Automated Data Labeling for Machine Learning Based Short-Term Voltage Stability Classification

K.D Dharmapala and A.D Rajapakse

University of Manitoba

Winnipeg, MB

Canada

Contents

1. Introduction

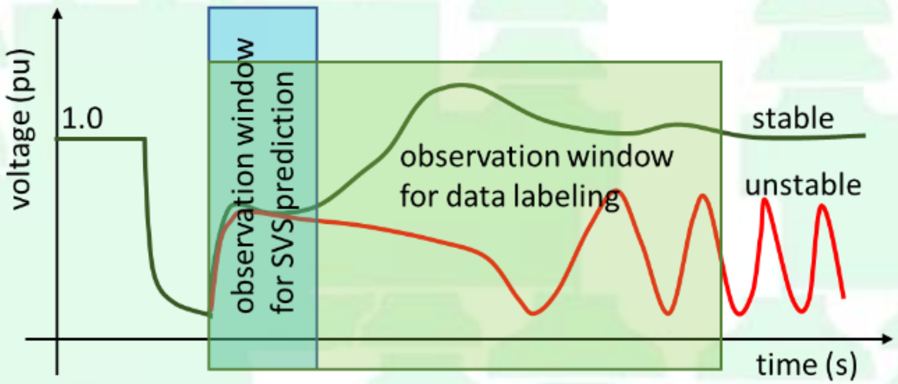
- Voltage stability is the ability of a power system to restore an acceptable steady state voltage at all the buses after the system been subjected to a disturbance [1].
- Short-term Voltage Stability (SVS) occurs within few seconds after the initiation disturbance.
- Modern power systems encounter more SVS issues due to:
  - ❖ Increasing demand of dynamic loads (eg: induction motors, power electronic loads)
  - ❖ High penetration of Inverter Based Resources (IBR) [1]-[4].
- Therefore, researchers have attempted to develop techniques to assess SVS status using real-time measurements.

1. Introduction

- Approaches for SVS status assessment:
  - ❖ Control theory based
  - ❖ Voltage curve based
  - ❖ Stability boundary based
  - ❖ Machine Learning (ML) based
- ML-based classification models show promising results
  - ❖ However, to use supervised learning, a large set of training data in the form of inputs and the corresponding output is required.

1. Introduction

- **Problem statement:** Label generation according to domain knowledge by an expert is a cumbersome process which inevitably consumes significant amount of engineering time, and it may cause human errors while labeling.
  - ❖ Note that data labeling differs from real time SVS status prediction
  - ❖ Data labeling is an off-line process and can use longer observation window to determine SVS status



