

COGS 9 – Discussion Section A01 and A02

Kunal Rustagi (TA): Mon 9AM ([Zoom](#))
Connor McManigal (IA)
Yupei Sun(IA)



Part 1: Algorithmic Decision Making

Overview

The article discusses the accountability in algorithmic decision making, particularly in the context of automated content generation and algorithmic curation of news. It highlights the errors that can occur in machine-generated articles and raises questions about their impact and root causes. The article calls for serious consideration of how algorithmically informed decisions should be accountable to the public, especially in the face of errors, discrimination, or unfair denials of services.

What is Algorithmic Decision Making?

- Prioritize
 - Algorithms prioritize information to cope with the overwhelming amount of data, but this prioritization inherently involves discrimination.
- Classify
 - Classification decisions mark a particular entity as belonging to a given class by considering key characteristics of that entity. The opportunities for bias, uncertainty, or outright mistakes are plentiful in automated classification.
- Associate
 - Association decisions revolve around creating relationships between entities. These associations lead to connotations in their human interpretation.
- Filter
 - Filtering decisions involve including or excluding information according to various rules or criteria.

Algorithmic Transparency Standard

- Human Involvement
 - Involve explaining the goal, purpose, and intent of the algorithm, including editorial goals and the human editorial process or social context crucible from which the algorithm was cast.
- Data
 - Communicate the quality of the data, including its accuracy, completeness, and uncertainty, as well as its timeliness
 - How was it defined, collected, transformed, vetted, and edited
- Model
 - What the model actually use for input?
 - What are the weights put on the features and what is the rationale for weighting?
 - What were the tools used to do modeling?

Algorithmic Transparency Standard Continued

- Inferencing
 - The inferences made by an algorithm, such as classifications or predictions, often leave questions about the accuracy or potential for error. Creator can disclose some key characteristics.
 - What is the accuracy rate, and how many false positives versus false negatives are there?
- Algorithmic presence
 - Disclose if and when an algorithm is being employed at all, particularly if personalization is in use
 - For example: Is A/B test being used?

Conclusion

- Algorithmic transparency requires further research to determine what information should be disclosed and how it should be communicated.
- Algorithms, used in different cases and environments, may not have same level of disclosure and transparency
- Transparency goals include scrutability (comprehensive), trust, effectiveness, and efficiency, but balancing transparency with user experience is a challenge.
- Collaboration between disciplines like human-computer interaction, machine learning, and software engineering is crucial for achieving algorithmic transparency.

Q1: What does data transparency involve?

- A) Communicating the accuracy and completeness of the data.
- B) Ensuring the privacy and security of the data.
- C) Disclosing the steps taken to remediate errors in the data.
- D) All of the above.

Q1: What does data transparency involve?

- A) Communicating the accuracy and completeness of the data.
- B) Ensuring the privacy and security of the data.
- C) Disclosing the steps taken to remediate errors in the data.
- D) All of the above.**

Explanation: Data transparency includes communicating the quality, accuracy, completeness, and privacy implications of the data, as well as disclosing steps taken to remediate errors.

Q2: What are the potential implications of tuning classification algorithms to reduce one type of mistake?

- A) Increased computational complexity
- B) Greater objectivity in decision-making
- C) Privilege granted to different stakeholders
- D) Enhanced accuracy of associations

Q2: What are the potential implications of tuning classification algorithms to reduce one type of mistake?

- A) Increased computational complexity
- B) Greater objectivity in decision-making
- C) Privilege granted to different stakeholders**
- D) Enhanced accuracy of associations

Explanation: Tuning classification algorithms to reduce one type of error can increase the occurrence of the other type of error. This tuning process can grant privilege to different stakeholders and outcomes in a decision.



Part 2: Machine Bias

Risk Assessment

Risk Assessment: Algorithm used to inform decisions about who can be set free at every stage of the criminal justice system:

- Assigns bond amounts
- Makes fundamental decisions about the defendants' freedom
- Given to judges during criminal sentencing

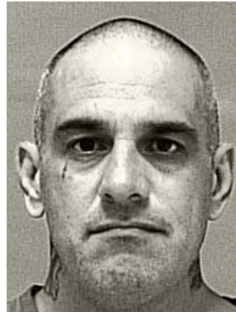
Used in Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington, and Wisconsin

- Justice Department's National Institute of Corrections encourages the use of these at every stage of the criminal justice process

Comparing the Cases of Borden and Prater

Prater:

- 41 year old repeat offender (armed robbery/attempted armed robbery/shoplifting)
- Multiple incidents of stealing
- Risk Score 3



VERNON PRATER
RISK3

Borden:

- 18 year old African American
Charged with burglary and petty theft
- Stole a scooter and bike
- Risk Score 8



BRISHA BORDEN
RISK8

ProPublica Study: Assessment of Risk Scores

Obtained 7,000 arrests from Broward County Florida to check and see how many were charged with new crimes over the next two years:

- Only 20% of people predicted to commit violent crimes actually went on to do so
- Algorithm was slightly better than chance at predicting future (less serious) charges
- Algorithm was 2x more likely to falsely flag black defendants as future criminals than white ones
- White defendants were more likely to be mislabeled with low risk scores

Statistical Analysis on Northpointe's Algorithm showed:

- Black defendants were still 77% more likely to be labeled as a higher risk of committing future violent crimes
- Black defendants were 45% more likely to be predicted to commit a future crime of any kind

Discussion Question: Talk Amongst Yourselves

Do you think risk assessment algorithms should be considered in the sentencing process? Why or why not? If not, what do you think they should be used for?



Should the algorithm even be considered for sentencing?

Boessenecker (who trains judges) warns his colleagues that the score doesn't necessarily reveal whether a person is dangerous or if they should go to prison

- Factors such as having a job or living in a home can carry more weight
- The algorithm has trouble with considering previous charges
- Northpointe's algorithm is often adopted without testing whether it works
- "The risk score alone should not determine the sentence of an offender."

Difficult to construct a score that doesn't include items that can be correlated with race - such as poverty, joblessness, and social marginalization

Discussion Question: Talk Amongst Yourself

What possible changes could be made to the algorithms(and/or the use of them in sentencing) to increase their effectiveness and predictive accuracy?



A Few Possible Solutions:

- Judges shouldn't use the algorithm when making decisions about sentencing, but instead use it for determining whether or not probation is an option
- Find other variables/measurements that don't possess multicollinearity and include them in the questions asked
- Adjust the weights of the questions accordingly:
 - Previous offenses and frequency of offenses should be the strongest indicators of a reoffender
 - Things like homelessness, joblessness, and economic status should carry less weight in determining a criminals risk score

Open Q/A + Public Office Hours

- If you weren't here for the attendance, you can come up now