2014

Graphical Models in Characterizing the Dependency Relationship in Wireless Networks and Social Networks

Phuoc Doan Huu Vu
Louisiana State University and Agricultural and Mechanical College

GRAPHICAL MODELS IN CHARACTERIZING THE DEPENDENCY RELATIONSHIP IN WIRELESS NETWORKS AND SOCIAL NETWORKS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering

in

The Division of Electrical and Computer Engineering

by
Phuoc Doan Huu Vu
B.S., Louisiana Tech University, 2012

August 2014

Acknowledgments

This thesis would not be possible without several contributions. First and foremost, I would like to express my sincerest gratitude to my advisor, Dr. Shuangqing Wei for his invaluable and insightful guidance throughout the course of my study at Louisiana State University. I am so grateful for his kind support and guidance that helped me clarify any misgivings about the most minor of details to truly understand insights in my research career.

I want to express my special gratitude to my thesis committee, Dr. Ramachandran Vaidyanathan, Dr. Jian Zhang for patiently attending my final exams and providing me with their constructive comments and suggestions.

Finally, I am obliged to many of my colleagues who supported me during my Ph.D. years. I thank Shuhang Wu, Ahsan- Abbas Ali, Ali Moharrer, Ben Carroll, Yahya Khiabani and Limeng Pu for their warm friendship and support and helped me clarify my thoughts throughout the on-going of this research.

This work was motivated by the study of semi Markov Process and Hidden Markov model proposed by Barbu and Liminios (2011).

This thesis is dedicated to my parents and my fiance for their support and encouragement throughout of my graduate study at LSU.

Table of Contents

List of Tables

List of Figures

Abstract

Semi-Markov processes have become increasingly important in probability and statistical modeling, which have found applications in traffic analysis, reliability and maintenance, survival analysis, performance evaluation, biology, DNA analysis, risk processes, insurance and finance, earthquake modeling, etc. In the first part of this thesis, we first present novel approaches to establishing statistics of the resulting random processes under the operations of down-sampling, superposition, and mislabeling on discrete time semi-Markov processes, respectively. We show that the resulting processes under the operations are still semi-Markov processes. Moreover, we prove that the statistics of the original semi-Markov sequence in terms of its sojourn time distribution, as well as its probability transition matrix can bother be restored given their counter parts under either superposition or mislabeling operations. As a contrast, we show that down-sampling creates singularity issues, thereby making it impossible to restore the original statistics. Simulation and numerical results further demonstrate the validity of our theoretical findings. Our results thus provide a more profound understanding on the limitation of applying semi-Markov models to characterizing and learning the dynamics of nodes activities in wireless networks.

In the second portion of the thesis a review is provided about several graphical models that have been widely used in literature recently to characterize the relationships between different users in social networks, the influence of the neighboring nodes in the networks or the semantic similarity in different contexts.

1

Introduction

The hidden Markov model and semi Markov model are two important models that were applied to a variety of domains, such as signal processing, machine learning, communications, and many more. The two models were first introduced in the late 1950s and later generalized by Baum and Petrie in 1966 [1]. There is much work in the literature focusing on the applications of Hidden Markov model to speech recognition [2], signal processing [3]. These applications in statistical signal processing and communications again reflect the power and flexibility of the model. Without the hidden part, on the other hand, semi Markov model is a generalization of Markov and of renewal processes and was shown to be a very powerful tool. Very recently, "Barbu and Limnios" [4] have introduced the semi Markov model in the discrete-type setting and pointed to various applications in reliability engineering and statistical learning. Nowadays, semi-Markov processes have become increasingly important in probability and statistical modeling. In their recent work, the authors in [5] proposed the problem of coverage intensity defined as the probability distribution of durations within which a target or an event is uncovered/unmonitored. They derived this distribution based on semi-Markov model, and the superposition of alternating renewal processes. In [6], the authors have studied the superposition of multiple independent semi Markov process and applications to bursty traffic sources and then derived the analysis for a statistical multiplexer model.

Recent years have witnessed growing attentions in applying semi Markov process to model on-off duty cycle of different nodes in wireless networks. In [7], the authors proposed

that semi Markov chain can be used to design lifetime model for each sensor node by considering the power consumption in different operational modes and the energy overheads incurred during transitions. Semi Markov chain was also a viable candidate for different types of Measurement-Based Model for Dynamic Spectrum Access in WLAN Channels. In [8], Kadiyala proposed a semi-Markov process based model to compute the network parameters such as saturation throughput, for the IEEE 802.11 Distributed Coordination Function employing the Binary Exponential Backoff.

Cherry was the first to introduce the superposition of two continuous time Markov renewal process [9] . In his paper, the structure of the interval process resulting from superposing two independent Markov renewal processes is characterized. The resulting stochastic process has a very large number of states that limit the applicability of the model to two processes. In [10] a mathematical model for the superposition of multiple independent continuous-time Markov renewal processes was introduced. The model records the times each process spent in the current state thereby limiting its applicability to cases where analytic expressions for the sojourn times between states can be found. Following similar methodology, Elsayed and etc. [6] presented an approximate model for characterizing the superposition process of $N > 2$ independent discrete-time binary Markov renewal process with the applications to bursty traffic sources and then derived the analysis for a statistical multiplexer model. The main disadvantage of their approach is the resulting state space is huge with the number of states in an order of $O(4^N)$ states where N is the number of independent semi Markov processes. Their work was followed by Hsin and Liu who have derived coverage intensity for wireless sensor network [11]. In both these papers the authors relied on the tuple (x_i, t_i) where x_i is the state of the i^{th} process and t_i the indicator state changes.

As can be seen from the previous examples, hidden Markov models and semi Markov models have been employed to model the dynamics of wireless network nodes. However, to the best of our knowledge, there is a lack of deep understanding in regard to how operations in a practical set-up such as sampling, superposition, and even mislabelling

due to near-far effect affect the restoration of the statistics of the original semi-Markov processes. Such issues were exemplified in one of our recent works [12]. In [12], we have adopted Bayesian Hidden Semi-Markov Model (HSMM) for detecting wireless RF devices. Specifically, we have employed multiple USRPs to simulate both coordinated and non-coordinated transmissions of wireless nodes in a small scale network. The generated RF traces were then collected via downsampling by a monitoring USRP node where an off-line non-parametric learning algorithm was executed to partition and label the collected RF traces. In our experimental study, we have noticed that the learning algorithm has done a decent job in segmenting RF traces into meaningful states. However, the identified post-sampling states transitions demonstrate some unseen patterns not evident in the original processes, which has thus prompted us to seek answers to such issues. Furthermore, and more importantly, the experimental works have prompted us to question if the statistics of semi-Markov processes are recoverable or not given those of the resulting discrete time sequences under the aforementioned operations.

It should be noted that our objective is quite different than that addressed by traditional sampling theorems, which are about estimating the original band-limited random process given its downsampled sequences. Rather, we are interested in only the original statistics which are captured by both states transition probabilities and sojourn time distributions for semi-Markov processes. Also notable is such recovery question becomes trivial if the original process is not Semi-Markov, but rather Markov. The findings in this paper could help us understand more profoundly the fundamental limitation in learning the nodes activity patterns in wireless networks under the widely used semi-Markov models when downsampling is necessary due to concerns of computational cost, as what we experienced in our experiments. Intuitively, superposition of semi- Markov processes are due to the collision of transmitting packages and mislabelling after the superposition is similar to that of near-far effect. The near-far problem is a condition in which a receiver captures a strong signal thereby making it impossible for the receiver to detect a weaker signal, a situation common in wireless communication systems [13]. Interestingly, differ-

ent from down-sampling operation, we can generally recover the statistics of the original sequence after superposition and mislabelling operations. We have presented some of the results dealing with down-sampling issues in [14], and we provide more detailed proofs here, as well as those findings related to other operations including superposition and mislabeling on semi-Markov processes. We report the results of other operations on semi-Markov processes such as super-position and false-labeling in addition to providing more detailed proofs for the results in [14]. We present both theoretical and numerical and simulation results on the effects of these operations on restoration of the original sequences' statistics in Chapter 2 of the Thesis.

In addition to characterizing the dependency of wireless networks, hidden Markov model also plays an important role in modelling the relationship of social network users. Many works have been done recently to follow this trend. Chapter 3 of this thesis reviews current literature on the coupled hidden markov model with application in social networks and hidden markov model tree dependency on understand the semantic relatedness and similarity.

2

Statistical properties of semi-Markov process

This chapter is organized as follows. Section II presents the set up and specifications of our experiments and the result from two coordinated users using OFDM transmissions. Section III first presents background notations on Semi Markov processes, and then provides analytical solutions to the statistics of downsampled sequences, as well as the justification on the singularity issue in restoration of the original semi-Markov processes. We also present the solution method for the superposition of two independent discrete time semi Markov processes and derives the formulation of the mislabelling problem. In Section IV, we compare the derived numerical results with those using simulations to further demonstrate the validity of our findings.

2.1 Experimental results from identification of the transmissions of wireless RF devices

2.1.1 Modelling and specification

In our experimental study, we have implemented the non-parametric learning algorithms proposed in [15] to learn the hidden sates of wireless RF devices under the framework of Hidden Semi-Markov Processes (HSMM). This was accomplished by programming several USRPs to transmit data according to semi-Markovian behavior as implemented through custom Python programs using GNU Radio. The Python programs enable two USRPs to coordinate their activity through the host PC so that no packet collisions are produced as they transmit data over the program's execution. A third USRP was inter-

fering the transmissions from the first two coordinated USRPs and the fouth USRP was then utilized to collect wireless RF traces of the generated activity to use as inputs to the Bayesian HSMM algorithm. It is the goal of the Bayesian HSMM algorithm to identify the number of devices present in each collected RF trace by examining the statistical properties of the received signal over time and to also identify collision instances when two USRPs attempt to transmit data packets simultaneously. Collection of the wireless traces over the ISM band was performed by running the USRP Python program. The inputs to this program specify the center frequency of interest, the sampling rate with which the received signal appearing at the antenna are digitized by the USRP's ADC. Due to the large amount of data stored within the file at the utilized sampling rate of 500 kHz for these experiments, the data was subsequently down-sampled to a rate of 1 kHz to allow the Bayesian HSMM algorithm to be conducted in reasonable amount of time. More descriptions of the experimental set up and configurations can be found in [12]. Here we present the experimental results for a case of two coordinated OFDM transmissions in the presence of interference from the third node.

2.1.2 Two coordinated transmissions in the presence of interference

We considered an experiment to assess the Bayesian HSMM algorithm's ability to discern between two coordinated USRPs with the third interfering USRP transmission. Both coordinated USRPs were chosen to send OFDM symbols with BPSK as the underlying modulation. A Python program was used to generate a realization of a Markov chain state sequence through the specification of an idle/busy state transition matrix P and a second uniformly distributed random variable $x \sim U(0, 1)$ used to select which USRP will be selected for transmission during any busy state. If $x < 0.5$, USRP 1 is chosen for transmitting packets, otherwise USRP 2 will send its own packets. The busy state durations corresponding to USRP 1/2's transmissions is governed by the amount of packets, packet size, and OFDM symbol bandwidth chosen for each USRP. Whenever USRP 1 is chosen to transmit its packets according to the generated Markov chain, it will send 5 packets with probability 0.5. Likewise, it will send 10 packets with equal probability. The

duration of the idle state D_{off} is generated through draws from an exponential random variable with a specified mean. The idle state durations are further bounded by minimum and maximum values to prevent extremely short durations or extremely long durations. As an example, suppose $D_{off} \sim Exp(0, 2)$ with bounds of $[0.3, 0.5]$. If the drawn value of D_{off} falls within $[0.3, 0.5]$, then the value is kept, otherwise, the idle duration will be the closest interval bound.

Table 2.1 presents a summary for the experimental setup, in which the interference introduced by USRP 3 is the result of GMSK transmissions over the wireless channel. Since GMSK is a means of frequency shift keying and is thus distinct from the OFDM transmission scheme and the underlying phase shift keying for USRPs 1 and 2, it is expected that the HSMM algorithm should be able to distinguish GMSK packet transmissions with better accuracy.

