BERT-Sort : A Zero-shot MLM Semantic Encoder on Ordinal Features for AutoML

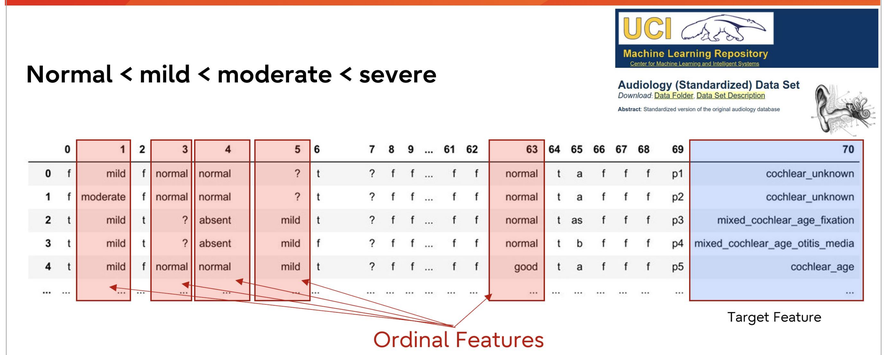
Mehdi Bahrami, Wei-Peng Chen, Lei Liu, Mukul Prasad
Fujitsu Research of America

FUJITSU

AutoML-Conf 2022

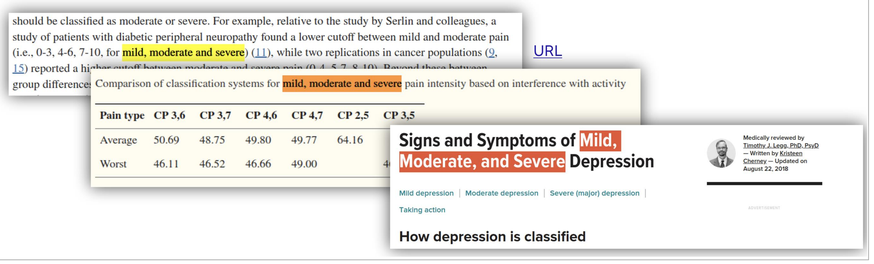
Background

Normal < mild < moderate < severe



Unsupervised Approach

- Processing large context to understand the order of context



Language Modeling

To identify empirically-derived cutoffs for mild, moderate, and severe pain

often developed based upon categorical ratings of pain (e.g., mild, moderate, severe).

