CHARLES Change-Aware Recovery of Latent Evolution Semantics in Relational Data

Shiyi He, Alexandra Meliou, Anna Fariha

THE UNIVERSITY OF UTAH

University of Massachusetts Amherst

How did my data change? Why?

name	gen	edu	exp	salary	bonus
Anne	F	PhD	2	$230,000	$23,000
Bob	M	PhD	3	$250,000	$25,000
Amber	F	MS	5	$160,000	$16,000
Allen	M	MS	1	$130,000	$13,000
Cathy	F	BS	2	$110,000	$11,000
Tom	M	MS	4	$150,000	$15,000
James	M	BS	3	$120,000	$12,000
Lucy	F	MS	4	$150,000	$15,000
Frank	M	PhD	1	$210,000	$21,000

2016 data

name	gen	edu	exp	salary	bonus
Anne	F	PhD	3	$230,000	$25,150
Bob	M	PhD	4	$250,000	$27,250
Amber	F	MS	6	$160,000	$17,440
Allen	M	MS	2	$130,000	$13,790
Cathy	F	BS	3	$110,000	$11,000
Tom	M	MS	5	$150,000	$16,400
James	M	BS	4	$120,000	$12,000
Lucy	F	MS	5	$150,000	$16,400
Frank	M	PhD	2	$210,000	$23,050

2017 data

Why is the **bonus** different in 2017? What happened?

I wonder why some people received a bonus increase, such as Anne got a $2150 increase, while others like Cathy didn't get any!

ChARLES explains changes in data

- Provides **semantic explanations**
- Each explanation consists of a series of **conditional transformations.**

Conditional Transformation

Condition → *Transformation*

Source dataset

Target dataset

Parameter Tuning

Diff Discovery Engine

Linear Model Tree (LMT)

Fit a linear regression line in each leaf

Ensemble

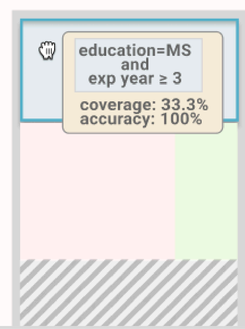
Multiple LMTs

Explanations

Ranker

Show the top 10 explanations

	Explanation		Details	Score	Accuracy	Interpretability
1	education=PhD	bonus_new = 1.05 X bonus_old + 1000	Interactive Visualization	89%	100%	78%
	education=MS and exp year ≥ 3	bonus_new = 1.04 X bonus_old + 800				
	education=MS and exp year < 3	bonus_new = 1.03 X bonus_old + 400				
2	education=PhD	bonus_new = 0.1 X salary + 0.01 X bonus_old + 1000		81%	97%	65%
	gender=F and education=MS	bonus_new = 0.1 X salary + 0.01 X bonus_old + 800				
	gender=M and education=MS	bonus_new = 0.13 X salary + 0.01 X bonus_old - 3175				



ChARLES balances accuracy and interpretability

Everyone receives about 6% increase

Less Accurate

More Interpretable

If PhD → *5% increase + $1000*

If MS and served at least 3 years → *4% increase + $800*

If MS and served less than 3 years → *3% increase + $400*

Reasonably Accurate

Reasonably Interpretable

If PhD and Female → *5% increase + $1000*

If MS and Female and served at least 3 years → *4% increase + $800*

If MS and Male and served at least 3 years → *4% increase + $700*

If BS and Female and served less than 3 years → *3% increase + $400*

If BS and Male and served less than 3 years → *3% increase + $300*

More Accurate

Less Interpretable

Looks like the company opted for a policy to reward **long-serving** employees and promote **educational advancement**.