1. Introduction

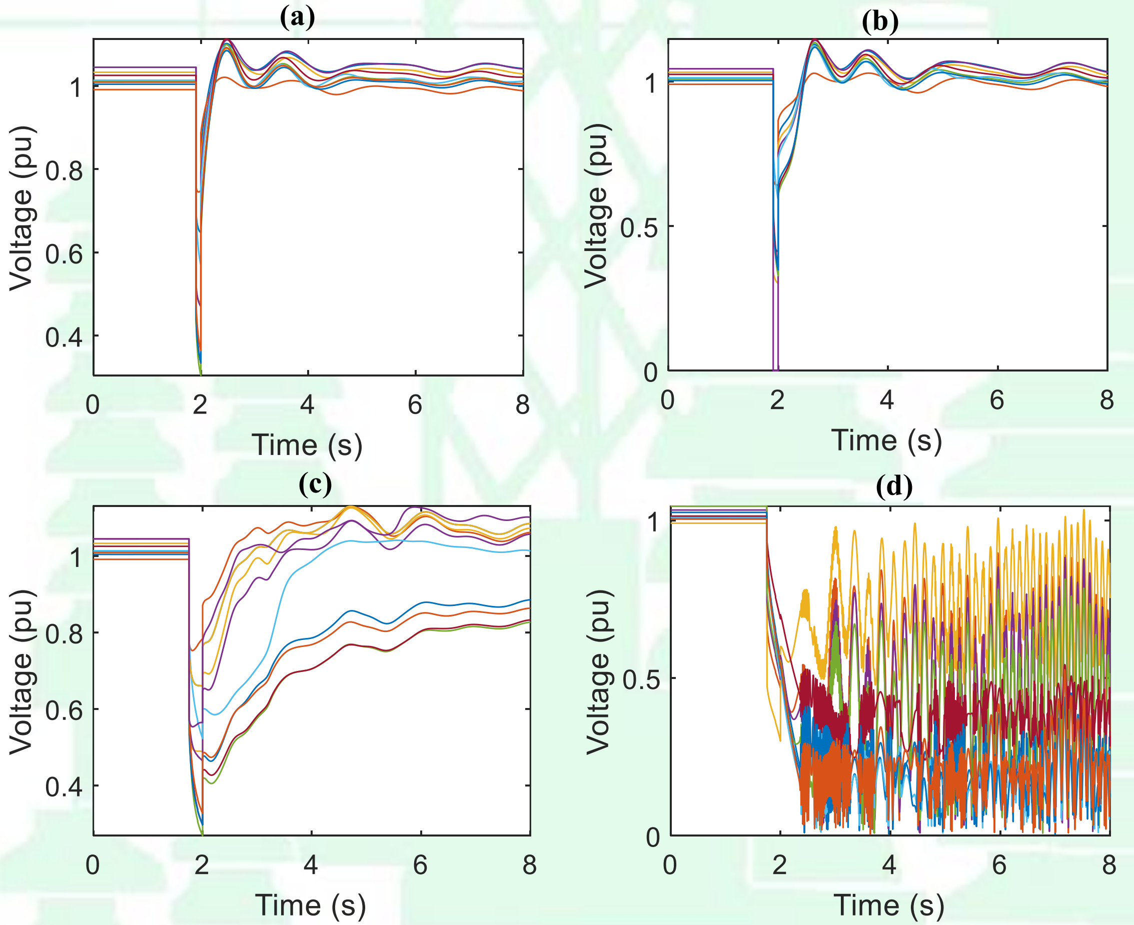
- **Hypothesis**: A semi-supervised learning algorithm can be adopted to generate labels where unlabeled data instances can be labeled using a small portion of labeled data instances as a guide.
- **Aim:** Make the label generation more efficient and accurate.
- **Proposed Approach:** Application of a graph based semi-supervised learning algorithm named “Label Propagation” with multiple features indicating SVS status.

2. Label Generation for SVS Assessment

- The voltage trajectory after a disturbance could be a:

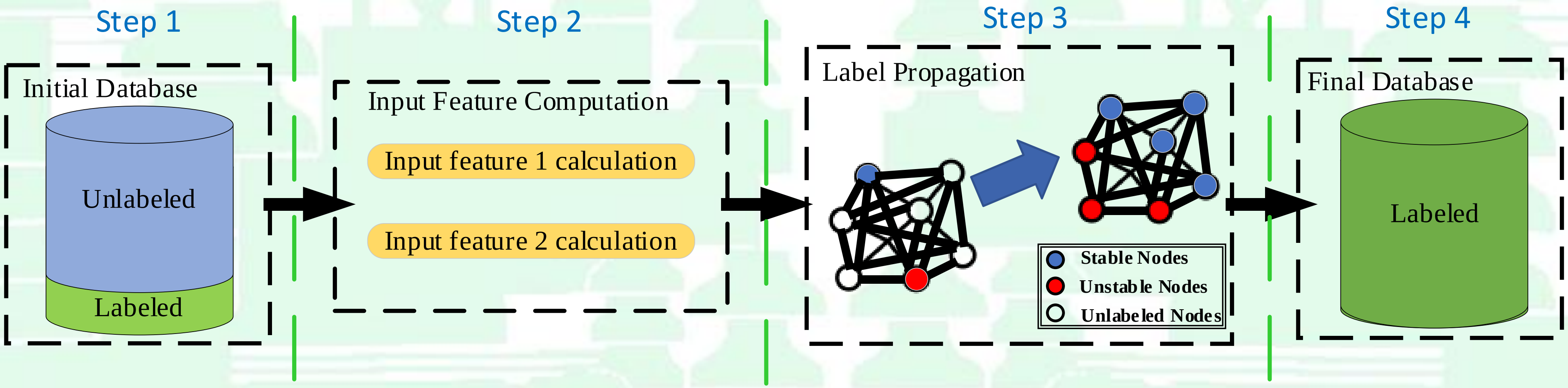
Stable
- a) Fast voltage recovery
- b) Fault Induced Delayed Voltage Recovery (FIDVR)