TABLE 2.1. Experimental set up for Experiment 2

	USRP 1	USRP 2	USRP 3
Modulation	OFDM-BPSK	OFDM-BPSK	GMSK
Symbol Bandwidth	250 kHz	250 kHz	-
Bit Rate			250 kHz
Number of Subcarriers	200	200	-
Packet Size	508 bytes	508 bytes	508 bytes
Packet Bunches	[5,10]	[5,10]	[5,10]
Packet Emission Probs.	[1/2, 1/2]	[1/2, 1/2]	[1/2, 1/2]
Transmission Gain	15.0	25.0	3.0
Idle State Duration	D_{off} ~Exp(0.2) sec	$D_{off} \in$ [0.30.5] sec	

Figure 2.1 depicts the labeled state sequence after the final iteration of the algorithm, along with the sample magnitudes for each point in the RF trace.

A legend for mapping each state to its corresponding can be found in the following. There are 5 states after the final iteration: State 1 with black color is USRP 3 state whose duration distribution follows Poiss($\lambda_1 = 65.25$); State 2(dark blue) is USRP 2 state whose duration distributions follow Poiss($\lambda_2 = 130.42$); State 3 with cyan color is USRP 1 and idle state with Poiss($\lambda_3 = 149.93$); State 4 with green color is Collisions and Transients with Poiss($\lambda_4 = 41.92$); State 5 with yellow color is USRP 3 and Collisions

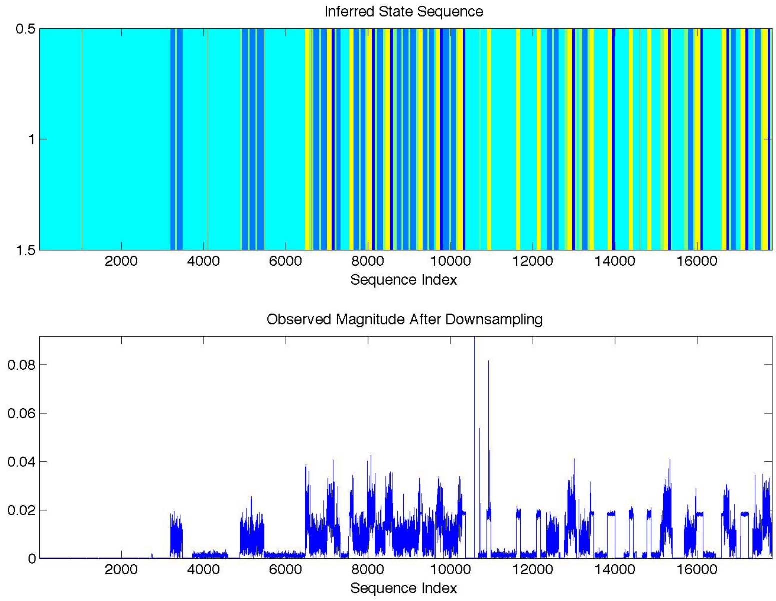

FIGURE 2.1. Labeled State Sequence for Experiment 3

with Poiss($\lambda_5 = 107.18$). At the completion of the final iteration, the resultant state sequence used only 5 model states. The learned state transition probability matrix is given by:

$$\mathbf{A} = \begin{bmatrix} 0.000 & 0.008 & 0.147 & 0.778 & 0.066 \\ 0.000 & 0.000 & 0.262 & 0.698 & 0.040 \\ 0.001 & 0.000 & 0.000 & 0.201 & 0.798 \\ 0.004 & 0.577 & 0.142 & 0.000 & 0.276 \\ 0.507 & 0.064 & 0.268 & 0.161 & 0.000 \end{bmatrix}$$

Once again, the final results in the state labeling show missed detection of USRP 1, as it is considered to belong to the same state as vacancies over the channel. Instances of GMSK's presence over the channel are consistently detected and considered as belonging to the same state, regardless of collision instances or undisturbed transmission. States 1 and 4 of the model also represent shorter duration states that seem to result from transient behavior as transmissions initiate from the USRP devices. As we can see that the learning algorithm mislabelled two states from USRP 1 and USRP 2, combining them into a single state. State 3 with cyan color is the mislabelling. From Figure 2.1, we

also observe many fast switching from USRP 1 to USRP 2 and the busy states are not always followed by the idle state any more but there is transition from a busy state to another busy state, i.e. the coordinated property is not fully restored after down-sampling. We can thus conclude based on the empirical results, not surprisingly, down-sampling, superposition due to collision, and mislabeling due to overshadowing effect from strong transmission, which could be some prevailing issues emerging in a practical set-up, all play a role in transforming the statistics of the original nodes activity patterns. What remains challenging then is to quantitatively demonstrate to what extent such operations affect the statistics, and whether such transformations are invertible in the sense of recovering the original sequences' statistics. We will address all these questions in the next a few sections.

2.2 An analytical approach for the problem: downsampling, superposition and mislabelling

We present an analytical approach to elaborating issues seen above from the experiments. Some notation and denitions are in order. All vectors and matrices are represented with lower and upper case boldface fonts respectively. Sets are represented with calligraphic fonts. Random variables are represented with italic fonts while their realisation is represented by lower case italic fonts. Let I_n be the identity matrix of size $n \times n$ while 1_n denotes the column vector of n ones. Throughout the paper, we reseve the lower case letter for probability distribution, for example, $h_i(k)$ denotes the sojourn time distribution of state i of the semi- Markov process. We denote letter with $\sim$ to define the cumulation distribution, for example, $\tilde{h}_i$ denotes the cumulative sojourn time distribution of state i. We keep upper case letter for the z-transform corresponding to the distribution in time domain, for example, $H_i(z)$ denotes the z-transform of the sojourn time distribution in state i. We say letter with hat implies the remaining distribution, for example, $\hat{h}_i(k)$ is the remaining life-time of the sojourn time distribution in state i. Finally the superscripts show the nature of the processes. For example, X_k^1 or X_k^2,.. show the component processes for N independent semi- Markov processes with the resulting X_k^S as the superposition

process. Similarly, $X_k^{S,M}$ is defined as the mislabelled semi- Markov process after superposition. And in the paper, we use X to shows the original sequence while Y defined as the down-sampled semi- Markov process.

2.2.1 Review of relevant background on semi- Markov processes

Define E as the state space of the semi-Markov process: $E = \{1, 2, 3, .., s\}$. Let $\mathbf{N}$ be the set of integers, i.e. $\mathbf{N} = \{0, 1, 2, 3, ..\}$. Define the set of non-negative matrices on $E \times E$ as μ_E. Consider a semi-Markov chain with state space E and let $\{\ i, j, k\ \}$ be the three states from the set E given by the following diagram in Figure 2.2. Here denote

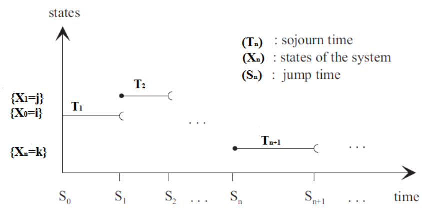

FIGURE 2.2. A sample path of a discrete time semi Markov chain

$\{X_n\}$ as the states of the chain at the n^{th} arrival and $\{T_{n+1}\}$ as the sojourn time of the semi-Markov process at that n^{th} arrival and $\{S_n\}$ be the corresponding jump time. Similar to the approach of [4], we define a discrete-time semi-Markov kernel: a matrix valued function $\mathbf{q} \subset \mu_{\mathbf{E}}$ is said to be discrete-time semi-Markov kernel if the following three conditions are met:

i. $0 \leq q_{ij}(k) \leq 1$

ii. $q_{ij}(0) = 0$; $\sum_{k=0}^{\infty} q_{ij}(k) \leq 1$ with $i, j \in E$

iii. $\sum_{k=0}^{\infty} \sum_{j \in E} q_{ij}(k) = 1$, for $i \in E$

Here the right continuous jump at S_n occurs at the state X_n, whose duration is T_{n+1}. We can define the element of the kernel $\mathbf{q}$ as

$$q_{ij}(k) = P(X_{n+1} = j, T_{n+1} = k \mid X_n = i). \tag{2.1}$$

Intuitively, $q_{ij}(k)$ is the probability that the semi-Markov chain jumps from state i to state j with the time spent during state i as k units of time. We want to make a remark about the subtle difference between the index of n and k, respectively. Here the index n

refers to the states of the Semi- Markov process and also refers to the arrival nature and the index k refers to time epochs of the Semi Markov chain/sequence or refer to the time nature. The transition matrix of the embedded Markov chain (X_n) defined by:

$$\tilde{p}_{ij} = P(X_{n+1} = j \mid X_n = i) \tag{2.2}$$

with $i, j \in E$ and $n \in \mathbf{N}$. Sojourn times distribution in a given state depends on the current state as well as the next state. For all $i, j \in E$ we denote $h_{ij}(k)$ be the sojourn time distribution in state i and the next state is j. We can write

$$h_{ij}(k) = P(T_{n+1} = k \mid X_n = i, X_{n+1} = j). \tag{2.3}$$

We have the following relationship:

$$q_{ij}(k) = \tilde{p}_{ij} h_{ij}(k) \tag{2.4}$$

The sojourn time distribution in a given state i can be written as:

$$h_i(k) = \sum_{\text{J} \in E} h_{ij}(k). \tag{2.5}$$

Define $Z = (Z_k)$, $k \in \mathbf{N}$, to be a semi-Markov chain with $Z_k = X_{N_k}$, $k \in \mathbf{N}$ with $N_k = max(n \in \mathbf{N} \mid S_n \leq k)$. Then N_k is the discrete counting process of the number of jumps in $[1, k]$ which is $\in \mathbf{N}$ and Z_k gives the system state at time k. Define the cumulative sojourn time distribution as:

$$\tilde{h}_i(k) = \sum_{\text{J} \in E} \sum_{l=1}^{k} h_{ij}(l) = \sum_{\text{J} \in E} \sum_{l=1}^{k} \frac{q_{ij}(l)}{\sum_{m=0}^{\infty} q_{ij}(m)}. \tag{2.6}$$

The transition function of semi-Markov chain Z is the matrix-valued function $\mathbf{P} \in \mu_E$ defined by

$$p_{ij}(k) = P(Z_k = j \mid Z_0 = i) \tag{2.7}$$

with $i, j \in E$. The transition function $\mathbf{P}$ can be computed as

$$p_{ij}(k) = I_{ij}(k)(1 - \tilde{h}_i(k)) + \sum_{r \in E} \sum_{l=0}^{k} q_{ir}(l) p_{rj}(k - l) \tag{2.8}$$

where $I_{ij}(k)$ is the indicator function, $I_{ij}(k) = 1$ if $i = j$ and $I_{ij}(k) = 0$ otherwise. We have in matrix form:

$$\mathbf{p} = \mathbf{I} - \tilde{\mathbf{h}} + \mathbf{q} * \mathbf{p}. \tag{2.9}$$

The cumulated semi-Markov kernel $\tilde{\mathbf{q}} = \tilde{q}_{ij}$ defined by:

$$\tilde{q}_{ij}(k) = P(X_{n+1} = j, T_{n+1} \leq k \mid X_n = i) = \sum_{l=0}^{k} q_{ij}(l). \tag{2.10}$$

We have the result for the elements of the transition matrix of the embedded markov chain as:

$$\tilde{p}_{ij} = \tilde{q}_{ij}(\infty) = \sum_{k=0}^{\infty} q_{ij}(k). \tag{2.11}$$

The stationary distribution of the semi-Markov process can be calculated as follows. Let $v = [v(1)v(2)...v(n)]$ be the stationary distribution of the embedded markov chain. In other word, $v = vP$ where P is transition matrix. We define the mean sojourn time in any state i as $m_i = E(S_1 \mid X_0 = i) = \sum_{k \geq 1} kh_i(k)$.
From that the j-element of the stationary distribution of the semi-Markov chain is given by: $\pi_j = \frac{v(j)m_j}{\sum_{\imath \in E} v(i)m_i}$. We have $\pi = (\pi_j)$, $j \in E$, is the stationary distribution of the semi-Markov chain.
The following figure (Figure 2.3) outlines the road map used for different operations of Semi- Markov Processes:

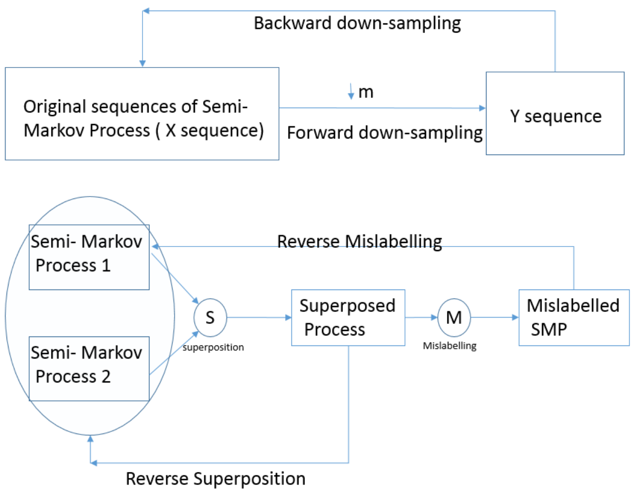

FIGURE 2.3. Problem Formulation and Outline

2.2.2 Statistical properties of down-sampled semi-Markov sequences

Main results for the down-sampling problem:

We first propose a result, which plays an important role in understanding the characterized behaviors of the down-sampled sequence:

Proposition 1. *The resulting process after downsampling a semi-Markov process is also a semi-Markov process.*

Here we present the outline of the proof. For a finite alphabet set of the state space, we define again our down-sampling as following. For the down-sample factor of $m > 1$ we keep the first letter and delete the next $m-1$ letters. One observation is that after the down-sampling the state space of the resulting sequence is the same as the state space of the original sequence. Our proof depends on the counting process of each state in semi-Markov chain. For more details, refer to [16].

