

The title of the dissertation using sentence
case: First word of subtitle capitalized

Student P. Name

A dissertation for the
Doctor of Philosophy degree
presented to the faculty of the
School of Data Science
University of Virginia
on 23 April 2024

Dissertation committee

John Conway, Advisor
Isaac Newton, Chair
Florence Nightengale
Katherine Johnson
Kermit T. Frog

Abstract

The title of the dissertation using sentence case: First word of subtitle capitalized

Student P. Name

The following abstract is from the dissertation of Natalie Kupperman, PhD. Athlete monitoring is the practice of collecting data in athletics to quantify load, both physical and mental, with the goal to reduce fatigue, optimize performance and mitigate injury risk. The influx in technology over past 15 years has given practitioners the ability to quantify these measures at an increased frequency. However, the subsequent research that followed has found limited associations between these measures and injury risk. There are many possible reasons for this paucity, one of which is the methods used to investigate these data. The purpose of this dissertation was to use three different methodological techniques to both introduce novel analysis to the field and also highlight the need for within-person design in athlete monitoring.

The purpose of manuscript 1 was to evaluate differences in external workload trends of athletes who sustained an overuse injury during sport. We did this using a case series format to evaluate differences in on-court volume, as measured by whole-body accelerometry, of three court-sport athletes and three position-matched healthy control athletes. Across sports, teams, and sexes, we found differences in accelerometer data between injured athletes and healthy matched controls in the 8 weeks leading up to the injury. The injured volleyball athlete had greater accumulated playerload per min (PL/min), jumps per minute and duration compared to the healthy athlete across the 8 weeks. For basketball, both injured athletes had greater PL/min compared to their controls, however, jumps per min were less than the healthy athletes over the 8 weeks.

The purpose of manuscript 2 was to apply longitudinal structural equation modeling framework to athlete monitoring data with the goal of better understanding between- and within-person differences. Longitudinal athlete sport participation status and self-reported wellness were collected from 16 volleyball athletes and aggregated into weekly observations over 5 weeks. Three models – univariate

latent curve model (LCM), LCM with structured residuals (SR), and LCM-SR with time varying covariates – are described. Then using simulated data based off the 16 athletes, results for each model are reported and substantively interpreted. The final model shows statistically significant effects of wellness on both the same and subsequent week athlete availability.

The purpose of manuscript 3 was to first introduce a research design using machine learning (ML) and explanatory model that is new the field of athlete monitoring and the use a case study in collegiate basketball to demonstrate the proposed research design. The research design contained 3 steps. First, data was processed using an ML algorithm designed for feature selection. Then based on the important features defined by the ML algorithm and researcher input, variables are selected, and a testable hypothesis is generated. The hypothesis was then tested using an explanatory model. The external load and readiness datasets performed best with ordinal forests, whereas the athlete self-report measures data did best with an support vector machine. Readiness had the best performance with MSE = 0.89. For the explanatory model, only the physical performance capability (PPC) model was statistically significant. This model found the odds having above average soreness decrease by 57% for every point increase in PPC (OR=0.57, 95% CI 0.32 to 0.97, $p<0.05$).

In undertaking these studies, the complexities of modeling injury were demonstrated. There is still much to be explored; however, by attempting new ways of analyzing athlete monitoring, we were able to highlight the need for continued forward thinking in methodologies, both at the level of analysis and the analytical tools employed.

Acknowledgements

From the dissertation of Phil Bourne. I am exceedingly grateful to Dr. M. R. Taylor for his supervision during my candidature for this degree. Special thanks must also go to Mr. J. Drennan and Ms M. J. McCall who were a continuous source of inspiration, and who will long be remembered. Thanks also to Drs. N. Latavalya and E. Summerville, Mr. J. Mohyla and Professor D. J.M. Bevan who provided useful comments and Mr. J. A. Westphalen who was always helpful in running XRAY-76 on the CDC7600 at The South Australian Institute of Technology.

People external to The Flinders University of South Australia who should be acknowledged are Drs. A. H. White and C. L. Raston for the 4-circle diffractometer data sets collected at The University of Western Australia and Prof. G. M. Sheldrick for his personal introduction to SHELX while solving the structure of [ZnC₁₄][cytosine.H]₂ at the International Summer School on Crystallographic Computing, Twente, The Netherlands, July, 1979.

Fast efficient technical assistance was always received from Mrs. A. Hermann. Thanks of a personal nature must go to Miss L. C. Clunes and Mr. F. C. D. Bourne.