Unstable
- c) Sustained low voltage without recovery
- d) Fast voltage collapse



3. Proposed Automatic Label Generation Scheme

Step	Task
Step 1	Initial database generation (with a small portion of labeled data)
Step 2	Computation of candidate input features
Step 3	Perform label propagation (semi-supervised learner)
Step 4	Generate a fully labeled database



3. Proposed Automatic Label Generation Scheme

3.1 Initial data base generation

- Dataset contains time series of bus voltage magnitudes of test system for different disturbances under different operating conditions.
- Dataset should cover all credible contingencies and credible operational conditions.
- In this study, 700 voltage trajectories under different conditions were generated using automated PSSE® dynamic simulations.
- Afterwards, a small portion of voltage trajectories are labeled manually by observing the plots of voltage trajectories.
- The labeled data instances have logical variable which indicates whether the data instance is stable, or unstable (-1 for unstable / 1 for stable) and unlabeled data instances have null values as the logical variable.

3. Proposed Automatic Label Generation Scheme

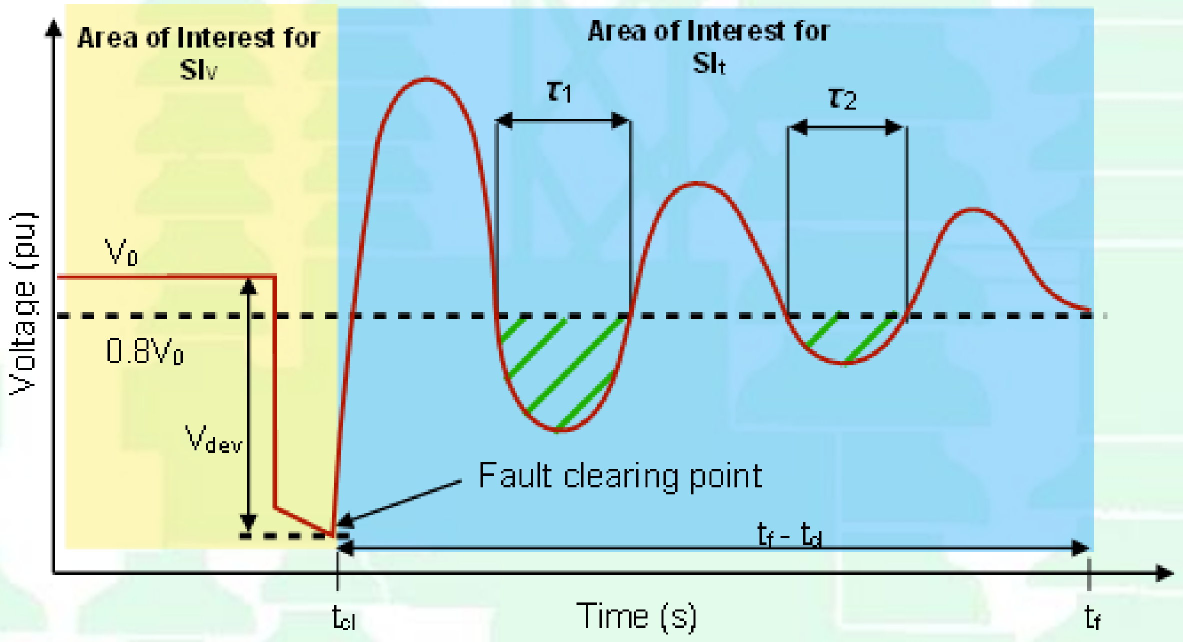
3.2 Input Feature Computation

(a) NERC Contingency Severity Index [5]