Next, since the down-sampled sequence is also a semi-Markov process, we are interested in finding the statistical properties of the resulting process. More specifically, we want to first find how downsampling is reflected in the statistics of the resulting sequence in terms of its sojourn time distribution and transition probabilities matrix. We next establish the foundations of our work in the following result. Only a simple, but non-trivial case with 3-state semi-Markov processes is given, whose results can be extended to more general cases in the similar manner. This case is also reflecting a set-up in our experiments when we only consider activity patterns of two nodes, together with the interlaced idle states.

Proposition 2. *For a given 3 states semi-Markov process, the relationship between the transition function $p_{ij}(k)$ and the semi-Markov kernel $Q_{ij}(k)$ for $i, j \in \{1, 2, 3\}$ in the transform domain are given by:*

$$Q_{12}(z) = \begin{vmatrix} P_{12} & P_{13} \\ P_{32} & P_{33} \end{vmatrix} \bigg/ \begin{vmatrix} P_{22} & P_{23} \\ P_{32} & P_{33} \end{vmatrix} \tag{2.12}$$

$$Q_{13}(z) = \begin{vmatrix} P_{12} & P_{13} \\ P_{22} & P_{23} \end{vmatrix} \bigg/ \begin{vmatrix} P_{22} & P_{23} \\ P_{32} & P_{33} \end{vmatrix} \tag{2.13}$$

where $|A|$ denotes determinant of a non-singular matrix A.

Similar expressions can be written for $Q_{21}(z), Q_{23}(z), Q_{31}(z), Q_{32}(z)$. The complete proof can be found in Appendix B. And from that we can find the relationships between a given semi-Markov kernel from state i to state j, $i, j \in \{0, 1, 2\}$ $Q_{ij}(z)$ and $P_{mn}(z)$, $m, n \in \{0, 1, 2\}$. The sojourn time distribution: $H_{ij}(z) = \frac{Q_{ij}(z)}{\tilde{p}_{ij}}$. We can get $\tilde{p}_{ij}$ by letting $\tilde{p}_{ij} = Q_{ij}(1)$. Finally, taking the numerical inverse z-transform to get $h_{ij}(k)$.

Suppose that the down-sampling factor is m. In other words, from the original semi-Markov chain $X = \{X_k\}$, $k \in N$, after down-sampling the sequence X by a factor m, we got a resulting semi Markov chain $Y = \{Y_k\}$, $k \in N$. We present next result that can help us understand the connection between sequence X and sequence Y in terms of the relationship between their respective z-transform of the transition function.

Proposition 3. *The relationship between the transition function $p^X_{ij}(k)$ and the transition function $p^Y_{ij}(k)$ in transform domain is given by:*

$$P^Y_{ij}(z) = \frac{1}{m}\left(P^X_{ij}(z^{\frac{1}{m}}) + \sum_{n=1}^{m-1} P^X_{ij}(z^{\frac{1}{m}} \exp(\frac{-i2\pi n}{m})\right) \tag{2.14}$$

Proof. Since the results are quite straight-forward and has been discussed in [17]. From the down-sampling we have:

$$p^Y_{ij}(k) = p^X_{ij}(mk) \tag{2.15}$$

with $i, j \in E$. Hence applying the frequency domain on both sides of the above equation, or in z-transform domain and the properties of the down-sampling with factor m, we obtain the immediate result for Equation (14). □

Now, we want to apply the propositions given to find the statistics of the down-sampled sequence Y given the statistics of the original sequence X. The method to use is to write all the quantities in terms of the semi-Markov kernel: $q^X_{ij}(k) = \tilde{p}^X_{ij} h^X_{ij}(k)$, with $i, j \in E$ where $\tilde{p}^X_{ij}$ denotes the ij- element of the embedded Markov chain transition matrix of $\{X_k\}$. Then from Proposition 3 we can find the transition function $P^Y(z)$. Solving for $P^Y_{ij}(z)$ and then from that we can find $Q^Y_{ij}(z)$ by the following relationships: $Q^Y_{01}(z) = \frac{P^Y_{01}(z)P^Y_{22}(z) - P^Y_{02}(z)P^Y_{21}(z)}{P^Y_{11}(z)P^Y_{22}(z) - P^Y_{12}(z)P^Y_{21}(z)}$

The rest of the expressions $Q^Y_{ij}(z)$ can be found explicitly from Proposition 2. So for each $Q^Y_{ij}(z)$ we can have two derivations. The transition probability matrix: $\tilde{p}^Y_{ij} = Q^Y_{ij}(1)$ in the z-domain from the above equations. And the sojourn time distribution: $H^Y_{ij}(z) = \frac{Q^Y_{ij}(z)}{p^Y_{ij}}$. Finally, taking the numerical inverse z-transform to get $h^Y_{ij}(k)$. Reverse downsampling problem:

In this subsection, we provide an analytical solution to answer the beginning question that given the observed random process after down-sampling, how much of the statistics of the original sequence we can restore.

Proposition 4. *There are infinitely many solutions for the reverse down-sampling problem, i.e. there is singularity issue and we cannot recover the statistics of the original sequence after downsampling.*

Proof. Suppose the statistical properties of the Y sequence are known, namely the transition probability matrix and the sojourn time distribution of the Y sequence. We want to find the statistics of the X sequence. We would like to find transition probability matrix and $h^X_{ij}(k)$ in terms of the down-sampling factor m and statistics of the Y sequence. Writing all the quantities in terms of the semi-Markov kernel from Equations (2) through (7) and using the z transform and the properties of the down-sampling with factor m by Proposition 3, we need to solve the functional equation (14). From Proposition 3 we solve for $P^X_{ij}(z)$ and then from that we can find $Q^X_{ij}(z)$ using the same method as in the forward problem. The first question to address involves the functional equation (14). Let $G_{ij}(z)$ be a function on complex domain that satisfies: $\frac{1}{m}(G_{ij}(z^{\frac{1}{m}}) + \sum_{n=1}^{m-1} G_{ij}(z^{\frac{1}{m}} \exp(\frac{-i2\pi n}{m})) = 0$. Also, from the complex domain, we always have

$$z^{\frac{1}{m}} + \sum_{n=1}^{m-1} z^{\frac{1}{m}} \exp(\frac{-i2\pi n}{m}) = 0. \tag{2.16}$$

We want to find all the Q functions that satisfy the above two properties. One possible function $G_{ij}(z)$ can be given by:

$$G_{ij}(z) = a_p z^p + a_{p-1} z^{p-1} + .. + a_1 z \tag{2.17}$$

where $p < m$. And this is one of the many solutions that we can find for Equation (14). Hence, we have one of the solutions for functional equation (14) as $P_{ij}^X(z) = P_{ij}^Y(z^m) + G_{ij}(z)$. Therefore, we can construct infinitely many solutions to the restoration problem on the statistics of the original semi-Markov processes, based on the downsampled sequence statistics only, thereby demonstrating the singularity issue in restoring the pre-downsampling statistics. $\square$

Relationship between time domain and frequency domain statistics:
The following figure (Figure 2.4) shows the relationship between time domain and frequency domain statistics, the framework under which we carry throughout the proofs of main results:

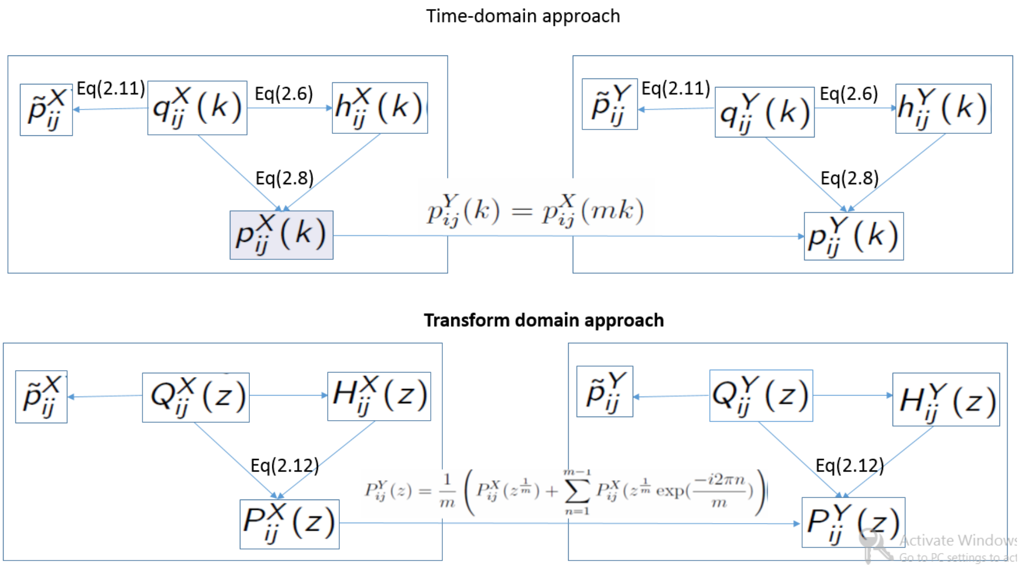

FIGURE 2.4. Relationship between time domain and frequency domain statistics

The outline of the proofs can be illustrated by the following Diagram (Figure 2.5):

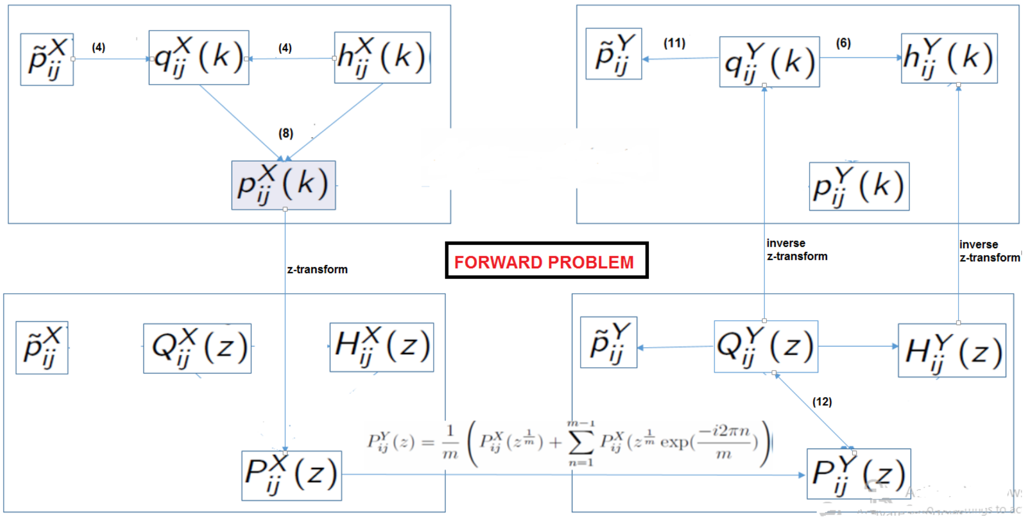

FIGURE 2.5. Relationship between time domain and frequency domain statistics

2.2.3 Superposition of two independent discrete-time semi-Markov processes

Main results for the superposition problem:

Here we only present the results on the superposition of two-processes for the ease of presentation, and the purpose of being consistent with our experimental set-up in which the transmission of a third USRP could collide with that from one of the coordinated USRPs. It should be noted that our proposed approach to establishing the statistics of superposed two independent semi-Markov processes could extended in a straightforward manner to the case of $N > 2$ independent processes. The size of state space under superposition in our approach is reduced to $O(2^N)$ as compared with $O(4^N)$ in [11] when each of the N processes has two states.

