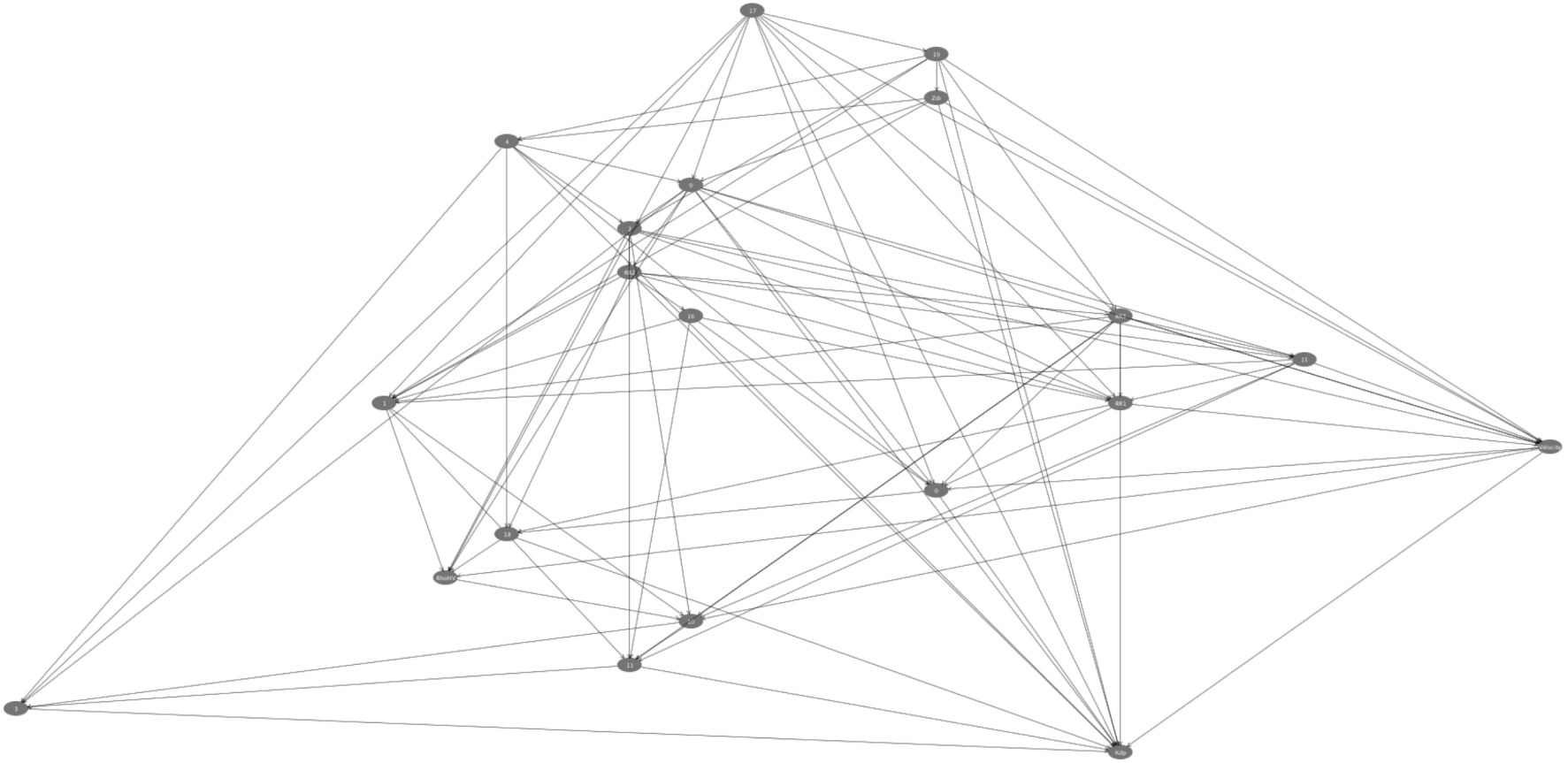
Manually designed structure



Learned structure

- Structure learning algorithms:
  - Structure scoring: Bayesian scoring with Dirichlet priors
  - Greedy: Hill Climbing, K2
  - Non-greedy: Simulated Annealing

Learned structure



Results

- Data and Setting
  - Subset of training set
  - Implementation using Weka
- Results
  - Continuous Ranked Probability

Algorithm	CRP
Logistic Regression	0.01096134
Naïve Bayes	0.01452578
Bayes Net - Manually Designed Structure	0.00965175
Bayes Net - Simulated Annealing	0.00864957

Conclusions

- Bayes networks proved to be an appropriate model:
  - Exploit (in)dependencies between features
- Domain knowledge can be used to build the network structure
- Structure learning is better when data is available and time is not an issue
- Future work:
  - Structure provided by domain expert
  - Aggregations functions for composite-valued features
  - Undirected models
  - Build model with full dataset