$$SI_v = \begin{cases} \dfrac{|V_0 - V_{dev}|}{V_0}, & \dfrac{|V_0 - V_{dev}|}{V_0} \geq \gamma \\ 0, & \dfrac{|V_0 - V_{dev}|}{V_0} < \gamma \end{cases}$$

$$Where\ \gamma = \begin{cases} 0.25 & : load\ buses \\ 0.3 & : other\ buses. \end{cases}$$

$$SI_t = \frac{\sum_{i=1} \tau_i}{t_f - t_{cl}}$$



3. Proposed Automatic Label Generation Scheme

3.2 Input Feature Computation

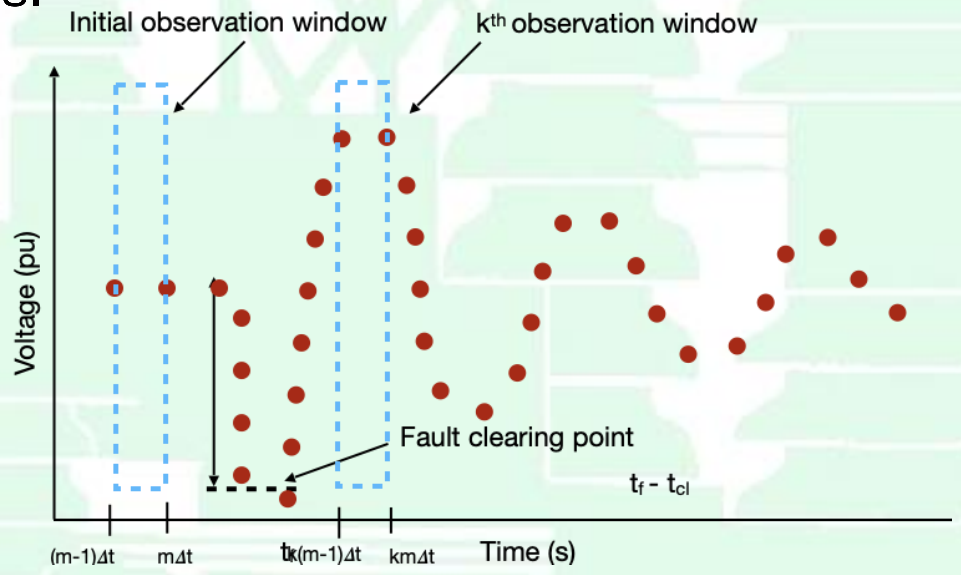
(b) Lyapunov Exponent [6]

- The Lyapunov Exponent (LE) is adopted from ergodic theory. LE has the potential to determine the chaotic nature of the system at any moment by analyzing the rate of divergence of the relevant dynamic system variables.

$$\lambda(k\Delta t) = \frac{1}{Nk\Delta t} \times \sum_{m=1}^{N} log \frac{\left\| V_{(k+m)\Delta t} - V_{(k+m-1)\Delta t} \right\|}{\left\| V_{m\Delta t} - V_{(m-1)\Delta t} \right\|}, k > N$$

Where integer N is chosen such that

$$\epsilon_1 < \left\| V_{m\Delta t} - V_{(m-1)\Delta t} \right\| < \epsilon_2 \; for \; m = 1,2,\ldots,N.$$