reported a higher cutoff between moderate and severe pain

$$P(w_t|context)\ \forall t \in V$$

$$P(w_t|w_{t-k}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+k})$$

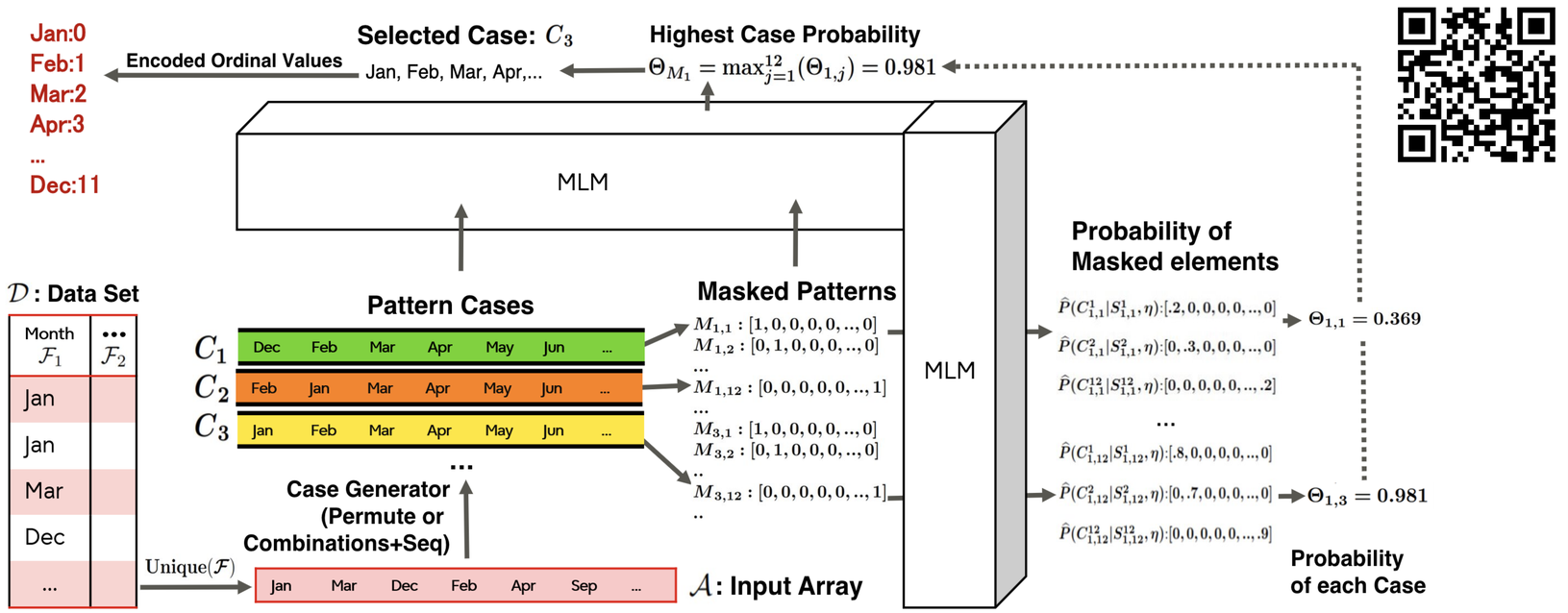
Evaluation

$$\Theta_{i,j} = \begin{cases} \frac{\sum_{k=1}^{||\mathcal{A}_i||} \widehat{P}(C_{i,j}^k|S_{i,j}^k,\eta)}{||\mathcal{A}_i||}, & \text{if } C_{i,j}^k \in \mathcal{W}_\eta \\ 0, & \text{otherwise} \end{cases}$$

$$Ord_{Acc} = \sum_{i=1}^{||A||} \frac{||A|| - |\mathcal{L}_i - \widehat{\mathcal{L}_i}|}{||A||^2}$$

#	Ranked Values	Acc	Ord_{Acc}
1	$[Feb \prec Jan \prec Mar \prec Apr]$	0.5	0.87
2	$[Mar \prec Feb \prec Jan \prec Apr]$	0.5	0.75

BERT-Sort



Semantic Evaluation

A Comparisons of semantic ordinal value evaluation of BERT-Sort with initiation on 4 different MLMs of DistilBERT, RoBERTa, XLM, BERT-base-uncased, and OrdinalEncoder

Evaluation	Ord_{Acc}					Acc				
Approach	BERT-Sort				OrdinalEncoder	BERT-Sort				OrdinalEncoder
Model Feature	M_1	M_2	M_3	M_4		M_1	M_2	M_3	M_4	
#champions	30	**35**	31	32	3	23	**31**	27	28	5
	w.r.t. OrdinalEncoder				w.r.t. M_2	w.r.t. OrdinalEncoder				w.r.t. M_2
Improvement	0.20	**0.27**	0.25	0.20	baseline	0.31	**0.55**	0.49	0.34	baseline

Multilingual Multi-Domain BERT Sort

#	Input	Model	BERT-Sort (Top 1)	OrdinalEncoder
2	[Lava Hot, Hot, Boiling Hot]	RoBERTa-large	[Hot<Boiling Hot <Lava Hot]	[Boiling Hot<Hot<Lava Hot]
3	[Eight, Four, Two, Six, Twelve]	RoBERTa-large	[Two < Four< Six< Eight < Twelve]	[Eight<Four<Six< Twelve<Two]
4	[Low, Medium, High]	RoBERTa-large	[Low < Medium < High]	[High < Low < Medium]
6	[Leukemia, Cancer, Melanoma]	RoBERTa-large	N/A	[Cancer < Leukemia < Melanoma]
7	[Leukemia, Cancer, Melanoma]	BioClinical BERT	[Melanoma < Leukemia < Cancer]	[Cancer < Leukemia < Melanoma]
8	[優れた, 貧しい, 良い]	Japanese BERT-MLM	[貧しい < 良い < 優れた]	[優れた <良い < 貧しい]
9	[Muy Buena, Normal, Buena]	Spanish BERT-MLM	[Normal < Buena < Muy Buena]	[Buena < Muy Buena < Normal]
10	[差, 好, 优秀]	Chinese BERT-WWM	[优秀 < 好 < 差]	[优秀 < 好 < 差]

AutoML Evaluation

Overall average F1 score and average Accuracy score performance of 8 original data sets, and its 4 other methods of ordinal value encoders on 4 AutoML platforms of AutoGluon, FLAML, H2O, and MLJAR with 4 different randomization experiences

Method	F1 Score	Accuracy Score
Encoded BERT	0.520	0.728
OrdinalEncoder	0.615	0.764
Original	0.625	0.769
BERT-Sort	**0.636**	**0.784**
Human Annotation	0.637	0.785

Performance of 11 ML algorithms on encoded the original UCI Car Evaluation data set through: i) BERT-Sort Encoder and ii) OrdinalEncoder