Fine-grained Attention in Hierarchical Transformers for Tabular Time-series

KDD'24
10th MiLeTS Workshop

Raphael Azorin, Zied Ben Houidi, Massimo Gallo,
Alessandro Finamore, Pietro Michiardi

Correspondence
azorin@eurecom.fr

Motivation

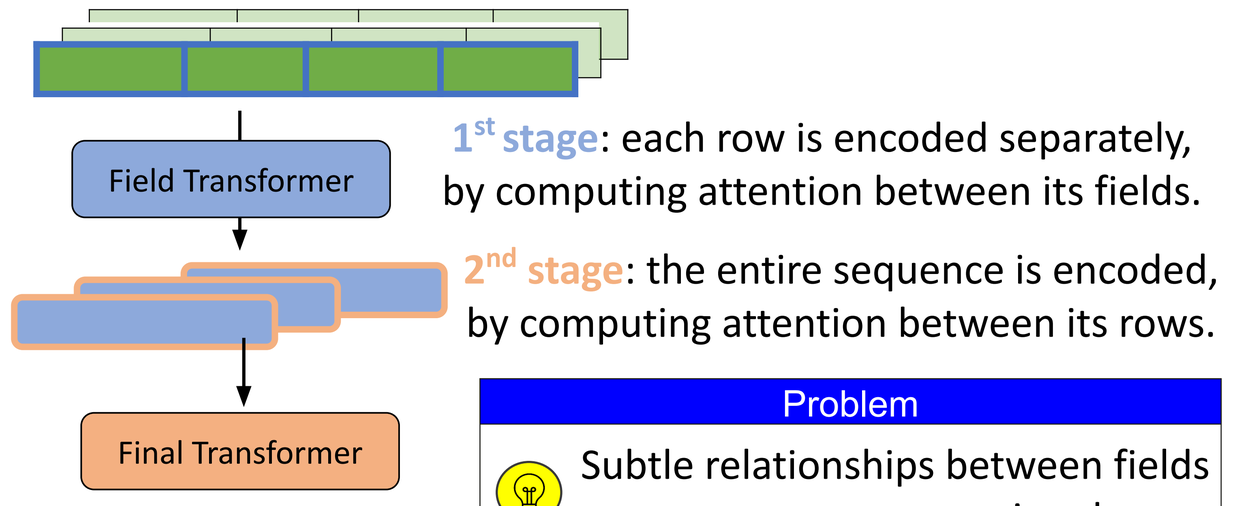
Hierarchical models struggle to capture relationships between fields across rows

Tabular time-series

Timestamp	Feature A	Feature B	Feature C

Unlike standard tabular data,
where each row corresponds to a single record,
a tabular time-series spans *multiple* time-dependent rows.

Hierarchical modeling - TabBERT [1]



1st stage: each row is encoded separately, by computing attention between its fields.

2nd stage: the entire sequence is encoded, by computing attention between its rows.

Problem

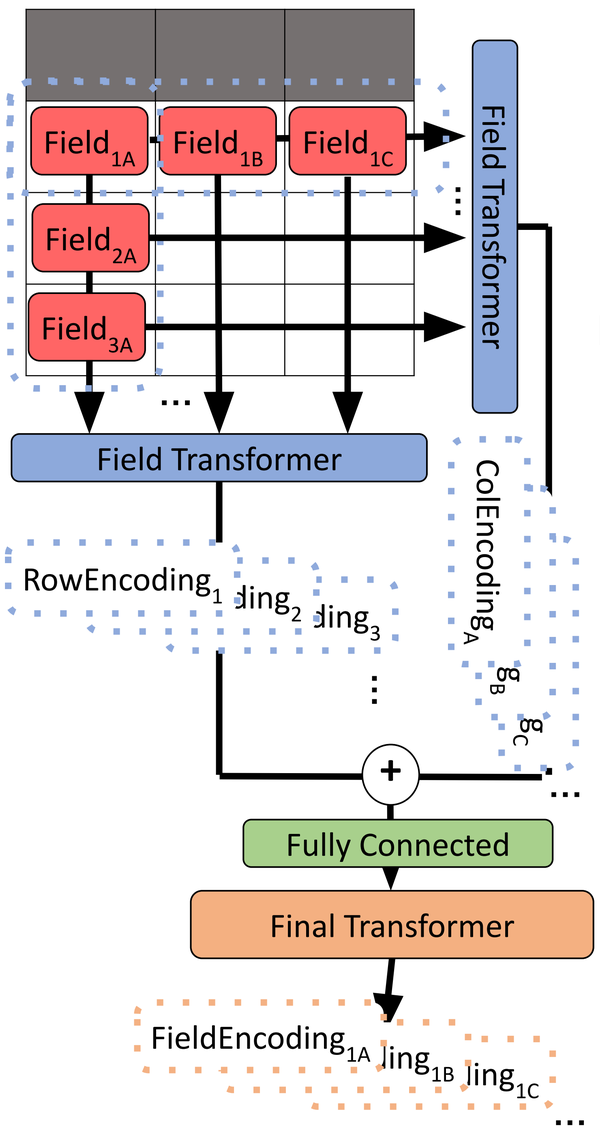
Subtle relationships between fields *across* rows are missed

Proposal

Fieldy learns relationships between fields across rows and columns

Fieldy is a novel hierarchical model that learns fields representations contextualized by both axes simultaneously.

Field-based hierarchical architecture



1st stage:
each row and each column
is encoded separately,
by computing attention between its fields [2].

Intermediate:
intersections are concatenated
and passed through a linear layer
to learn contextualized field
representations.

2nd stage:
the entire sequence is encoded,
by computing attention
between its contextualized fields.

Benefits and limitations

✓ Each field is related to every other field, to learn fine-grained interactions.

✓ Each field is contextualized by both row and column.

✗ Lack of table structure information

↓

Positional embeddings to inform on each field's column index and row position.

✗ Increased parameters and computations

↓

Reduce the size of the Field Transformers and near-linear attention optimization.

Results

Fieldy outperforms hierarchical SoTA

Evaluation on regression and classification tasks.

Model	Pollution [3] RMSE ↓	Loan [4] Avg. Precision ↑
XGBoost	50.74	0.36
TabBERT (SoTA)	21.05	0.46
Fieldy (ours)	**20.13**	**0.48**

Future work

Further evaluation and extensions

Evaluate *Fieldy* on more tasks and larger datasets, e.g., click-through rate, multivariate time-series.

Augment *Fieldy* with more sophisticated embeddings techniques and pre-training objectives.

[1] TabBERT [Padhi et al., ICASSP 2021]
[2] Tabbie [Iida et al., arXiv:2105.02584 2021]
[3] Beijing pollution prediction [Chen, UCI Repository 2019]
[4] Loan default prediction [Berka, KDD 1999]