

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Motivation | 3 |
| 1.2 | General Objective | 3 |
| 1.3 | Specific Objectives | 4 |
| 1.4 | Resources Available for this Thesis | 4 |
| 2 | Methodology | 5 |
| 3 | Theoretical Background | 7 |
| 3.1 | Machine Learning | 7 |
| 3.1.1 | Classification and Regression Problems | 7 |
| 3.1.2 | Supervised and Unsupervised Learning | 8 |
| 3.1.3 | Semi-Supervised Learning | 8 |
| 3.2 | Deep Learning | 9 |
| 3.2.1 | Traditional Neural Networks | 9 |
| 3.2.2 | Recurrent Neural Networks | 9 |
| 3.2.3 | Training of Neural Networks | 11 |
| 3.3 | Variational Inference | 13 |
| 3.4 | Variational Auto-Encoders | 16 |
| 3.4.1 | Prior over the latent variables | 16 |
| 3.4.2 | The decoder distribution | 17 |
| 3.4.3 | The encoder distribution | 17 |
| 3.4.4 | Forward pass of the VAE model | 18 |
| 3.4.5 | ELBO function for the VAE model | 18 |
| 3.4.6 | Putting the pieces together | 19 |
| 3.4.7 | Sequential Variational Auto-Encoders in Literature | 21 |
| 3.5 | Normalizing Flows | 21 |
| 3.5.1 | Masked Auto-Encoders for Density Estimation (MADEs) | 23 |
| 3.5.2 | Inverse Auto-Regressive Flows (IAFs) | 24 |
| 3.5.3 | IAFs in Traditional Variational Auto-Encoders | 25 |
| 4 | Proposed Temporal Variational Auto-Encoders | 28 |
| 4.1 | Temporal independent tVAE (tVAE1) | 30 |
| 4.1.1 | Implementation details for the tVAE1 model | 32 |
| 4.1.2 | Forward pass for the tVAE1 | 33 |
| 4.2 | Past and present dependent tVAE (tVAE2) | 33 |

| | | |
|-----------|---|-----------|
| 4.2.1 | Implementation details for tVAE2 model | 35 |
| 4.2.2 | Forward pass for the tVAE2 | 36 |
| 4.3 | Full temporal dependent tVAE (tVAE3) | 36 |
| 4.3.1 | Implementation details for the full temporal dependent tVAE | 38 |
| 4.3.2 | Forward pass for the tVAE3 | 38 |
| 4.4 | Normalizing Flows in tVAEs | 39 |
| 5 | Proposed Approach for Semi-Supervised Learning with tVAEs | 41 |
| 6 | Experimental Setting | 44 |
| 6.1 | Baseline Models | 44 |
| 6.1.1 | Baseline 1: Ablation Model | 44 |
| 6.1.2 | Baseline 2: Decoupled Model | 45 |
| 6.2 | Description of the Experiment | 47 |
| 6.3 | Model’s Neural network Architectures | 48 |
| 7 | Case Study 1: CMAPSS dataset for Turbofan Engines Remaining Useful Life Estimation | 50 |
| 7.1 | Data Preparation | 50 |
| 7.2 | Results and Discussion for the FD004 Dataset | 53 |
| 7.2.1 | Assessing the different time-structures of the encoder and decoder | 53 |
| 7.2.2 | Assessing the improvements achieved by using Inverse Auto-Regressive Flows | 56 |
| 7.2.3 | Assessing the improvements achieved by the coupled training process | 57 |
| 7.2.4 | Results from the Ablation Study | 60 |
| 8 | Case Study 2: Pumping system dataset for Diagnosis of Health State | 62 |
| 8.1 | Data Preparation | 62 |
| 8.2 | Results and Discussion for the Pumping System Dataset | 65 |
| 8.2.1 | Assessing the different time-structures of the encoder and decoder | 65 |
| 8.2.2 | Assessing the improvements achieved by using Inverse Auto-Regressive Flows | 67 |
| 8.2.3 | Assessing the improvements achieved by the coupled training process | 69 |
| 8.2.4 | Results from the Ablation Study | 71 |
| 9 | Concluding Remarks | 73 |
| 10 | Appendix | 76 |
| | Bibliography | 79 |