This work was undertaken during the tenure of a Commonwealth Postgraduate Research Award.

Table of Contents

Committee	ii
Abstract	iii
Acknowledgements	v
List of Figures	vii
List of Tables	viii
Chapters	
1 Introduction	1
1.1 Motivation	1
2 Literature Review	2
3 Title of first paper	4
4 Title of second paper	5
5 Title of third paper	6
6 Conclusion	7
References	7
Code and supplementary files	9

List of Figures

2.1	Back-propagated fine-tuning process. [2]	3
4.1	Why use <code>input</code> when rendering your dissertation	5

List of Tables

3.1 US Schools of Data Science and Their Founding Dates	4
---	---

1 Introduction

This is the introduction to your dissertation. This introduction came from D. E. Brown, “Text Mining the Contributors to Rail Accidents,” in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 2, pp. 346-355, Feb. 2016.

In the 11 years from 2001 to 2012 the U.S. had more than 40,000 rail accidents with a total cost of \$45.9 M. These accidents resulted in 671 deaths and 7061 injuries. Since 1975 the Federal Railroad Administration (FRA) has collected data to understand and find ways to reduce the numbers and severity of these accidents. The FRA has set “an ultimate goal of zero tolerance for rail-related accidents, injuries, and fatalities” [1].

A review of the data collected by the FRA shows a variety of accident types from derailments to truncheon bar entanglements. Most of the accidents are not serious; since, they cause little damage and no injuries. However, there are some that cause over \$1M in damages, deaths of crew and passengers, and many injuries. The problem is to understand the characteristics of these accidents that may inform both system design and policies to improve safety.

After each accident a report is completed and submitted to the FRA by the railroad companies involved. This report has a number of fields that include characteristics of the train or trains, the personnel on the trains, the environmental conditions (e.g., temperature and precipitation), operational conditions (e.g., speed at the time of accident, highest speed before the accident, number of cars, and weight), and the primary cause of the accident. Cause is a four character, coded entry based on based on 5 overall categories (discussed in Section IV). The FRA also collects data on the costs of each accident decomposed into damages to track and equipment to include the number of hazardous material cars damaged. Additionally, they report the number of injuries and deaths from each accident.

1.1 Motivation

This is a section on motivation. Text is ragged right justified, left aligned.

2 Literature Review

Your literature review goes here. Here we show a little mathematics and figure from [2].

$$W_{i+1} = W_i + \eta(\langle v_i h_i \rangle_{P(h|v)} - \langle v_i h_i \rangle_{\text{recon}}) \quad (2.1)$$

$$b_{i+1} = b_i + \eta(\langle v_i \rangle_{P(h|v)} - \langle v_i \rangle_{\text{recon}}) \quad (2.2)$$

$$a_{i+1} = a_i + \eta(\langle h_i \rangle_{P(h|v)} - \langle h_i \rangle_{\text{recon}}) \quad (2.3)$$

Note that the equations are labeled so that they can be referenced in the text. For example, 2.3 refers to the 3rd equation above. Figures are also referencable, like this: 2.1.

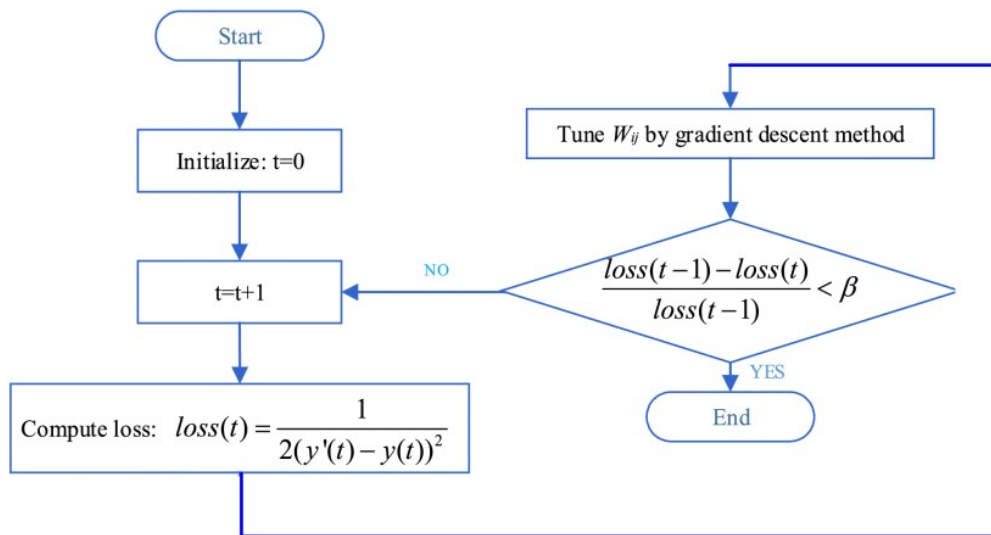


Figure 2.1: Back-propagated fine-tuning process. [2]

3 Title of first paper

This chapter has a table.

