Classifying Rainfall Regions in Weather Radar Mosaics

ALEX HABERLIE
LOUISIANA STATE UNIVERSITY
AHABERLIE1@LSU.EDU

PROJECT COLLABORATORS: WALKER ASHLEY (NIU), VICTOR GENSIINI (NIU), RUSS SCHUMACHER (CSU)
Useful Information

https://github.com/ahaberlie/unidata-workshop-2018

Haberlie and Ashley (2018)

McGovern et al. (2017) – Great overview of current machine learning trends in Meteorology

Storm Identification and Feature Extraction
  ◦ WDSS-II (Lakshmanan et al. 2007)
  ◦ TITAN (Han et al. 2009)
  ◦ THoR (Houston 2015)
  ◦ Hagelslag (Gagne 2018)

Machine Learning and Forecasting:
  ◦ MCS initiation (Ahijevych et al. 2016)
  ◦ Damaging Straight-Line Wind Prediction (Lagerquist et al. 2017)
  ◦ Heavy rain forecasting (Herman and Schumacher 2018)
Mesoscale Convective Systems

Source: NOAA Storm Prediction Center

Source: Walker Ashley

Source: Arturo Fernandez/Rockford Register Star via AP.

Source: NOAA Storm Prediction Center
Parker and Johnson (2000)  

Gallus et al. (2008)
Are These MCSs?
What about these?
Problem Description

How do you generate a climatology of mesoscale convective systems (MCSs) with a huge dataset of composite radar mosaics?

- NOWrad (~2 km)
- Over 95% of 15 minute periods from 1996-2017
- ~10^6 images
- Many well-known issues, but the analyses can be useful (Fabry et al. 2017)

Parker and Johnson (2000) objective definition:

- Convective cells organized on a horizontal scale of at least 100 km
- Must last for at least 3 hours

Computing Resources:

- Ryzen 1700 (8 C, 16 T), nVidia GTX 1070, 32 gb RAM
Why Machine Learning?

“Reducing time to science”

~5.5 million “MCS Snapshots”

Automate classification of MCSs and four common false positives after segmentation

- Tropical Systems
- Synoptic Systems
- Unorganized clusters
- Ground Clutter / Noise / Etc.
Related Work

Baldwin et al. 2005
- Linear, cellular, stratiform

Gagne et al. 2009
- Pulse, multicellular, MCSs

Lack and Fox 2012
- Supercell, QLCS, rotating storms, pulse, etc.

Hobson et al. 2012
- Supercell, pulse, multicellular, linear
Sample Training Workflow

1) Ask yourself a few questions:
   ◦ What are the classes you want to identify?
   ◦ What are distinguishing features of each class?
   ◦ What data do you need to gather samples?
   ◦ What algorithm should I use?

2) Identify class examples

3) Extract features
   ◦ Area, Shape, Intensity, etc.

4) Generate training and testing data

5) Train machine learning model
   ◦ Always test model performance on data not used to generate model
Notebook Examples

Training Process

Extraction Process
MCS: 93.0%

- MCS
- Unorganized
- Tropical
- Synoptic
- Clutter
Machine Learning vs. Manual

Manual MCS slice positions (2003-2013)

Automated Approach (2015)

- 0.95 threshold 40-hr isopleth
- 0.5 threshold 40-hr fill
- No threshold 40-hr isopleth
Quasi-linear convective systems (QLCSs) can produce severe weather (Trapp et al. 2005)

Convection-permitting models have trouble with QLCSs (Lawson and Gallus 2016)

Implications for people, weather forecasting, and high resolution climate simulations
Visual differences

QLCS

Non-QLCS
First Try

Select 3000 random high-probability MCS “snapshots”

Label as QLCS or Non-QLCS based only on their visual features
  ◦ Subjective
  ◦ Looking for common visual traits

Use features to train tree-based ensemble
  ◦ ~70% accuracy
Second Try

Employ a convolutional neural network (Krizhevsky et al. 2012)

Inspiration / model configuration came from astronomy (Dieleman et al. 2015)

Much harder to generate training / testing data for CNNs
Training data approach

What is visually important for classifying this as a QLCS?

1) The entire structure?
2) Stratiform features?
3) Convection features?

“I’m looking for an intense line with a strong reflectivity gradient.”
Training / testing data creation

All images must be the same size

Find largest contiguous region of 50+ dBZ

Center a box on the intensity-weighted centroid

Extract intensity information within box
Data Augmentation

Addressing overfitting
- Only ~3000 samples

Keras ImageDataGenerator
- Randomly apply slight modifications to images during training

Physically Reasonable?
- Scale is important
- Orientation might be important

Notebook Examples

QLCS Detection
Application

QLCS Occurrence
June – August (2001-2013)

Percent of MCS events that were QLCSs
June – August (2001-2013)
References


Thank You