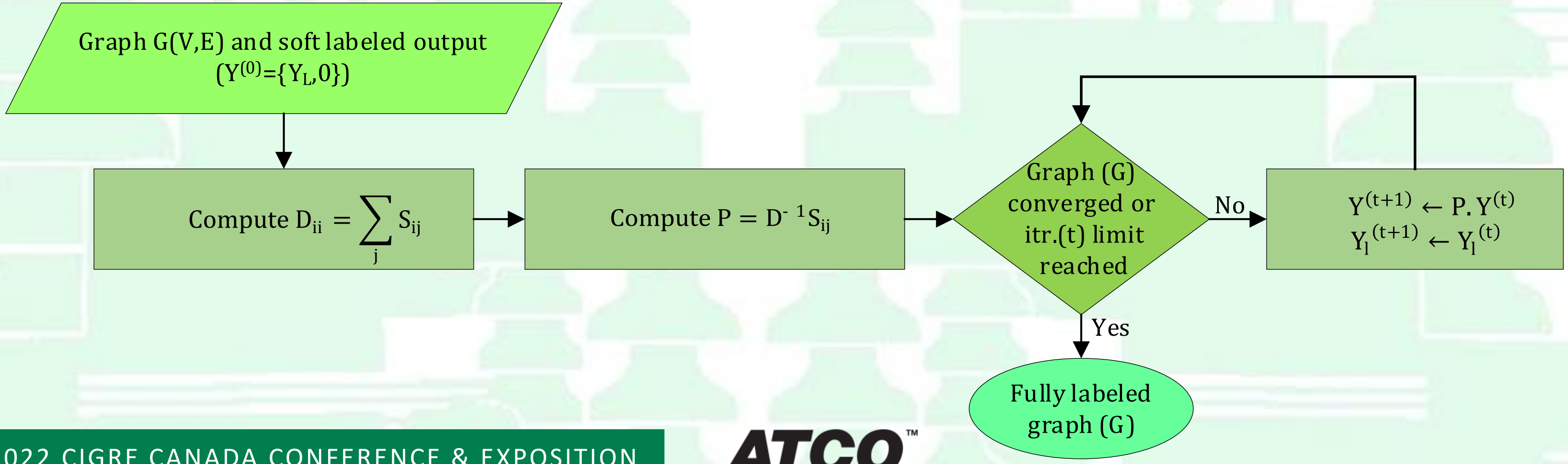
3. Proposed Automatic Label Generation Scheme

3.3 Semi-supervised Label Propagation

- Label Propagation is a graph based semi supervised learning algorithm [7].

$X=\{(x_1^1 .. x_N^1),...,(x_1^n .. x_N^n), .. ,(x_1^M .. x_N^M)\}$,with M number of features and N number of nodes.

$$S_{ij} = \exp\left(-\frac{\sum_{n=1}^{M}\left(x_i^n - x_j^n\right)^2}{\sigma^2}\right). \qquad P_{ij} = P(j \to i) = \frac{S_{ij}}{\sum_{k=1}^{L+U} S_{kj}}$$

