BIG DATA IN ECOLOGY

By

MORTEZA SHAHRIARI NIA

A DISSERTATION PROPOSAL PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHYLOSOPHY

UNIVERSITY OF FLORIDA

2015

I dedicate this to my parents who selflessly devoted the best of their lives to raise their kids to be hard working, humble, selfless and to take steps towards the greater human wisdom

ACKNOWLEDGMENTS

I would like to thank Dr. Daisy Zhe Wang for believing in me and providing the great opportunity of tackling the whole real of big data. I would also like to thank Dr. Yuguang Fang for his great support and commitment. Dr. Paul Gader and Dr. Stephanie Bohlman were great mentors, without contributions of whom this work would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

Table page

LIST OF FIGURES

Figure page

Abstract of Dissertation Proposal Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Phylosophy

BIG DATA IN ECOLOGY

By

Morteza Shahriari Nia

April 2015

Chair: Dr. Daisy Zhe Wang
Major: Electronics and Computer Engineering

Ecological sciences benefit from the huge diversity of plant species which play an important role in large scale ecological aspects such as global warming, land cover change, CO^2 emission, invasive species, fire hazard, and etc. State-of-the-art species classification techniques utilize remote sensing data such as hyperspectral and LiDAR, however this task involves plenty of field data collection which is both highly time consuming, costly and can only be accomplished by ecological experts. Among thousands of the most commomly found plant species there is huge similarities between them from a remote sensing point of view which makes the task of species classification very daunting; therefore we see a whole body of literature specifically dedicated to this issue which is yet far from real world scenarios with thousands of possible species. While this is an indicator of the importance and complexity of the issue, little has been done to tackle the problem from a computational point of view harnessing the power of "big data". Periodic airborne campaigns can generate terrabytes of data on vast swaths of land. To tackle these problems we propose to use probabilistic knowledge bases and deep learning both of which work best when there is lots and lots of data. Probabilistic knowledge base captures ecological expert knowledge in terms of probabilistic rules, which will be maped to remote sensing data and used to infer new facts and therefore enhance species classification accuracy. Deep learning on the other hand as a semi-supervised algorithm will benefit

from the vast amounts of data available and capture intrinsic features of data through its layered architecture and thus help in reducing the amount of labeled data required.

CHAPTER 1
INTRODUCTION

Understanding the dynamics of ecological structures has important impacts on climate change, land cover change, invasive species, fire hazards,

CHAPTER 2
REMOTE SENSING

CHAPTER 3
MASSIVE DATA MINING

CHAPTER 4
PRELIMINARY RESULTS

CHAPTER 5
INFORMATION EXTRACTION

CHAPTER 6
CONCLUSION

REFERENCES

BIOGRAPHICAL SKETCH

This section is where your biographical sketch is typed in the bio.tex file. It should be in third person, past tense. Do not put personal details such as your birthday in the file. Again, to make a full paragraph you must write at least three sentences.