More specifically, our main contribution for the superposition problem is that we can show how to construct the transition from any state i to j in state space E of the superposition process by explicitly taking into account both the age and residual life time of each component process as a result of the superposition. Consequently, as a comparison with the existing works on superposition of semi-Markov processes, we do not take into account t_i, the indicator state changes. Other methods have relied on the tuple (x_i, t_i) where x_i is the state of the i^{th} process and t_i the indicator state changes. This results in a state space whose size scales in the order of $O(4^N)$ states where N is

the number of independent binary semi Markov processes. Hence when extending our results to a large number of independent binary semi Markov processes, the size of the resulting state space is in the order of $O(2^N)$, rendering less computationally expensive. Also notable is that we have taken a time-domain approach to tacking the superposition problem, as demonstrated next, rather the transform domain one used in coping with down-sampling issues previously because of the need to consider both elapsed time and remaining time during superposition operations.

We define $q_{ij}(k)^m$ as the probability of making a transition from the current state i to the next state j, with the remaining time k slots in the current state i, where i and j are in E_m, the state space of a particular semi-Markov process m. Let $\tilde{p}_{ij}^m$ denote the transition probability from state i to state j of a particular semi Markov process m, $1 < m \leq N$ and we have

$$\tilde{p}_{ij}^m = \sum_{k \geq 1} q_{ij}^m(k) \tag{2.18}$$

also $\sum_{j \in E_m} \tilde{p}_{ij}^m = 1$, and we define $h_{ij}^m(k)$ as the sojourn time distribution in state i, next state j of the process m

$$h_{ij}^m(k) = Pr(X_{n+1}^m = j, T_n^m = k | X_n^m = i) \tag{2.19}$$

where T_n^m is the inter-arrival time of state i and X_n^m is the state at the n arrival.

We have $\sum_{k \geq 1} h_{ij}^m(k) = 1$ and $q_{ij}^m(k) = \tilde{p}_{ij}^m h_{ij}^m(k)$ for all $k > 0$ and $i, j \in E_m$ Let $\tilde{h}_{ij}^m(k) = \sum_{l=1}^{l=k} h_{ij}^m(k)$ be the associated cumulative probability density function of length up to k for state i of process m. Define the mean duration or mean inter-arrival time for state i is

$$M_i = \sum_{k \geq 1} k h_{ij}^m(k) \tag{2.20}$$

Residual life-time probability mass function

Suppose that $N(t)$ is the number of arrivals of the semi Markov process up to time t. Suppose that S_n, $n = 1, 2, ..$ be the arrival times. The following denotations will be used

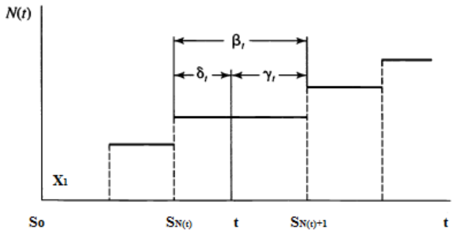

FIGURE 2.6. Residual life-time, current life (age) and total life-time

as the residual life-time and age of the semi Markov process. We define

$$\gamma_t = S_{N(t)+1} - t \tag{2.21}$$

be the excess or residual life at time t. Define

$$\beta_t = t - S_{N(t)} \tag{2.22}$$

as the current life or age at time t. The probability mass distribution of β_t as $t \to \infty$ is the sojourn time distribution of the state i at time t for the semi Markov process.From renewal theory, the probability mass distribution of the residual life time γ_t^i of state i for a particular process m can be shown to be

$$\hat{h}_{ij}^m(k) = Pr(\gamma_t^i = k) = \frac{1 - \tilde{h}_{ij}^m(k)}{M_i} \tag{2.23}$$

We have $\sum_{k \geq 1} \hat{h}_{ij}^m(k) = 1$ and $\hat{h}_{ij}^m(k)$ is the residual life-time probability mass function associated with sojourn time distribution $h_{ij}^m(k)$. We define the residual life-time semi Markov kernel $\hat{\mathbf{q}}_{ij}^m$ associated with the semi Markov kernel $\hat{q}_{ij}^m(k)$ which has the property:

$$\hat{q}_{ij}^m(k) = Pr(X_{n+1}^m = j, \gamma_t^i = k | X_{n+1}^m = i). \tag{2.24}$$

$\hat{q}_{ij}^m(k) = \tilde{p}_{ij}^m \frac{1 - \tilde{h}_{ij}^m(k)}{M_i}$ Similar to the sojourn time distribution, we define the cumulative residual life-time probability distribution associated with $\hat{q}_{ij}^m(k)$ as $\tilde{\hat{q}}_{ij}^m(k)$.

Let $T(X_j)$ be the time that process j has spent so far in state X_j. Due to the independence of the two processes, the distribution of $T(X_j)$ will be equal to the sojourn distribution or the current life distribution in state X_j. The time $T'(X_j)$ that it takes process j to undergo a state change would have a distribution equal to the residual life-time

distribution at state X_j. This is true when we take all possible realization of the above event and assuming that all processes are independent. Define $t_i \in \{0, 1\}$ denote whether process i has changed state or not, with $t_i = 1$ if process i has changed state. For a particular state $u = (X_1, X_2)$ we need to have $\sum_{i=1}^{2} t_i > 0$ because an arrival happens in the superposition when one or more of the component processes makes a transition. Next, we propose a result which plays an important role in understanding the characterized behaviors of the superposition process:

Proposition 5. *The superposition of two independent discrete-time semi-Markov processes is also a semi-Markov process.*

The proof of this proposition is given in the Appendix A.5.

Let the probability of going from u to v in k slots be denoted by $q^S_{uv}(k)$ or the probability that knowing the current state is $\mathbf{u} = (X^1_1, X^2_1)$, the next transition is to $\mathbf{v} = (X^1_2, X^2_2)$ and it takes k units to go from $\mathbf{u}$ to $\mathbf{v}$ hence $\mathbf{q_{uv}}(k)$ is the semi Markov kernel of the superposition process. We have the function $q^S_{uv}(k)$ depending on the value of $t_i(v)$ for all processes i. So there are 2 cases: Case I. Only process i changes state at v or $t_i(v) = 1, t_j(v) = 0$. The probability of this event to happen is equal to the probability that the residual life-time in state $x_i(u)$ is equal k can be given by

$$\phi_{uv}(k)^i = \hat{q}^m_{x_i(u)x_i(v)}(k). \tag{2.25}$$

We have process j not changing state at v, the probability of this event to occur is equal to the probability that the sojourn time in state $x_j(u)$ is greater k can be given by

$$\phi_{uv}(k)^j = 1 - \tilde{q}^m_{x_j(u)x_j(v)}(k). \tag{2.26}$$

Due to the independence of these component processes,

$$q^S_{uv}(k) = \phi_i(u, v, k)\phi_j(u, v, k) = \hat{q}^m_{x_i(u)x_i(v)}(k)(1 - \tilde{q}^m_{x_j(u)x_j(v)}(k)) \tag{2.27}$$

The transition probability from state $\mathbf{u}$ to state $\mathbf{v}$ is $\tilde{p}_{\mathbf{uv}}$ and equals to

$$\tilde{p}_{\mathbf{uv}} = \sum_{k\geq 1} \mathbf{q^S_{uv}(k)}. \tag{2.28}$$

Case II. Both processes i and j change state at v or $t_i(v) = t_j(v) = 1$. The probability of this event to happen is equal to the probability that the residual life-time in state $x_i(u)$ and $x_j(u)$ is equal k can be given by

$$q^S_{uv}(k) = \hat{q}^m_{x_i(u)x_i(v)}(k)\hat{q}^m_{x_j(u)x_j(v)}(k). \tag{2.29}$$

As a case study, we next apply the above method to find the transition probability matrix and sojourn time distribution of two independent ON-OFF discrete time semi Markov processes(SMP) in order to capture the collision between transmissions of two non-coordinated nodes, e.g. USRP1 and USRP3 in our experiment. SMP 1 has 2 states (idle1, A) and SMP 2 has 2 states (idle2,B). Suppose that the sojourn time distribution of each of the two states for SMP 1 and SMP 2 specified in Table 2.2:

TABLE 2.2. Duration parameters

State	State notation	Duration distribution Parameter(s)
idle1	0_A	$h^{0_A}(k)$
A	1_A	$h^{1_A}(k)$
idle2	0_B	$h^{0_B}(k)$
B	1_B	$h^{1_B}(k)$

The resulting superposition process is a semi Markov process and due to the combination of the two semi Markov chain as above, we would have 4 states as specified in Table 2.3.

TABLE 2.3. Superposition results

State	idle1	A
idle2	0=(0_A,0_B)	1=(1_A,0_B)
B	2=(0_A,1_B)	3=(1_A,1_B)

Analytical solutions

Our main contribution for the superposition problem is that we can show how to construct the transition from any state i to j in state space E of the superposition process. More importantly, when extending our results to superposition of $N > 2$ number of independent semi Markov processes, the size of the resulting state space is in the order of $O(2^N)$, less than $O(4^N)$ as needed using approaches in existing works [6].

Proposition 6. *From the above analysis, the semi Markov kernels of the superposition from two independent semi- Markov processes each with sojourn time distributions* $h^{0_A}(k), h^{1_A}(k), h^{0_B}(k), h^{1_B}(k)$ *can be given by:*

$$
\begin{aligned}
q^S_{01}(k) &= \frac{1 - \sum_{l=1}^{k} h^{0_A}(k)}{\sum_{k=0}^{\infty} kh^{0_A}(k)}(1 - \sum_{l=1}^{k} \frac{1 - \sum_{m=1}^{l-1} h^{0_B}(k)}{\sum_{k=0}^{\infty} kh^{0_B}(k)}) \\
q^S_{03}(k) &= \frac{1 - \sum_{l=1}^{k} h^{0_A}(k)}{\sum_{k=0}^{\infty} kh^{0_A}(k)} \frac{1 - \sum_{l=1}^{k} h^{0_B}(k)}{\sum_{k=0}^{\infty} kh^{0_B}(k)}
\end{aligned} \tag{2.30}
$$

where the rest of the expressions for $q^S_{uv}(k)$ *can be written similarly and can be found in the Appendix A.6.*

Reverse superposition problem:

Suppose that there are two independent ON-OFF discrete time semi Markov processes(SMP): SMP 1 has 2 states (idle1, A) and SMP 2 has 2 states (idle2,B) Suppose that the sojourn time distribution of each of the two states for SMP 1 and SMP 2 follows their own distributions with parameters specified in the Table 2.2. The resulting superposition process is a semi Markov process and due to the superposition of the two semi Markov chain as above, we would have 4 states as from Table 2.3. Suppose that after the superposition we get the Y sequence and the statistical properties of the Y sequence are known, namely the transition probability matrix and the sojourn time distribution of the Y sequence. We want to find the statistics of the two original sequences.

Follow the results of superposition process we obtain the semi-Markov kernels from Equation (2.30). The unknowns for us are the functions $h^i(k), i \in \{0_A, 0_B, 1_A, 1_B\}$. In order to solve for $h^i(k), i \in \{0_A, 0_B, 1_A, 1_B\}$ of the original two sequences, we need to find conditions the resulting superposed process has to satisfy in terms of its semi-Markov kernels.

Proposition 7. *After superposition of two independent semi-Markov processes, the semi-Markov kernels of the resulting superposed process satisfy the following conditions:*

$$
\frac{q^S_{02}(k)}{q^S_{03}(k)} = \frac{q^S_{20}(k)}{q^S_{21}(k)}; \frac{q^S_{13}(k)}{q^S_{12}(k)} = \frac{q^S_{31}(k)}{q^S_{30}(k)}; \frac{q^S_{01}(k)}{q^S_{03}(k)} = \frac{q^S_{10}(k)}{q^S_{12}(k)}; \frac{q^S_{23}(k)}{q^S_{21}(k)} = \frac{q^S_{32}(k)}{q^S_{30}(k)} \tag{2.31}
$$

If one of the four conditions is violated then we cannot find $h^i(k), i = 0_A, 0_B, 1_A, 1_B$,i.e. there is no superposition process.

The proof is given in Appendix E.