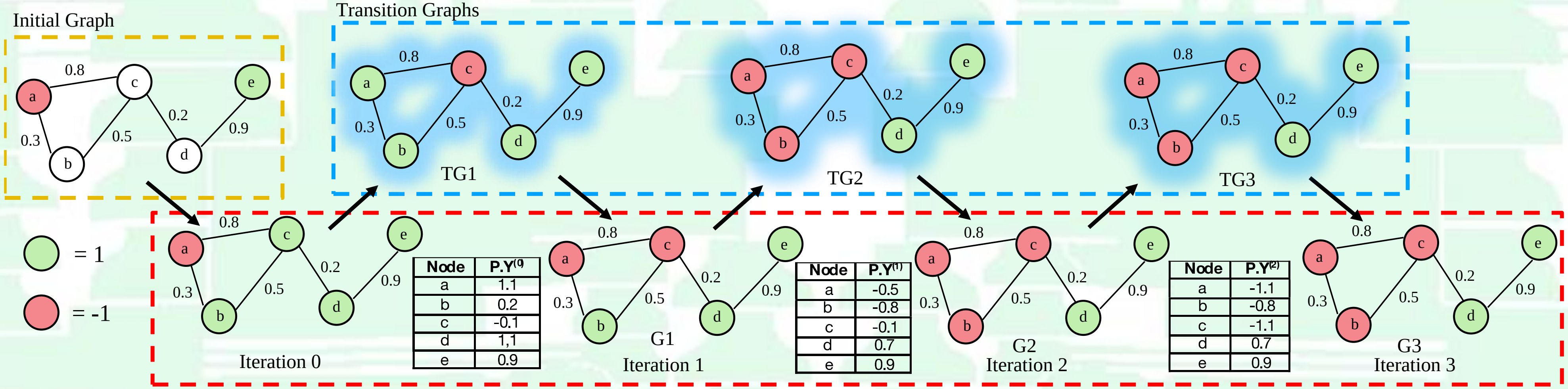


3. Proposed Automatic Label Generation Scheme

3.3 Semi-supervised Label Propagation (example)

- Assume an initial graph with nodes ***a, b, c, d*** and ***e***. Node ***a*** and ***e*** are labeled with class “red” and “green” which are numerically represented by “-1” and “1” respectively and other nodes are unlabeled.

- The respective P values are denoted on each vertex. P values of nodes which are not connected are assumed to be negligible.



Node	$P.Y^{(0)}$
a	1.1
b	0.2
c	-0.1
d	1,1
e	0.9

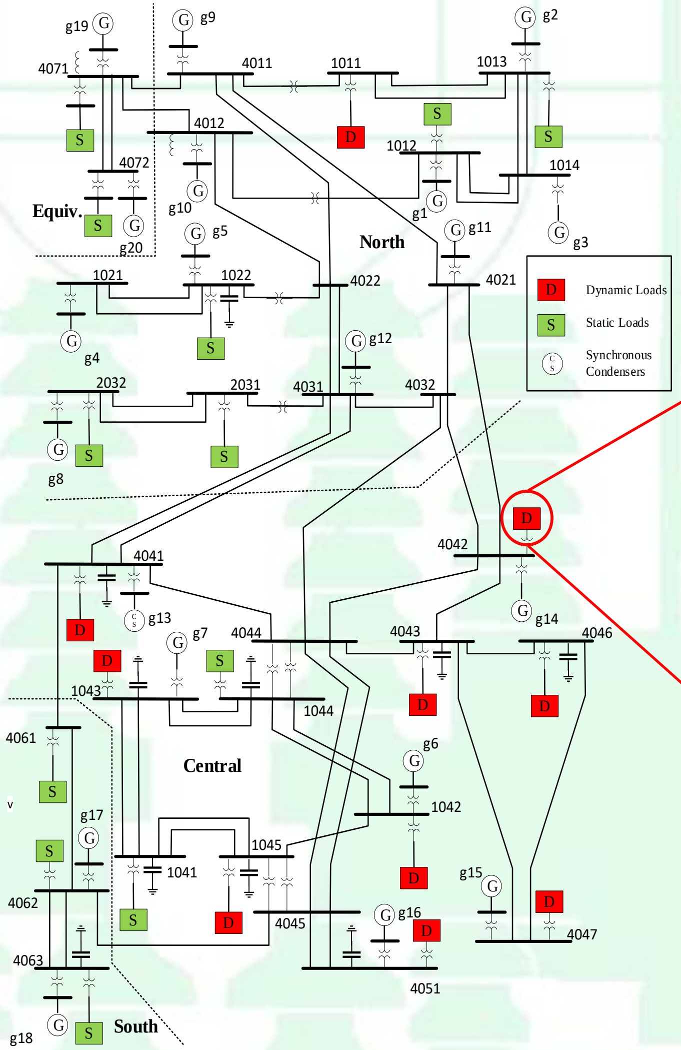
Node	$P.Y^{(1)}$
a	-0.5
b	-0.8
c	-0.1
d	0.7
e	0.9

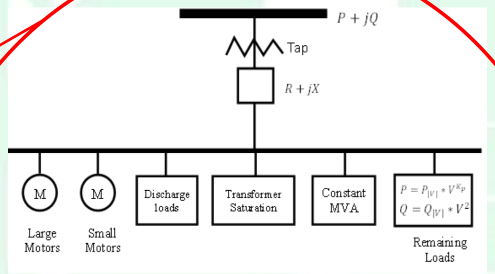
Node	$P.Y^{(2)}$
a	-1.1
b	-0.8
c	-1.1
d	0.7
e	0.9

4. Case Study

Test System

- IEEE 32 bus Nordic test system [8].
- Dynamic loads such as induction motors are the main contributor of short-term voltage instability after faults [9].
- Therefore, some of the static loads at random locations were replaced by dynamic loads.
- In this study, Complex Load Model or CLOD was used as the dynamic load model [10].





