COE TTHP
Center of Excellence for Technical Training & Human Performance

The UNIVERSITY of OKLAHOMA

LEARNER DATA MANAGEMENT

- **PRINCIPAL INVESTIGATORS: Dr. Christan Grant (cgrant@ou.edu), Dr. Dean Hougen (hougen@ou.edu)**
- **STUDENTS: Keerti Banweer, Santosh Malgireddy, Chenguang Xu**
- **SPONSOR: Shawn Mansfield**
- **TECHNICAL MONITOR: Dr. Hezekiah Braxton**

WHAT

This project applies text analytics, knowledge extraction, and machine learning to integrate data from FAA databases and transform it into usable information for more efficient and effective management of Aviation Safety.

GOAL

The goal is to empower FAA management with data-driven insights into connections between training and performance, reduce the data management burden, and improve the collection and analysis of learning management.

HOW

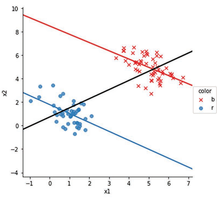
We create a prototype system that integrates existing FAA data sources, accommodates new data sources, allows appropriate access to data by a wide variety of users, and incorporates flexible analytics that produce insights into the data to encourage data-driven decision making.

WHY

The categorization of course documents allows the FAA to manage training requirements and create training plans for individual students. This helps the FAA to specify required courses and separate the electives, which helps the FAA academy to organize training courses.

DETECTING SIMPSON'S PARADOX

We implement methods to find Simpson's paradox anomalies in data. In this phenomenon, an association trend in the whole population reverses within subpopulations defined by a categorical variable. Detecting Simpson's paradox reveals surprising and interesting patterns of the data set for users.



Algorithm 1 Simpson's Paradox Detection Algorithm

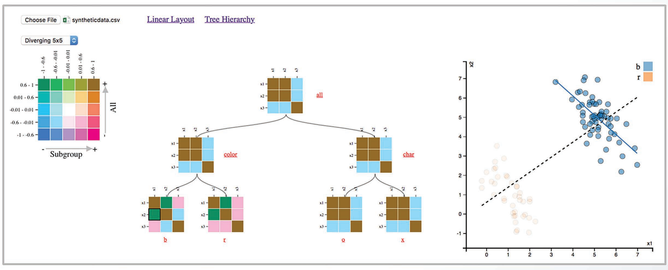
```
INPUT: Relational Table R
con_col ← detectTypes(R)
cat_col ← detectTypes(R)
for all (col1,col2) ∈ con_col do
  corrMatrix1 ← computeCorrelation(col1,col2)
end for
for col ← cat_col do
  subgroups ← R.groupby(col)
  for group ← subgroups do
    for all (col1,col2) ∈ con_col do
      corrMatrix2 ← computeCorrelation(col1,col2)
    end for
    if isReverse(corrMatrix1, corrMatrix2) then
      SP_result ← subgroup_info
    end if
  end for
end for
```

		Attribute 1	*Attribute 2*
Blue	*Attribute 1*	1.0000	−0.6190
	Attribute 2	−0.6190	1.0000
Red	*Attribute 1*	1.0000	−0.6160
	Attribute 2	−0.6160	1.0000

allCorr	attr1	attr2	revCorr	catAttr	subgroup
0.7710	attribute 1	attribute 2	−0.6190	color	b
0.7710	attribute 1	attribute 2	−0.6160	color	r

Chenguang Xu, Sarah M. Brown, Christan Grant. *Detecting Simpson's Paradox*. The 31st International Florida Artificial Intelligence Research Society (FLAIRS) Conference. Melbourne, Florida. 2018.

We develop a visualization using visual and interactive techniques to facilitate exploration of Simpson's paradox.



CLUSTERING COURSE DOCUMENTS

We categorize course documents based on document similarity and topics of interest for semantic analysis. Organizing course materials for air traffic controller training helps in prioritizing courses, discontinuing courses, and/or introducing new required courses. Using machine learning and text analytics, we cluster similar trainings together, which will allow us to categorize student skills. This helps us in further analysis of student performance and their strengths and weaknesses.

IMPACT

Our analysis suggests tracking student progress with detailed information of the scores each receives at each stage of training and in each course. This allows (1) an understanding of strengths and weakness of individual trainees, (2) finding areas where improvement is needed.

DISCUSSION

- The existing FAA databases are distributed across various locations and a large portion of the training details are stored locally.
- Student performance is stored as pass or fail for the training without details on strengths or weakness of trainees.

OU DAL

ENGINEERING

Evolution Adaptation Learning Robotics REAL

School of Computer Science
College of Engineering

Federal Aviation Administration