Proposition 8. *There is a unique solution for the reverse superposition problem.*

Now, suppose the initial sojourn time distributions $h^i(k), i = 0_A, 0_B, 1_A, 1B$ all satisfy the above conditions. From the relationship: $\frac{A_1}{A_2} = \frac{q^S_{03}(k)}{q^S_{02}(k)}$ and $A_2(k) = \sum_{l=1}^{k} A_1$, we have $A_2(k) = \frac{q^S_{02}(k)}{q^S_{03}(k)} A_1(k) = \sum_{l=1}^{k} A_1$. When $k = 0$ we have $A_2(0) = 0$ and we have $A_1(0) = \frac{q^S_{03}(k)}{q^S_{02}(k)}$. We can solve for $A_1(k)$ iteratively from the above equation and from that we can write $A_2(k)$ as: $A_2(k) = \frac{q^S_{02}(k)}{q^S_{03}(k)} A_1(k)$ so $A_2(k)$ can be computed from $A_1(k)$ directly. From that then applying the z-transform for equations (5) and (6) and solving for $H_{0A}(z)$ we have:

$$H_{0A}(z) = \frac{\frac{z}{z-1} - A_2(z)}{(\frac{z}{z-1})^2} \frac{(z-1)^2}{z} + \frac{z}{z-1} A_1(z) \tag{2.32}$$

Similarly, we can have the same relationship for $H_{1_A}(z)$, $H_{0_B}(z)$, $H_{1_B}(z)$ as following: $H_{1_A}(z) = \frac{\frac{z}{z-1} - B_2(z)}{(\frac{z}{z-1})^2} \frac{(z-1)^2}{z} + \frac{z}{z-1} B_1(z)$; $H_{0_B}(z) = \frac{\frac{z}{z-1} - C_2(z)}{(\frac{z}{z-1})^2} \frac{(z-1)^2}{z} + \frac{z}{z-1} C_1(z)$; $H_{1_B}(z) = \frac{\frac{z}{z-1} - D_2(z)}{(\frac{z}{z-1})^2} \frac{(z-1)^2}{z} + \frac{z}{z-1} D_1(z)$.

We also can write the following relation: $A_1(z) = \frac{z}{z-1} - \frac{z}{z-1} A_2(z)$. So $H_{0_A}(z)$ only depends on $A_1(z)$. Taking the inverse z-transform we have the form of $h^i(k), i \in \{0_A, 0_B, 1_A, 1_B\}$ which satisfy the condition (26).

2.2.4 Mislabelling two states of the superposition two independent discrete-time semi-Markov processes

It has been shown in the previous section how superposition affects the resulting semi-Markov process's statistics. In the presence of both collision which results in superposition, and near-far effect due to un-evenly distributed receiving powers from colliding active transmission, which causes mislabeling, we next address the problem of recovering the statistics of the original component semi-Markov processes. More specifically, in accordance with our experimental set-up, we still consider superposition of two indepen-

dent binary state semi-Markov processes as in the last section. Our goal is to find the transition probabilities and sojourn time distributions of the mislabeled process first, and then the corresponding statistics of the component ones given the mislabeled ones in the reverse problem.

Forward mislabelling problem:
We remark that our assumption is B process has a much higher received power at the monitoring node than that of node A, and thus state 2, and state 3 are both labelled as state 5. So after the mislabelling, the new semi Markov process has three states denoted as:

$$\text{state } 0 = (0_A, 0_B)$$

$$\text{state } 1 = (1_A, 0_B)$$

$$\text{state } 5 = (0_A, 1_B) \cup (1_A, 1_B)$$

Proposition 9. *For the mislabeled process, the relationship between the semi-Markov kernel function $q_{ij}^{S,M}(k)$ of the mislabelling process and the semi-Markov kernel function $q_{ij}^{S}(k)$ of the superposition is given by:*

$$\begin{aligned} q_{01}^{S,M}(k) = q_{01}^{S}(k); q_{05}^{S,M}(k) = q_{02}^{S}(k) + q_{03}^{S}(k); q_{10}^{S,M}(k) = q_{10}^{S}(k); q_{15}^{S,M}(k) = q_{12}^{S}(k) + q_{13}^{S}(k) \\ q_{50}^{S,M}(k) = \frac{q_{20}^{S}(k)\pi_2 + q_{30}^{S}(k)\pi_3}{\pi_2 + \pi_3}; q_{51}^{S,M}(k) = \frac{q_{21}^{S}(k)\pi_2 + q_{31}^{S}(k)\pi_3}{\pi_2 + \pi_3} \end{aligned} \tag{2.33}$$

The proof is given in the appendix. The analytical formula for the transition probability matrix of the process is given by

$$\mathbf{P}^{\mathbf{S,M}}_{\mathbf{analytical}} = \begin{bmatrix} 0 & \tilde{p}_{01}^{S} & \tilde{p}_{02}^{S} + \tilde{p}_{03}^{S} \\ \tilde{p}_{10}^{S} & 0 & \tilde{p}_{12}^{S} + \tilde{p}_{13}^{S} \\ \frac{\tilde{p}_{20}^{S}\pi_2 + \tilde{p}_{30}^{S}\pi_3}{\pi_2+\pi_3} & \frac{\tilde{p}_{21}^{S}\pi_2 + \tilde{p}_{32}^{S}\pi_3}{\pi_2+\pi_3} & 0 \end{bmatrix}$$

This derivation is given from the fact that $\tilde{p}_{ij}^{S,M} = \tilde{q}_{ij}^{S,M}(\infty) = \sum_{k=0}^{\infty} q_{ij}^{S,M}(k)$. The analytical formula for the duration distribution of the combination process is given by

$q_{ij}^{S,M}(k) = \tilde{p}_{ij}^{S,M} h_{ij}^{S,M}(k)$. From that the sojourn time distributions of the three states $0, 1, 5$ can be given analytically by:

$$q_{01}^{S,M}(k) = \frac{1 - \sum_{l=1}^{k} h^{0_A}(k)}{\sum_{k=0}^{\infty} k h^{0_A}(k)} (1 - \sum_{l=1}^{k} \frac{1 - \sum_{m=1}^{l-1} h^{0_B}(k)}{\sum_{k=0}^{\infty} k h^{0_B}(k)}) \tag{2.34}$$

Similar equations can be derived in the same way and given in the Appendix A.7.

Reverse mislabelling problem

We want to find the statistics of the two original sequences.

Follow the results of superposition process we obtain the semi-Markov kernels from Eq. (30). The unknowns for us are the functions $h^i(k), i = 0_A, 0_B, 1_A, 1_B$ for the superposition process.

Proposition 10. *There is a unique solution for the reverse mislabelling problem.*

Now from the forward mislabelling problem, we have:

$$q_{05}^{S,M}(k) = q_{02}^{S}(k) + q_{03}^{S}(k) = C_1(k); q_{15}^{S,M}(k) = q_{12}^{S}(k) + q_{13}^{S}(k) = B_1(k) \tag{2.35}$$

Similarly, we can find the expressions for $A_1(k)$ and $D_1(k)$. We also can write the following relation:$A_1(z) = \frac{z}{z-1} - \frac{z}{z-1} A_2(z)$. So $H_0(z)$ only depends on $A_1(z)$. Taking the inverse z-transform of $A_1(z), B_1(z), C_1(z), D_1(z)$ we have the form of $h^i(k), i = 0_A, 0_B, 1_A, 1_B$ which satisfy the condition (26).

Difference between a coordinated 3-state semi-Makov, and a mislabeled 3-state Markov out of 4-state superposed from two independent Markov processes:

Next, we provide some comments on the differences between a coordinated 3-state Semi-Makov process, and a mislabeled 3-state Markov process out of 4-state superposed from two independent Markov processes with and without down-sampling. As mentioned from our problem formulation, one question raised in our studies is given the observed random process after these operations, how much we can restore the statistics of the original sequence? This section summarizes the previous results from down-sampling problem and superposition problem to help answer this question directly.

Remarks: One observation about the difference between a coordinated 3-state Semi-Markov, and a mislabeled 3-state Markov out of 4-state superposed from two independent

Markov processes is that for the mislabelling the superposition of two processes, the sojourn time distribution and semi Markov kernel follow the form as specified by the equations in (31) and (39). The conditions of the beginning two superposition processes also need to be satisfied as in (26). These boundary conditions tell us the differences of the class of mislabelling superposition process and the class of coordinated 3-state Semi-Markov processes. With down-sampling issue, however, we cannot tell the difference between two sets. Since down-sampling operation completely destroys the statistics of the original sequence, we could have a result with the statistics of a given coordinated 3-state Semi-Markov the same as the statistics of another mislabeled 3-state Markov out of 4-state superposed from two independent Markov processes. It is thus impossible to recover the statistics of the original sequence, namely the probability transition matrix of the embedded Markov chain and the sojourn time distribution before down-sampling given the observed semi Markov model.

2.3 Simulation and numerical results

In this section, we provide case studies using simulation and numerical results to demonstrate the validity of our theoretical studies regarding the effects of downsampling, superposition, and mislabeling on the statistics of involved semi-Markov processes in both forward and reverse scenarios.

2.3.1 Downsampling

A case study for the dowm-sampling problem:
We apply our approach for a case study of 3-state semi-Markov process. Again our results can be applied to general n states semi-Markov process but we present here a simple case to illustrate our theoretical results. Suppose that a semi-Markov process with 3 states are given by the following transition matrix:

$$\mathbf{P} = \begin{bmatrix} 0 & 0.4 & 0.6 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

There are 3 states in the process: state A and B are busy states and C is an idle state. A, B, and C are states whose duration or number of symbols is subject to duration distribution. Suppose the duration distribution of each state is given by the following table (Table 2.4):

TABLE 2.4. Duration parameters

State	Duration distribution	Parameter(s)
A	Poisson	λ=15
B	Poisson	λ=20
idle	Poisson	λ=9

The probability mass function or the sojourn time distribution of state i can be given by $h_i(k) = \frac{\lambda_i^k exp(-\lambda_i)}{k!}$ From Equations (1) through (7) we can write the $\mathbf{q}$ matrix as:

$$\mathbf{q(k)} = \begin{bmatrix} 0 & 0.4\frac{15^k exp(-15)}{k!} & 0.6\frac{15^k exp(-15)}{k!} \\ \frac{20^k exp(-20)}{k!} & 0 & 0 \\ \frac{9^k exp(-9)}{k!} & 0 & 0 \end{bmatrix}$$

For the downsampling factor of 4, the analytical solution for the transition matrix of the embedded Markov chain for the downsampled sequence (Y) is

$$\mathbf{P^Y_{analytical}} = \begin{bmatrix} 0 & 0.7375 & 0.2625 \\ 0.945 & 0 & 0.055 \\ 0.872 & 0.128 & 0 \end{bmatrix} \tag{2.36}$$

The stationary distribution of the semi Markov chain $Y(k)$ is π= [0.1784 0.4161 0.4055]. We set up the simulation using MATLAB. First we generate the X sequence from its transition probability matrix and sojourn time distribution as specified from above. Then we perform the down-sampling operation by keeping the first symbol, deleting the next 3 symbols and so on to form the X sequence. By estimation, we find the following transition matrix for simulation result of the Y sequence

$$\mathbf{P^Y_{sim}} = \begin{bmatrix} 0 & 0.7651 & 0.2349 \\ 0.9101 & 0 & 0.0899 \\ 0.8912 & 0.1188 & 0 \end{bmatrix}$$

We adopt the Frobenius norm as the metric to measure the distance between two matrices $\mathbf{P}^{\mathbf{Y}}_{\mathbf{analytical}}$ and $\mathbf{P}^{\mathbf{Y}}_{\mathbf{sim}}$ [18]. The squared or Frobenius norm of a matrix $\mathbf{A}_{n\times n}$ is defined as $\parallel \mathbf{A} \parallel_F$ or Frobenius norm of A

$$\parallel \mathbf{A} \parallel_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = trace(\mathbf{A}^T A).$$

where $\mathbf{A}^T$ is the transpose of matrix $\mathbf{A}$. Our result indicates that the Frobenius norm of $\mathbf{P}^{\mathbf{Y}}_{\mathbf{analytical}}$ is 1.0197 and the Frobenius norm of $\mathbf{P}^{\mathbf{Y}}_{\mathbf{analytical}} - \mathbf{P}^{\mathbf{Y}}_{\mathbf{sim}}$ equals to 0.075 or 7.36% relative error. Next, we compared the numerical sojourn time distribution of the down-sampled sequence Y with the simulation results from the analytical solution above. The following diagram (Figure 2.7) show the histogram of the analytical result and the simulation results for the case $m = 4$.

