# Review of Papers by Subasish Das, former student of Xiaoduan Sun

Brad Burkman

27 March 2021

## Contents



# 1 *Using Deep Learning in Severity Analysis of At-Fault Motorcycle Rider Crashes*

Das, Dutta, Dixon, Minjares-Kyle, and Gillette (2018)

## 1.1 Abstract

- That Das compared motorcycle and passenger car fatalities *per mile traveled* seems a good basis of comparison.

- Comparing number of motorcycle fatalities in 2014 and 1997 with no corresponding comparison of miles traveled or number of motorcycles on the road, or something else to compare data seventeen years apart, seems sketchy. Motorcycle ridership may have increased dramatically over that time, and an increase indicates not only more riders but more inexperienced riders.
- I Googled "curved aligned roadways," from the abstract, and this paper is the only usage of the term. I don't know what it means, other than, from context, it's dangerous for motorcycles.
- Is a "deep learning framework" not a "machine learning algorithm"?
- "The intensity of severities was found to be more likely associated with rider ejection, two-way roadways with no physical separation, curved aligned roadways, and weekends."

These correlations were obvious before this study.

| | |
|---|---|
| Rider ejection | What motorcycle accidents over 20 mph don't result in rider ejection? Ejection may correlate with speed, which would correlate with severity. |
| Two-way roadways | Can't have a head-on collision if it's a one-way street. |
| Curved aligned roadways | Yep. Lots of accidents happen on curves. |
| Weekends | Motorcycle ridership is way up on weekends, and the number of recreational (less experienced) riders is up. |

### 1.2 Introduction

Again, the sketchy use of data, not saying whether the comparison is per roadway mile, per number of drivers, or total.

"Thus, it is essential to determine the influence of patterns of associated factors with crash severity." In my work, I need to find a clearer way to say this, that my analysis will be significantly different from other methods.

### 1.3 Literature Review

What are *geometric variables*, and why are they called that? "Several geometric variables have been identified in the literature to contribute to motorcycle crash-severity outcomes including roadway type, light conditions, posted speeds, roadway features such as curves or T-junctions, and weather."

The term *farm to market road* is specific to Texas.

Definition of *loop and lollipop routes*: `https://desktop.arcgis.com/en/arcmap/latest/extensions/roads-and-highways/loop-and-lollipop-routes.htm`

"Motorcycle crashes have been found to be more severe during daytime weather versus wet or rainy weather." Another possible reason for this correlation is that recreational riders don't go out in wet or rainy weather.

This article, so far, makes no mention of the patterns of recreational use of motorcycles by less experienced riders.

### 1.3.1 Overview of Models Used to Analyze Crash Severity

**Multinomial ordered probit and logit models** order the crashes by category of injury-severity levels, from no injury to fatal. Limitations of the model are "constrained effects resulting from ordered modeling" and underreporting of crashes with no severe injury.

**Multinomial logit models (MNL)** Consider three or more outcomes, without incorporating the ordering into the model.

**Mixed Logit Analysis Models**

**Log Linear Modeling**

**Empirical Bayesian Analysis**

**Stepwise Logic Regression**

Key final statement, telling why other methods are needed: "It is important to note that conventional statistical models are good at statistical inference. The common limitation of these methods is poor prediction accuracy. In addition, the models are based on assumptions. Violations of any of the assumption will produce biased results."

### 1.3.2 Overview of Deep Learning Method Analysis

Two approaches: Data modeling v/s Algorithmic modeling.

Data modeling (statistical modeling) assumes that the data was generated by an underlying stochastic process, and our job is to uncover that stochastic process that maps the independent variables (predictor variables) to the dependent variables (response variables). To evaluate the model, apply goodness-of-fit tests.

Algorithmic modeling (machine or deep learning): Understanding the unknown by minimizing the error rate through a black box algorithmic model.

"Imputation": "traffic data imputation," methods for cleaning up dirty and incomplete data.

Convolutional neural network with a joint Bayesian network methods used to classify faces, applied in one work to classifying vehicles.

## 1.4 Data Processing

Dataset had only 6,853 crashes. 2010 - 2014, in Louisiana, where the motorcyclist was at fault.