Institution	Founding Date
University of Virginia School of Data Science	2019
Indiana University Luddy School of Informatics, Computing, and Engineering	2015
University of California, Berkeley Division of Data Sciences	2017
NYU Center for Data Science	2013
University of Rochester Goergen Institute for Data Science	2015

Table 3.1: US Schools of Data Science and Their Founding Dates

4 Title of second paper

By placing the text of each chapter in its own file, you can use the file for the dissertation and for individual publications. The formatting will automatically change depending on which document is being rendered.

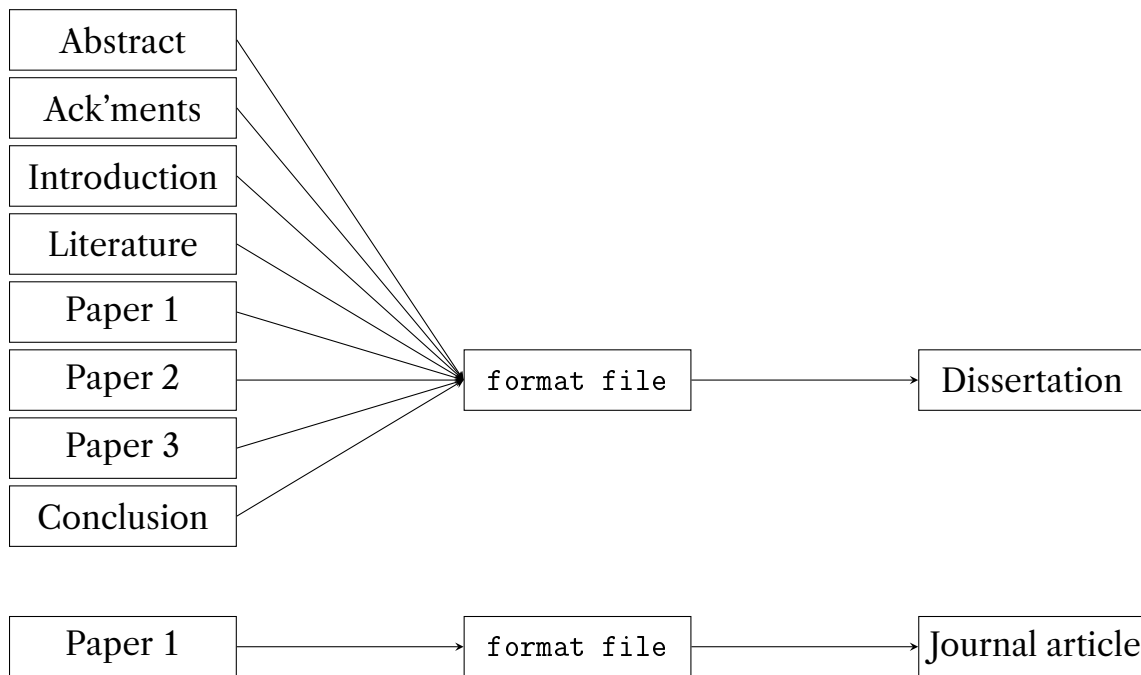


Figure 4.1: Why use input when rendering your dissertation

5 Title of third paper

6 Conclusion

References

- [1] Federal Railroad Administration. Railroad safety statistics 2009 annual report. Technical report, U.S. Department of Transportation, Washington, DC, 2011. Last updated: April 1, 2011.
- [2] Tinghui Ouyang, Yusen He, Huajin Li, Zhiyu Sun, and Stephen Baek. Modeling and forecasting short-term power load with copula model and deep belief network. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3(2):127–136, 2019.

Code and supplementary files

Code scripts and supplementary files for this dissertation are available at the following locations.

Files	Location
Paper 1	github.com/username/reponame1
Paper 2	github.com/username/reponame2
Paper 2	github.com/username/reponame3