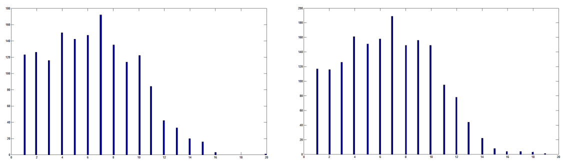

FIGURE 2.7. Histogram of sojourn time distribution for simulation vs analytical

We illustrate above histogram of the sojourn time distribution only for a particular transition from state C to state A, as an example and the rest of the sojourn time distributions can be compared similarly. Then we perform the comparison using two-sample Kolmogorov-Smirnov test (KS test). The KS test returns a test decision for the null hypothesis that the data in vectors from analytical and simulations are from the same distribution while the alternative hypothesis asserts that they are from different distributions [19]. The remaining histograms of the sojourn time distribution h^Y_{ij}, $i, j \in \{0, 1, 2\}$ can be compared in the similar manner and the KS test results accept the null hypothesis that the data in vectors from analytical and simulations are from the same distribution. We can confirm that our analytical results together with the simulation results are agreed within 5% confidence level. A case study for the reverse problem:

Here we set up our study by starting from the down-sampled sequence, namely Y sequence. We apply our analytical results from section 2.3 to find two X sequences. Then

from each of these sequences, we compare their corresponding statistics, i.e. the transition probability matrix and sojourn time distribution. We want to show an example that there are two X sequences with different statistics (different transition probability matrix and sojourn time distribution) that after downsampling can get the same Y sequence. The method to use is: from Equations (2) to (7) we want to show that there are two functions $P_{ij}^X(z)$ that satisfies functional equation (2.35) and each gives a unique transition probability matrix and sojourn time distribution for the X sequence. The transition probability of the embedded Markov chain of the downsampled sequence $\{Y\}$ is given by Equation (18). Given that the sojourn time distribution of each of the three states as calculated from the forward case of the above example. From that following the previous steps we can compute the transition function matrix $P_{ij}^Y(z)$ for the Y sequence. From Equation (14) we propose two functions $Q_{ij}(z)$'s that is $Q_{ij}^1(z) = 0$ and $Q_{ij}^2(z) = z^3$. For the first case we have $P_{ij}^X(z) = P_{ij}^Y(z^2)$ and then the transition matrix is:

$$\mathbf{P^{X1}_{analytical}} = \begin{bmatrix} 0 & 0.4315 & 0.5685 \\ 0.9768 & 0 & 0232 \\ 0.9102 & 0.0898 & 0 \end{bmatrix}$$

For the second case we have $P_{ij}^X(z) = P_{ij}^Y(z^2) + z^3$ and then the transition matrix is:

$$\mathbf{P^{X2}_{analytical}} = \begin{bmatrix} 0 & 0.7054 & 0.2946 \\ 0.3426 & 0 & 0.6574 \\ 0.8901 & 0.1099 & 0 \end{bmatrix}$$

Using this analytical results, the simulation with down-sample 4 and 200 number of runs, taking the average value of results, the transition probability matrix for the Y sequence by simulation results are very close to the analytical solution with the Frobenius norm of $\mathbf{P^Y_{analytical}} - \mathbf{P^Y_{sim1}}$ equals to 0.045 or 2.36% relative error, the Frobenius norm of $\mathbf{P^Y_{analytical}} - \mathbf{P^Y_{sim2}}$ equals to 0.045 or 1.65% relative error. The KS test confirms that the sojourn time distribution are the from the same distribution with 5% confidence interval for the two Y sequence.

Next, we perform KS test for the two X sequences. The returned value of $h = 1$ indicates that KS test rejects the null hypothesis at the default 5% significance level. And the rest of the comparison show that the KS test rejects the null hypothesis that the data in vectors from two X sequences are from the same distribution. We also find that the Frobenius norm of $\mathbf{P^{X1}} - \mathbf{P^{X2}}$ equals to 1.045 or 39.65% relative error.

Remark We can confirm that there are at least two X sequences for the resulting Y sequence with downsampling factor of 4. Such singularity issues persists for any downsampling factor $m > 1$, which means that the statistics of the original discrete time semi-Markov process can not be recovered with a unique solution given its down-sampled semi-Markov sequence, thereby creating a singularity issue due to downsampling.

2.3.2 Superposition

A case study for the superposition problem:

SMP 1 has 2 states (idle1, A) and SMP 2 has 2 states (idle2,B). Suppose that the sojourn time distribution of each of the two states for SMP 1 and SMP 2 follows Poisson distribution with parameters specified in the following table (Table 2.5):

TABLE 2.5. Duration parameters

State	Duration distribution	Parameter(s)
A	Poisson	λ= 15
B	Poisson	λ= 18
idle1	Poisson	λ= 9
idle2	Poisson	λ=12

The resulting superposition process is a semi Markov process and due to the combination of the two semi Markov chain as above, we would have 4 states as Table 2.3 suggests. From the above analysis, the semi Markov kernels of the superposition can be given by: $q^S_{01}(k) = \frac{1-\sum_{l=1}^{k}\frac{e^{-9}9^l}{l!}}{9}(1-\sum_{l=1}^{k}\frac{1-\sum_{m=1}^{l-1}\frac{e^{-12}12^l}{l!}}{12})$. The rest of the expressions $q^S_{ij}(k)$ can be found explicitly from the equation (31) and (32). So for each $q^S_{01}(k)$ we can have two derivations.

The transition probability matrix of the superposition process can be derived by the analytical formulas given as above and is shown to be:

$$\mathbf{P}_{\textbf{analytical}} = \begin{bmatrix} 0 & 0.532 & 0.3911 & 0.0677 \\ 0.4057 & 0 & 0.0561 & 0.5326 \\ 0.3049 & 0.0489 & 0 & 0.6401 \\ 0.0478 & 0.4176 & 0.5309 & 0 \end{bmatrix} \tag{2.37}$$

We do the simulation of the superposition by first simulation Semi Markov process 1 and semi Markov process 2 and make the new denotation as listed in the Table 2.3, there are 4 states for the superposition process. The resulting transition probability matrix of the superposition is given by

$$\mathbf{P}_{\textbf{simulation}} = \begin{bmatrix} 0 & 0.5310 & 0.4208 & 0.0482 \\ 0.4366 & 0 & 0.0414 & 0.5219 \\ 0.3101 & 0.0328 & 0 & 0.6571 \\ 0.029 & 0.4508 & 0.5202 & 0 \end{bmatrix}$$

Our result indicates that the Frobenius norm of $\mathbf{P}^{\textbf{simulation}}_{\textbf{analytical}}$ is equal to 0.095 or 9.36% relative error. Next, we compared the numerical sojourn time distribution of the down-sampled sequence Y with the simulation results from the analytical solution above. The following diagram (Figure 2.8) show the histogram of the analytical result and the simulation results for the case $m = 4$. We illustrate above histogram of the sojourn time

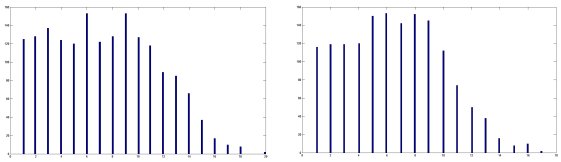

FIGURE 2.8. Histogram of sojourn time distribution for simulation vs analytical

distribution only for a particular transition from state D to state A, as an example and the rest of the sojourn time distributions can be compared similarly. The remaining histograms of the sojourn time distribution h^Y_{ij}, $i, j \in \{0, 1, 2\}$ can be compared in the similar manner and the KS test results accept the null hypothesis that the data in vectors from analytical and simulations are from the same distribution. We can confirm that our analytical results together with the simulation results are agreed within 5% confidence level.

A case study for reverse superposition problem

Suppose that the transition probability matrix of the superposition process can be derived by the analytical formulas given from equation (47). And the sojourn time distribution of the superposition process can be given analytically from the numerical example in the above section (the forward case). So we want to find the sojourn time distributions for the original 2-state semi Markov processes $h^i(k), i = 0_A, 0_B, 1_A, 1_B$. First, we check all the condition equation (35) to equation (37) for the solution to be existed. We apply our analytical results from section IV to find the original 2-state semi Markov sequences analytically. Then from each of these sequences, we compare their corresponding statistics, i.e. the transition probability matrix and sojourn time distribution.

Using this analytical results, the simulation with 200 number of runs, taking the average value of results, the transition probability matrix for the Y sequence by simulation results are very close to the analytical solution with the Frobenius norm of $\mathbf{P}^{\mathbf{Y}}_{\mathbf{analytical}} - \mathbf{P}^{\mathbf{Y}}_{\mathbf{sim}}$ equals to 0.0295 or 7.35% relative error. The KS test confirms that the sojourn time distribution are the from the same distribution with 5% confidence interval for the Y sequence. Figure 2.9 shows the histogram of the sojourn time distribution:

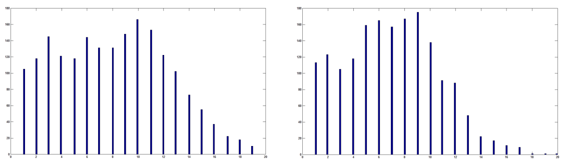

FIGURE 2.9. Histogram of the simulation vs analytical sojourn time distribution Y

2.3.3 A case study for mislabelling problem

Suppose that there are 2 independent discrete time semi Markov processes. SMP 1 has 2 states (idle1, A) and SMP 2 has 2 states (idle2,B). Suppose that the sojourn time distribution of each of the two states for SMP 1 and SMP 2 follows Poisson distribution with parameters specified in the Table 2.4. The resulting superposition process is a semi Markov process and due to the combination of the two semi Markov chain as above, we would have 4 states as the Table 2.3 suggests. The transition probability matrix of the

superposition process can be derived by the analytical formulas as shown from Proposition 9. Hence instead of having 4 states as suggested by the table, we have only 3 states: idle (0), $(idle, A)$ (1) and D (5). We would like to find the transition probability matrix and the sojourn time distribution for the modified semi Markov process as outlined in this part. The analytical solution for the transition probability matrix of the combination process is given by

$$\mathbf{P'}_{\textbf{analytical}} = \begin{bmatrix} 0 & 0.3911 & 0.6089 \\ 0.3049 & 0 & 0.6951 \\ 0.4721 & 0.5279 & 0 \end{bmatrix}$$

Simulation result:

$$\mathbf{P'}_{\textbf{simulation}} = \begin{bmatrix} 0 & 0.4208 & 0.5792 \\ 0.3101 & 0 & 0.6899 \\ 0.4826 & 0.5174 & 0 \end{bmatrix}$$

Next, we compared the numerical sojourn time distribution of the superposition of two semi Markov processes with the simulation results from the analytical solution above. The following diagram (Figure 2.10) shows the histogram of the analytical results and the simulation results.

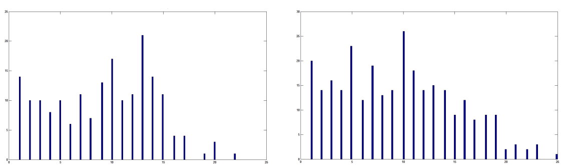

FIGURE 2.10. Histogram of sojourn time distribution for simulation vs analytical

We set up the simulation with 200 number of runs, taking the average value of results, the transition probability matrix for the Y sequence by simulation results are very close to the analytical solution with the Frobenius norm of $\mathbf{P}^{\mathbf{Y}}_{\textbf{analytical}} - \mathbf{P}^{\mathbf{Y}}_{\textbf{sim}}$ equals to 0.0486 or 6.54% relative error. The KS test confirms that the sojourn time distribution are the from the same distribution with 5% confidence interval for the Y sequence.

3

Hidden Markov model, coupled hidden Markov models and applications in social networks

This chapter summarizes and compares different hidden Markov models for modelling the interactions between different users in a dynamic social network.

3.1 Hidden Markov models

As a statistical modelling tool, hidden Markov models were first described in the classic paper by Baum [1]. Shortly afterwards, they were applied to automatic speech recognition independently at CMU and IBM . Until recently, HMMs have become the predominant approach in speech recognition, subsuming dynamic time warping and outperforming neural networks in most speech recognition tasks.