KABCO Injury Scale (this is weird)

**K** Fatality within 30 days

**A** Incapacitating injury, preventing the injured person from normal daily work.

**B** Non-Incapacitating Injury

**C** Possible/Compliant Injury, complains of pains or stresses with no physical evidence.

**O** No-injury crashes, also known as Property Damage Only (PDO)

Do we know whether the injuries were suffered by the motorcycle riders or the others, particularly pedestrians?

The discussion of median age of motorcyclists in crashes does not have any corresponding analysis of the ages of motorcycle riders, just a general "higher number of middle aged motorcycle riders." Where is the data?

"Sixty-two percent of crashes occurred more frequently on two-lane rural roadways that did not have a dividing barrier, which is similar to previous findings in the literature." What does it mean that 62% occurred more frequently?

"Motorcycle crashes were also more likely to occur on straight level roadways (74.9%), but it is important to note that 15.6% occurred on curve level roadways, a feature which has been shown to impact on crash severity." Yes, but can we compare those percentages with the percentage of miles of roadway classified as "straight level" versus "curve level"? How do those percentages for motorcycle crashes compare with car crashes?

The data lacks context.

## 1.5 Deep Learning

### 1.5.1 History

Perceptron was the first example of an artificial neural network (ANN), but failed to approximate nonlinear decision functions. Multilayer perceptron approximates nonlinear decision functions by stacking multiple layers of linear classifiers, and is considered "deep" if it has more than two layers.

### 1.5.2 Theory

Doesn't say what $\tilde{x}$ is, other than an estimation of $x^{(i)}$.

The equations don't make sense. Here's what they are, v/s what I think they should be.

$$
\begin{aligned}
z^{(i)} &= W_1 x^{(i)} + b_1 \\
\tilde{x}^{(i)} &= W_2 x^{(i)} + b_2 \\
J(W_i, b_i, W_2, b_2) &= \sum_{i=1}^{m} \left( \tilde{x}^{(i)} - x^{(i)} \right)^2 \\
&= \sum_{i=1}^{m} \left( W_2 x^{(i)} + b_2 - x^{(i)} \right)^2 \\
&= \sum_{i=1}^{m} \left( W_2 \left( W_1 x^{(i)} + b_1 \right) + b_2 - x^{(i)} \right)^2
\end{aligned}
$$

$$
\begin{aligned}
z^{(i)} &= W_1 x^{(i)} + b_1 \\
\tilde{x}^{(i)} &= W_2 z^{(i)} + b_2 \\
J(W_i, b_i, W_2, b_2) &= \sum_{i=1}^{m} \left( \tilde{x}^{(i)} - x^{(i)} \right)^2 \\
&= \sum_{i=1}^{m} \left( W_2 z^{(i)} + b_2 - x^{(i)} \right)^2 \\
&= \sum_{i=1}^{m} \left( W_2 \left( W_1 x^{(i)} + b_1 \right) + b_2 - x^{(i)} \right)^2
\end{aligned}
$$

Minimize the objective function, $J$, through stochastic gradient descent.

## 1.6 Model Development

Notes to myself so I can replicate the method.

1. Divided the dataset into training (50%), validation (25%), and test (25%) sets.
2. Made a cheap and dirty model (Model 1) with one epoch and two hidden layers.
   Note that the paper doesn't say what they did with the results of that model, whether they reworked their data or something.
3. For model 2, apply stopping criteria (what is this?), 100,000 epochs, 32 hidden layers.
4. Check when accuracy stops improving (after 100 epochs). Is that what the "stopping criteria" is, that they didn't actually go to 100,000 epochs?
5. For model 3, tuned with adaptive learning algorithm, tuning on $\rho$ and $\epsilon$, to balance the global and local search efficiencies. Is there a "use the adaptive learning rate algorithm" switch that they flipped, or did they do this by hand? Since they don't give their values for $\rho$ and $\epsilon$, I assume it's just a switch.
6. 100 epochs, 128 hidden layers, annealing rate $2 \times 10^{-6}$, 350,000 samples. I know what the epochs, hidden layers, and annealing rate are, but I'm not sure about the samples.
7. The "confusion matrix" has K, A, B, C, and O as both the rows and columns. The rows represent the true values, and the columns represent the classified values. Numbers on the main diagonal represent the number of each correctly classified, and everything else represents a misclassification.
8. Run the models on the validation set and the test set. I don't see how they used the validation set and test set differently, unless they used the validation set to train the adaptive learning rate algorithm.

