Estimating Train Delays in a Large Rail Network

Ramashish Gaurav, Biplav Srivastava

June 2018

Problem

- Given a train and its route information, predict the delay in minutes at an in-line station during its journey on a valid date.

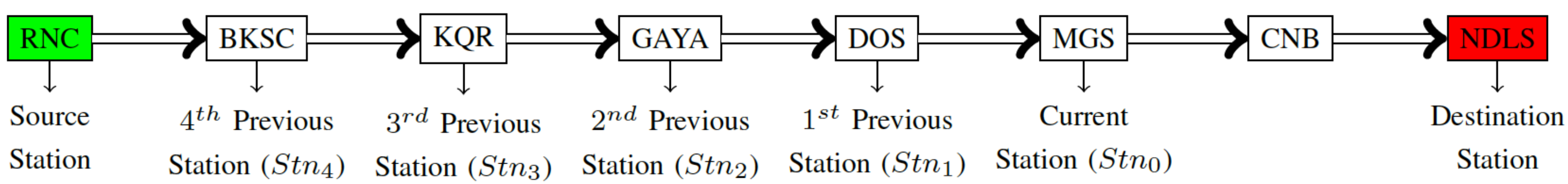


Fig. 2. Train Route of Train 12439. The above figure shows the route of train 12439 which starts at the station RNC and ends at the station $NDLS$. For current station MGS, 4 previous stations are considered; whose information we can use for preparing a 4-$prev$-stn data-frame (Table III). Stn_i notation for i^{th} previous station is used throughout this paper.

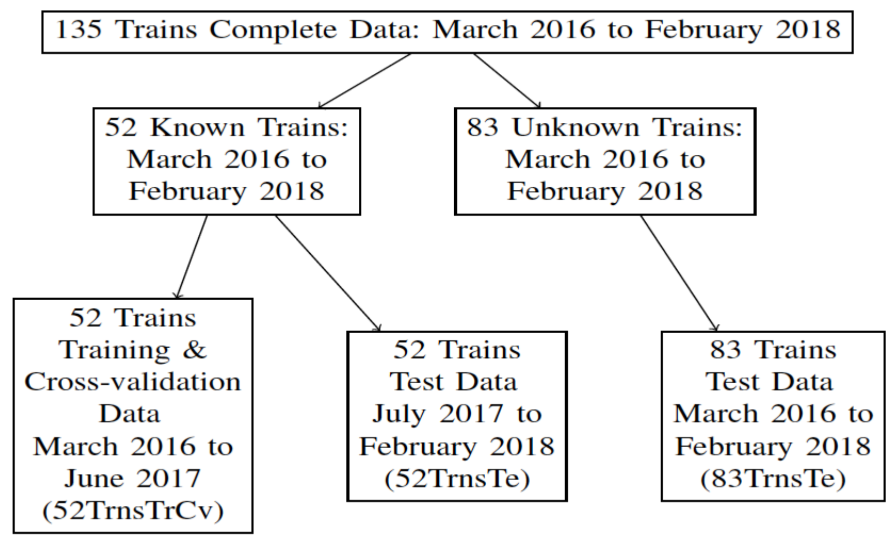
Contributions

- Collected data for two years for135 Indian trains' running status information (which captures delays along stations).
  - Focus is on trains touching the busy Mughal Sarai station (MGS; renamed Deen Dayal Upadhyay Station in 2018)
- Studied delays using n -order Markov Process Regression models to find the correct order of the Markov Process. Most of the 135 trains follow 1-order Markovian Process.
  - Do Akaike Information Criterion (AIC) and Schwartz Bayesian Information Criterion (BIC) analysis to find the correct order of the Markov Process.
  - Information useful to build accurate models with just the right data needed.
- Built a scalable, train-agnostic, accurate and Zero-Shot competent framework for predicting train arrival delays
  - learning from a fixed set of trains and transferring the knowledge to an unknown set of trains.
  - Able to predict late minutes at in-line stations of Known Trains with 62% accuracy within 95% CI
  - Method works for unknown trains (i.e., whose delay data is not known) but accuracy is lower
- Discuss how the train-agnostic framework can leverage different types of trained models and be deployed in real time to predict the late minutes at an in-line station.

Train Data Set

Data Statistics for 135 Trains Complete Data

Total number of trains considered	135
Total number of unique stations covered	819
Maximum number of journeys made by a train	334
Average number of journeys made by a train	48
Maximum number of stations in a train's route	129
Average number of stations in a train's route	30



Sample Performance of Prediction Model

TABLE V
PREDICTED LATE MINUTES FOR *Known Train* 22811 TEST DATA (OBTAINED FROM 4-OMLMPF WITH RFR MODELS)

Stations:	BBS	CTC	JJKR	BHC	BLS	KGP	BQA	ADRA	GMO	KQR	GAYA	MGS	CNB	NDLS
Actual Late Minutes:	0	2	8	-1	13	25	19	18	2	9	-21	-5	6	15
Predicted Late Minutes:	0	2.75	6.83	0.01	17.44	16.52	11.22	17.65	1.94	16.01	-8.77	-0.25	12.26	23.10

TABLE VI
PREDICTED LATE MINUTES FOR *Known Train* 12326 TEST DATA (OBTAINED FROM 4-OMLMPF WITH RFR MODELS)

Stations:	NLDM	ANSB	RPAR	SIR	UMB	SRE	MB	BE	LKO	BSB	MGS	PNBE	KIUL	JAJ	JSME	ASN	KOAA
Actual Late Minutes:	0	3	4	-11	0	-6	15	55	30	10	18	10	11	0	7	3	5
Predicted Late Minutes:	0	9.38	7.87	-2.43	3.61	0.50	26.13	36.14	29.42	32.14	20.38	3.296	6.87	-3.80	17.55	14.30	13.91

TABLE VII
PREDICTED LATE MINUTES FOR *Unknown Train* 12356 TEST DATA WITH 3 *Unknown Stations* (OBTAINED FROM 4-OMLMPF WITH RFR MODELS)

Stations:	JAT	PTKC	JRC	LDH	UMB	SRE	MB	BE	LKO	RBL	JAIS	AME	PBH	BOY	BSB	MGS	DNR	PNBE	RJPB
Actual Late Minutes:	0	8	3	0	-5	-15	-10	-1	30	41	51	57	74	111	75	123	130	120	120
Predicted Late Minutes:	0	10.19	10.74	10.17	11.60	11.97	27.24	34.63	28.45	40.15	41.29	42.94	60.71	72.51	75.25	70.50	74.45	67.95	71.80

Future Work

- Expand the data collection and extend the analysis to trains India-wide
- Explore other approaches like time series prediction and neural networks. In particular, Recurrent Neural Networks (RNN) with their property of memorizing past details and predicting the next state. changing behavior of railway network and delays.
- Predict delay of trains in other countries