Loading Scenario	Load Type			
	Large motor	Small motor	Constant MVA	Remaining loads
Scenario A	0%	0%	100%	10%
Scenario B	5%	10%	50%	35%
Scenario C	10%	10%	40%	40%
Scenario D	10%	20%	30%	40%
Scenario E	30%	20%	10%	40%

5. Results

- All 700 data instances were labeled manually for the purpose of validation.

$$TA = \frac{Number\ of\ correctly\ labled\ data\ instances}{Number\ of\ unlabeled\ data\ instances} \times 100\%$$

SVS Indices used as input features	TA (%) under different percentages of labeled data		
	10%	**20%**	**30%**
NERC CSI (SI_v & SI_t)	80.6	82.5	82.8
Lyapunov Exponent (λ)	94.4	98.5	98.5
All indices (SI_v , SI_t & λ)	98.2	99.5	99.5

6. Conclusions

- The results show that the proposed automated data labeling approach can more accurately label the input data compared to labeling data using a single indicator such as SI_v , SI_t or λ.
- When the percentage of manually labeled data instances increases, the accuracy of labeling process has increased.
- Furthermore, when all the indices are considered as input features the level of accuracy increases since the contribution of each index will be ensembled.
- In this study, using three SVS indices (SI_v , SI_t & λ) as input features, labels are assigned with an accuracy of 99.5 % when only 20% of the data are manually labeled.
- This semi-supervised learning process can be used for other applications, for example to screen the results of automated contingency simulations done for large networks which are currently done manually.

Bibliography

[1] P. Kundur, J. Paserba, M. Electric, P. Products, and N. D. Hatziargyriou, “Definition and Classification of Power System Stability IEEE/CIGRE Joint Task Force on Stability Terms and Definitions,” IEEE Trans. Power Syst., vol. 19, no. 3, pp. 1387–1401, 2004.

[2] J. A. Diaz de Leon and C. W. Taylor, "Understanding and solving short-term voltage stability problems," IEEE Power Engineering Society Summer Meeting, 2002, pp. 745-752 vol.2, PESS.2002

[3] C. Dwivedi, "Literature Survey on Short-Term Voltage Stability Effect, Cause and Control," 2018 IEEE Green Technologies Conference (GreenTech), 2018, pp. 15-20.

[4] A. Alshareef, R. Shah, N. Mithulananthan and S. Alzahrani, "A New Global Index for Short Term Voltage Stability Assessment," in IEEE Access, vol. 9, pp. 36114-36124, 2021.

[5] NERC/WECC. Planning Standards. 2003. Available online: https://www.scribd.com/document/81304318/WECC-NERC-Planning-Standards

[6] Dasgupta, S.; Paramasivam, M.; Vaidya, U.; Ajjarapu, V. Real-time monitoring of short-term voltage stability using PMU data. IEEE Trans. Power Syst.2013,28, pp. 3702–371.

[7] Xiaojin Zhu and Zoubin Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation.

[8] Test Systems for Voltage Stability Analysis and Security Assessment. Accessed: Jun. 2020. [Online]. Available:https://resourcecenter.ieee-pes.org/publications/technical reports/PESTR19.html

[9] K. Wagstaff, S. Rogers, and S. Schroedl, “Constrained-means clustering with background knowledge,” in Proc. 8th Int. Conf. Machine Learning, 2001, pp. 577–584.

[10] Comprehensive Load Modeling for System Planning Studies. EPRI, Palo Alto, CA: 2009.1015999

Thank You !