## 1.7 Parts I need to Understand

- Stopping criteria
- Samples
- Root Mean Squared Error
- How to use the validation set and test set differently

## 1.8 Summary of Brad's Thoughts, and Questions

### 1.8.1 Basic Questions

1. In the abstract, Das claims that a deep learning model is not a machine learning algorithm. In what way is that true?
2. Does getting a misclassification rate of 0% in the training set under Model 3 likely indicate overfitting? Especially if the misclassification rate on the validation set goes up from Model 2 to Model 3?

### 1.8.2 Minor Critiques and Observations

These are mostly thoughts about how I should (or should not) model my papers after this one.

1. It seems that every paper has a cute name for its innovation, like *DeepScooter*.
2. This paper reads like a homework assignment for a first ML course.
3. That Das takes several pages to explain what deep learning is indicates either that the audience is not in ML but knows enough math to follow, or that the author needed to fill pages.
4. Doesn't anybody proofread papers before publication?
    - Four times, Dal refers to "text data," as in "training, validation, and text data," which I assume should be "test data."
    - On page 5, in equation (2), the $x$ should be a $z$, and this mistake propagates into equation (3).
5. I understood the paper on the first reading. That either indicates that I've learned more, that this paper covers an application that I understand (roads and motorcycles), so it's easier for me to grasp, or that the paper is really superficial.
6. In Figure 3a (page 10), the accuracy is only given to one digit of accuracy. As a reader, I'd like to know how different the 8%, 7%, and 6% are, and how different the three 6% values are. In the text, page 7, lower right, Das says that the misclassification rates for the validation set go from 5.79% in Model 2 to 6.25% in Model 3.
7. In the last paragraph of page 7, Das mentions for the first and only time that they have also applied a statistical model (multinominal logistic regression) for comparison. Should they not have mentioned this in the abstract? They give no details for how they got the 78% from the MLR.
8. DO NOT claim that your model "can estimate accuracy up to 100%" (page 11) if you're talking about the training data. Overfitting is not good.

### 1.8.3 Big Questions

1. The conclusions were known before the study. What has ML contributed?
2. The paper doesn't incorporate how motorcycle ridership is different from car drivership. Many (most?) motorcyclists are recreational riders. On weekends, car traffic goes down, and motorcycle traffic goes up. When the weather is bad, motorcycle ridership goes down, because most riders weren't going anywhere they needed to go, and most who had to go somewhere also have a car. A disproportionately large number of motorcycle miles are ridden by retired people, especially men. The ridership skews strongly male.
3. A useful study needs normalizing data. When comparing days of the week, you not only need the number of motorcycle crashes on those days, but also the number of miles ridden by motorcyclists on those days. Perhaps crashes per mile are actually lower on the weekends. Yes, 69.1% of crashes occurred in daylight, but whether that's high or low depends on what

percentage of motorcycle riding occurred in daylight. When comparing number of motorcycle fatalities in 2014 and 1997, also need a comparison of the number of motorcycle miles ridden in those years. This paper gives no basis for comparison.

4.

## 2 Papers I haven't Read Yet

Das, Avelar, Dixon, and Sun (2018)
(Das et al., 2019)

## References


Das, S., Avelar, R., Dixon, K., & Sun, X. (2018). Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis and Prevention*, *111*, 43-55. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0001457517304037` doi: https://doi.org/10.1016/j.aap.2017.11.016

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., & Jalayer, M. (2019). Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*, *23*(1), 30-48. Retrieved from `https://doi.org/10.1080/12265934.2018.1431146` doi: 10.1080/12265934.2018.1431146

Das, S., Dutta, A., Dixon, K., Minjares-Kyle, L., & Gillette, G. (2018). Using deep learning in severity analysis of at-fault motorcycle rider crashes. *Transportation Research Record*, *2672*(34), 122-134. Retrieved from `https://doi.org/10.1177/0361198118797212` doi: 10.1177/0361198118797212