An HMM can be used to represent a specific unit of speech, such as a phoneme or a word. Most speech recognition systems use phonetic HMMs. However, for small vocabulary tasks in which data collection does not require heavy efforts, word HMMs can be used efficiently to obtain high-accuracy speech recognition systems. Formally, an HMM is defined as

$$\lambda \; = < S, I, F, \mathbf{A}, \mathbf{B}, \mathbf{\Pi} >$$

where S is a set of states with transition arcs defined between the states. Associated with each transition from state i to state j is an output distribution, $b_{ij}(\mathbf{x}) \in \mathbf{B}$, which defines how likely a certain event $\mathbf{x}$ in the observation space is going to happen when the

transition is taken [1]. That is, whenever a transition is taken, a piece of data is observed (or "output") with a certain probability. In addition, a transition probability $a_{ij} \in \mathbf{A}$ is associated with each arc and specifies the likelihood of transit to state j, given that we are currently at state i. I is the set of initial states, F is the set of final states. Any sequence of observations is output by starting from one of the initial states and ending with one of the final states. $\pi_i \; \forall \; i \in I$ is the probability that we start from state i. Since $\mathbf{A}$, $\mathbf{B}$, $\mathbf{\Pi}$ are all probabilities,

$$\begin{aligned} \sum_{j \in S} a_{ij} &= 1 \quad \forall \; i \; \in S, \\ \int_{\mathbf{X}} b_{ij}(\mathbf{x}) d\mathbf{x} &= 1 \quad \text{for all transitions } i \text{ to } j \;\; \text{if } \mathbf{x} \text{ is a continuous vector,} \\ \sum_{k} b_{ij}(k) &= 1 \quad \text{for all transitions } i \text{ to } j \;\; \text{if } \mathbf{x} = \mathrm{k} \text{ is a discrete symbol,} \\ \sum_{i \in I} \pi_i &= 1 \end{aligned} \tag{3.1}$$

Starting from a certain initial state (unknown or hidden), we observe a sequence of data. We know these data are emitted (output) by the HMM and that the stochastic process ends at one of the final states. The observed event is the sequence of data output by the HMM; the *hidden* part is the sequence of states the observed data have gone through. In other words, we don't know which sequence of transitions output the observed data.

3.2 Coupled hidden Markov model architectures and related backgrounds

Two hidden Markov models are coupled by introducing the conditional probabilities between their hidden state variables. The state of one model at time t depends on the states of all models (including itself) at time $t-1$. For C hidden Markov models coupled together, the state transition probability is described as $Pr(S_t^C | S_{t-1}^1, S_{t-1}^2, .., S_{t-1}^C)$ instead of $Pr(S_t^C | S_{t-1}^C)$ where the superscript C represents the C^{th} hidden Markov model. The state transition matrix is described by a $C+1$ dimensional matrix and the number of the parameters for this transition probability matrix is N^C where N is the number of hidden

[1] Output distributions may also be associated with states instead of arcs. The theory is exactly the same, with only slight modification on notations.

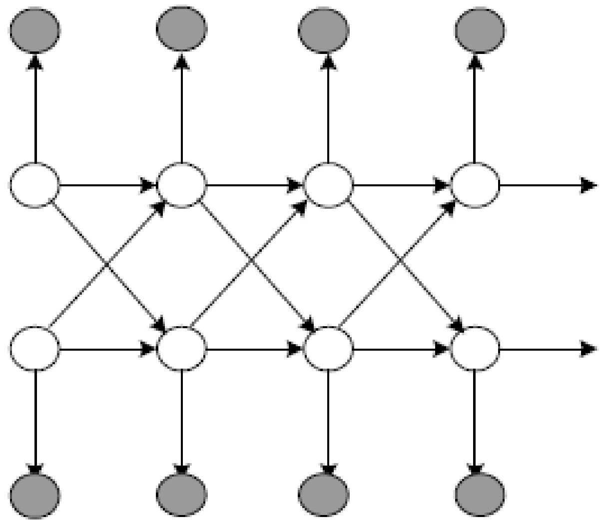

FIGURE 3.1. Fully- coupled hidden Markov model

states. Different variants of coupled HMMs have been used in diverse settings including models for complex human actions and behaviors, freeway traffic , audio-visual speech , EEG classification , spread of infection in social networks , etc. The following graph (Figure 3.1) shows the fully-coupled hidden Markov model as derived from [20] and [21].

Follow these models, Raghavan and colleagues [21] developed the coupled hidden Markov model as the inter-connected dynamics of user activity in a social network. The individual dynamics of each user is coupled to the aggregated activity profile of the neighbors (namely friends or followers) in the network. To model activity profile of a specific user in a social network, his proposed model is illustrated below (Figure 3.2):

Here let T_i, $i = 0, 1, 2, .., N$ denote the time-stamps of a specific user's tweets over the period of interest. The following parameters are relevant from the proposed model:

- **Observations**: Define the inter-tweet duration δ_i, $i = 0, 1, 2, .., N$ as

$$\delta_i = T_i - T_{i-1} \quad (1)$$

- **Hidden states**: Suppose that a variable Q_i, $i = 0, 1, 2, .., N$ reflect the state of the user of interest. For the simplest 2 state Hidden Markov model, $Q_i \in (0, 1)$ where

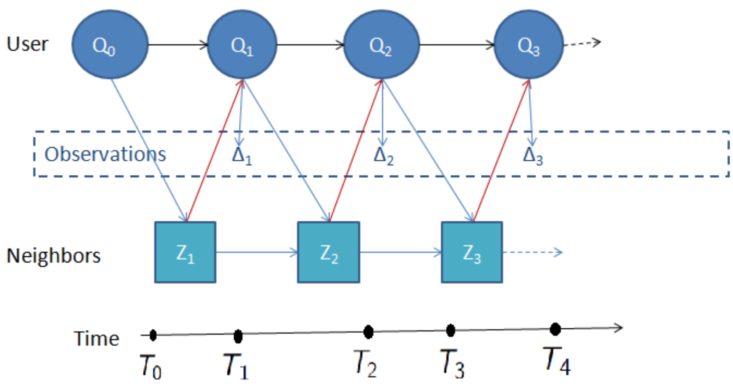

FIGURE 3.2. The coupled hidden Markov model with mentions from social network

$Q_i = 0$ denotes that the user is in an Inactive state between two consecutive tweets T_i and T_{i-1} and $Q_i = 1$ denotes that the user is in an Active state.

The state transition probability matrix is given by

$$P[m, n] = Pr(Q_i = n|Q_{i-1} = m) \quad (1)$$

where $m, n \in (0, 1)$

The prior probability of the initial state Q_0 is $Pr(Q_0 = j) = \pi_j$ where $j \in (0, 1)$. In general, Q_i is hidden (unobservable) and we can only observe δ_i, $i = 0, 1, 2, .., N$ or equivalently T_i. Suppose that the duration distribution of the observations are dependent on the state Q_i and we assume that

$\delta_i \sim f_1(.)$ if $Q_i = 1$

$\delta_i \sim f_0(.)$ if $Q_i = 0$

For the experimental setup, the distribution $f(.)$ are either Weibull or Gamma density.

- **Influence from the neighbors**: Let the variable Z_i, $i = 0, 1, 2, .., N$ capture the influence of the neighbors' tweets on the user of interest. In the context Z_i denotes either

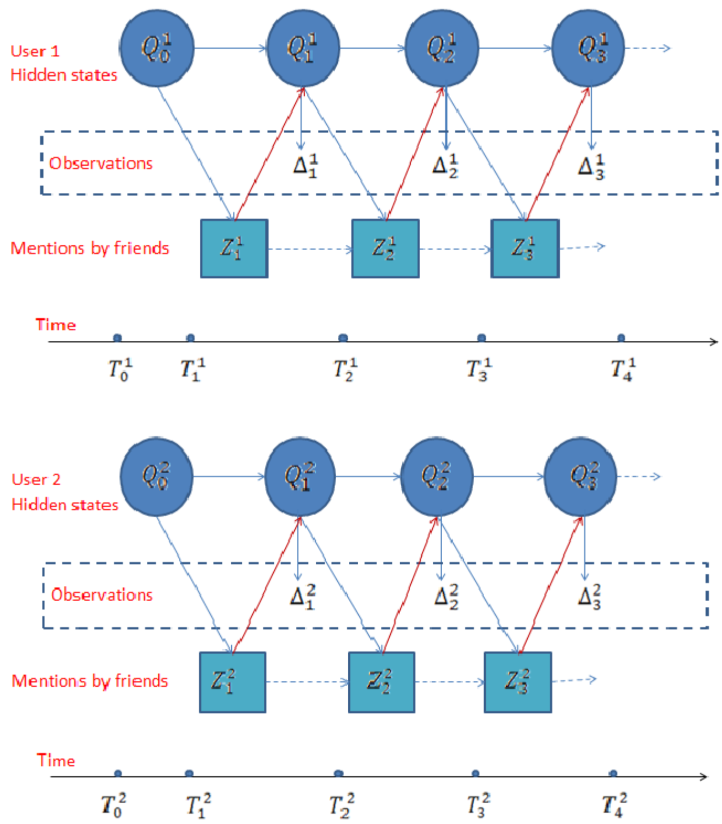

FIGURE 3.3. Two independent users using coupled hidden Markov model

(1) the binary value 1 or 0 that there is a mention from the neighbor's tweets about the specific user or not where $Z_i = 0$ means that there is no mention about the user from his social network

(2) Z_i denotes the number of mentions for that particular user based on total traffic (aggregated activity) of the friends of the user

Noting that Z_i is a function of the activity of all the neighbors (and not a specific user) and let the probability density function of Z_i be given as

$Z_i \sim g_1(.)$ if $Q_{i-1} = 1$

$Z_i \sim g_0(.)$ if $Q_{i-1} = 0$

The following figure (Figure 3.5) mentions the model of two users in a social network: The coupling between Q_i and Z_i is simplified by the Markovian that $Pr(Q_i|Q_1^{i-1}, Z_1^i) = Pr(Q_i|Q_{i-1}, Z_i)$. Suppose that the number of mentions Z_i is

captured by the summary statistic $\phi(Z_i)$ such that:

$$Pr(Q_i|Q_{i-1}, Z_i) = P_0(Q_i|Q_{i-1})(1 - \phi(Z_i)) + P_1(Q_i|Q_{i-1})\phi(Z_i) \quad (2)$$

with $P_k[m,n] = P_k(Q_i = n|Q_{i-1} = m)$ and $k = 0,1$. In particular, the choice $\phi(Z_i) = 1_{Z_i>\tau}$ for a suitable threshold τ implies that the user switches from the transition probability matrix P_0 to P_1 depending on the magnitude of the influence structure. To paraphrase, the user evolves according to a baseline dynamics corresponding to P_0 if his network activity is below a certain threshold and evolves according to an elevated dynamics corresponding to P_1 if his network activity exceeds that threshold.

Z_i is a function of the activity of all the neighbors then hypothesize that: $Pr(Z_i|Q_1^{i-1}, Z_1^{i-1}) = Pr(Z_i|Q_{i-1})$

Hence, the joint density of the observations δ_i, the influence structure Z_i, and the state Q_i can be simplified as

$$Pr(\delta_1^n, Q_0^n, Z_1^n) = Pr(Q_0)\prod_{i=1}^{N} Pr(Z_i|Q_{i-1})\prod_{i=1}^{N} Pr(Q_i|Q_{i-1}, Z_i)\prod_{i=1}^{N} Pr(\delta_i|Q_i) \quad (3)$$

3.3 Proposed coupled hidden Markov model for two known users in a dynamic social network

Remarks on the limitations of the above model and our contribution

- One of the main problems in the modelling is the time line issue. Since there are 2 users and T_i^1 and T_i^2 represent the times of the i^{th} tweets on the same time line, we need to take into account those parameters for all the distributions when derived them. The existing model arose causality issues when using only meta-data, i.e. regardless of the contents of each tweets that the user posts on his Tweeter, the timing and mentioning relationship are the only parameters that needed to be considered. The time evolution of the activities of two users can be characterized by the following diagram (Figure 3.4):

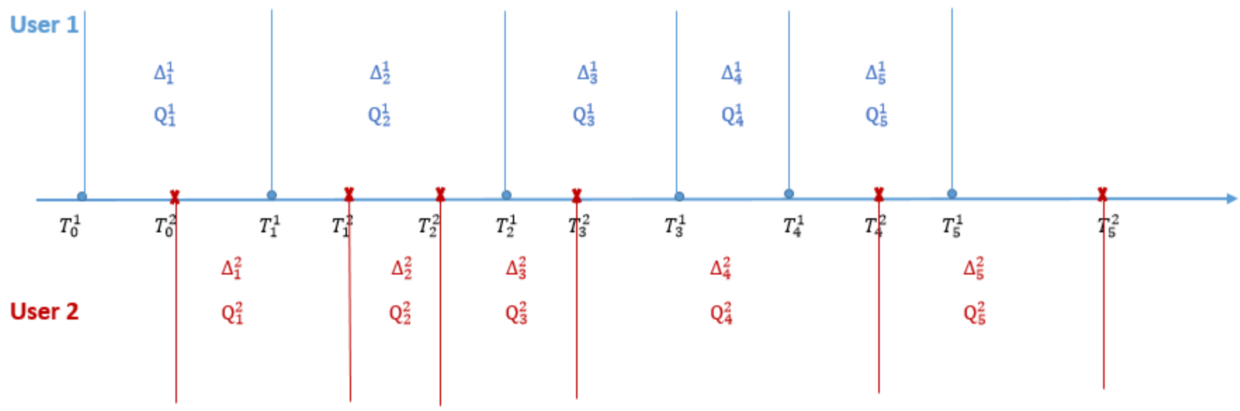

FIGURE 3.4. Time evolution of the hidden states and arrival times

- From the compound processes $N^2(T_i^1)$ and $N^1(T_j^2)$ as derived from the conditional probabilities given above, we want to analytical verifying whether the resulting processes (individual process for user 1 and user 2) are Markov Process.

- This problem can be considered as the superposition of 2 hidden Markov models with arrival times index follow the previous conditions. Again for any sojourn time distribution of user 1 and user 2 and the counting process $N^2(T_i^1)$ and $N^1(T_j^2)$, we want to verify that this process is Markov chain.

  We want to model a coupled hidden Markov model where the mutual interactions between two particular users with the influence from the rest of their social networks. For example, the following figure (Figure 3.5) shows the two tables of a dialogue between two users and their activities in an aggregated network.

  We proposed the following model as the mutual interactions between 2 users versus the rest of their friends in the social networks (Figure 3.6). We define some notations from the model. Let T_i^1, $i = 0, 1, 2, .., N$ denote the time-stamps of a user 1's tweets over the period of interest. Similarly, T_i^2, $i = 0, 1, 2, .., M$ denote the time-stamps of a user 2's tweets over the period of interest. M and N denotes the total number of tweets for user 2 and user 1, respectively, during the period of interest. Define the

Time	User	Tweet
23:56	**User1**	@**User2** why don't you get a car my friends
00:00	**User2**	@**User1** cause my cars transmission blew before i left remember..i may get a new one when i come back
00:01	**User1**	@**User2** ohh and when you come back we must go to chipotle together when is that?
00:03	**User2**	@**User1** DEFINITELY im going there and in an out every dayyyy
00:09	**User1**	@**User2** do you not have a in n out to??
00:16	**User2**	@**User1** no we do but its hella far :(i come back in december
00:19	**User1**	@**User2** my birthday!! I'll drive??
00:23	**User2**	@**User1** im sooo down..my parents wanna get me a convertible bug lol

Table 1: Example two-way dialogue.

ID	Time	User	Tweet
(A)	18:55	**User1**	@**User7** MQM is THE MAFIA! The organized crime in Karachi! .... Now, please! @**User2**
(A)	18:57	**User1**	@**User2** bro, please stop misstating me. I love Karachi and the people of Karachi. But ...
(A)	19:00	**User2**	@**User1** Mafia of MQM makes 70% of Karachi, then wht do U luv hre? The remaing 30%? ...
(B)	19:01	**User3**	@**User4** @**User5** @**User2** v hope 4 a political d judical systm in our country! whoevr fulfil ...
(A)	19:04	**User1**	@**User2** 70% of Karachi is MQM? Really?? Is that how you learn your other 'facts,' too? :) ...
(C)	19:05	**User2**	@**User6** @**User1** @**User5** @**User3** MQM waloon ki Qabar Karachi mai hi hoti hai. Kon apnay ...
(C)	19:06	**User2**	@**User1** ok then app battaa do... Laikin baat tou puri karo ...
(C)	19:09	**User4**	@**User3** @**User5** @**User2** :) No Maseeha other than Imam Mehdi&ESA(AS) will come, ...
(C)	19:10	**User2**	@**User5** @**User6** @**User1** @**User3** wrong, unn ki maiyat gaoo jaati hai, as per tradition. There...
(C)	19:11	**User2**	@**User4** @**User3** @**User5** no doubt about that, but until then we make way.
(C)	19:13	**User1**	@**User2** :) Your 'facts' tell me this discussion would go nowhere. Besides, ...

Table 2: Example of merging two dialogues. User2 was active in dialogue (A) and implicit in dialogue (B) as he was mentioned. The user tied the dialogues together by "mentioning" users from both of these dialogues.

FIGURE 3.5. Examples of dialogues

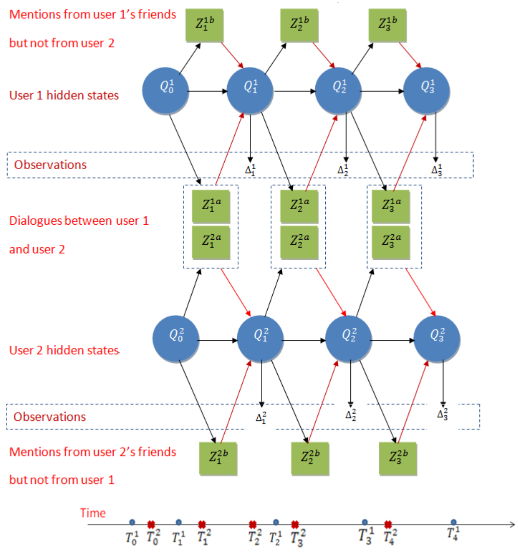

FIGURE 3.6. Proposed 2 users coupled hidden Markov model

inter-tweet duration δ_i^1, $i = 0, 1, 2, .., N$ for user 1 and δ_i^2, $i = 0, 1, 2, .., M$ for user 2 as

$$\delta_i^1 = T_i^1 - T_{i-1}^1 \quad (1)$$

And similar for user 2.

- **Hidden states**: Suppose that a variable Q_i^1, $i = 0, 1, 2, .., N$ reflect the state of user 1. For the simplest 2 state Hidden Markov model, $Q_i^1 \in (0, 1)$ where $Q_i^1 = 0$ denotes that the user 1 is in an Inactive state between two consecutive tweets T_i^1 and T_{i-1}^1 and $Q_i^1 = 1$ denotes that the user is in an Active state.

  The state transition probability matrix is given by

  $$P[m, n] = Pr(Q_i^1 = n | Q_{i-1}^1 = m)$$

  where $m, n \in (0, 1)$
  Similar derivation is for user 2.

- **Influence from the neighbors**: Let the variable Z_i^1, $i = 0, 1, 2, .., N$ capture the influence of the neighbors' tweets on the user 1. In the context Z_i^1 denotes the number of mentions for that particular user based on total traffic (aggregated activity) of the friends of the user. We further split Z_i^1 as two components:

  1. Denote Z_i^{1a} as the number of mentions from USER 2 about user 1
  2. Denote Z_i^{1b} as the number of mentions from his other friends rather than user 2

  Similar definitions for user 2

  1. Denote Z_i^{2a} as the number of mentions from USER 1 about user 2
  2. Denote Z_i^{2b} as the number of mentions from his other friends rather than user 1

The coupling between Q_i^1 and Z_i^1 is by the Markovian condition that $Pr(Q_i^1|Q_{i-1}^1, Z_i^{1a}, Z_i^{1b}, Z_i^{2a})$. Similarly, the coupling between Q_i^2 and Z_i^2 is by the condition that
$Pr(Q_i^2|Q_{i-1}^2, Z_i^{2a}, Z_i^{2b}, Z_i^{1a})$

Since Z_i^k, $k = 1$ or $k = 2$ is a function of the activity of all the neighbors then we have:
$Pr(Z_i^{1a}, Z_i^{1b}, Z_i^{2a}, Z_i^{2b}|Q_{i-1}^1, Q_{i-1}^2)$

We want to write the joint probability in terms of different distribution:

$Pr(Q_0^1, Q_1^1, .., Q_n^1, T_0^1, T_1^1, .., T_n^1, T_0^2, T_1^2, .., T_n^2, Z_1^1, Z_2^1, .., Z_n^1)$

One of the main problem in the modelling is the time line issues or causality problems. Since there are 2 users and T_i^1 and T_i^2 represents the time of the i^{th} tweets on the same time line we need to take into account those parameters for all the distribution when derived them. With the empirical results to justify the latter two assumptions. For this, we start with three typical users (denoted as User-I and User-II) whose activity over the thirty-day period consists of: i) 807 tweets, 260 mentions, and 16,935 tweets from his social network of 62 friends, and ii) 1,914 tweets, 1,108 mentions, and 10,281 tweets from his social network of 92 friends iii) an extreme case of a highly active user (denoted as User-III) whose activity over the thirty-day period consists of 2,387 tweets, 2,872 mentions, and 58,810 tweets from his social network of 206 friends. Users-I and II do not appear to be popular public figures, whereas User-III is a popular journalist, advocate on many political issues, and an activist.

With this data, [21] use the generalized Baum-Welch algorithm to learn model parameters for a coupled HMM with the number of mentions as the influence structure. They also use the model parameters learned with the generalized Baum-Welch algorithm in the generalized Viterbi algorithm to estimate the most likely state sequence corresponding to the observations.

3.4 Hidden Markov tree models for semantic class induction

The authors start by introducing a new unsupervised method for semantic classes induction. This is achieved by defining a generative model of sentences with latent variables, which aims at capturing semantic roles of words. They require method to be scalable, in order to learn models on large datasets containing tens of millions of sentences [22]. Figure 3.7 shows an example of a dependency tree.

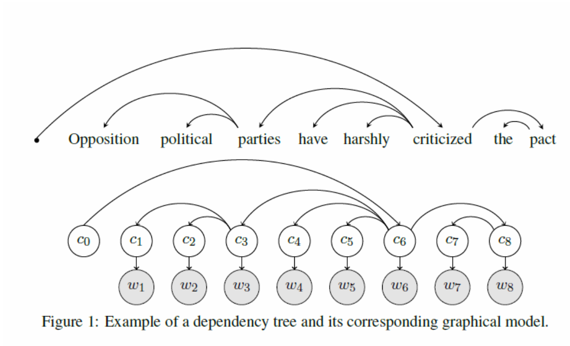

FIGURE 3.7. Example of a dependency tree

In the paper, the authors [22] introduce our probabilistic generative model of sentences. They start by setting up some notations. A sentence is represented by a K-tuple $w = (w_1, w_2, .., w_k)$ where $w_k \in V$ is an integer representing a word and V is the size of the vocabulary. The goal will be to infer a K-tuple $c = (c_1, c_2, ..c_k)$ of semantic classes, where each $c_k \in \{1, 2, .., C\}$ is an integer representing a semantic class, corresponding to the word w_k. The Markov process used to generate the semantic classes will take into account selectional preference. Since to model homonymy, each word can be generated by multiple classes

To describe the Markov process they propose to generate the semantic classes. They assume that we are given a directed tree defined by the function represents the unique parent of the node k and 0 is the root of the tree. Each node, except the root, corresponds to a word of the sentence. First, they generate the semantic class corresponding to the root of the tree and then generate recursively the class for the other nodes. The classes are conditionally independent given the classes of their parents. Using the language of probabilistic graphical models, this means that the distribution of the semantic classes factorizes in the tree.

The results are Hidden Markov tree models also outperform hidden Markov chain models, except for supersense tagging on verbs. We believe that this drop in performance on verbs can be explained because in English the word order (Subject-Verb-Object) is strict, and thus, the chain model is able to differentiate between subject and object, while the tree model treats subject and object in the same way (both are children of the verb).

Moreover, in the tree model, verbs have a lot of children, such as adverbial clauses and auxiliary verbs, which share their transition probability distribution with the subject and the object. These two effects make the disambiguation of verbs more noisy for trees than for chains. Another possible explanation of this drop of performance is that it is due to errors made by the syntactic parser.

We want to give a review about the current literature on Hidden Markov Tree Model and its applications. The recent view of the Hidden Markov Model as a particular case of Bayesian networks has provided new theoretical insights and helped conceiving extensions of the standard model in a sound and formally elegant framework. Whereas standard HMMs are commonly employed for learning in sequential domains, the extension can learn probability distributions on labeled trees which are called Hidden Markov Tree [23]. One of the main advantage of using Hidden Markov Tree model is that unless sharing mechanisms are defined, the parameters of Hidden Markov Tree model vary with the node being considered, yielding a large total number of parameters that may quickly lead to overfitting problems. In HTMM, several forms of stationarity can be assumed, each associated with a different form of parameter sharing. We say that a HTMM is fully stationary if it is both transition and emission stationary. Since the model is a special case of Bayesian network, the two main algorithms (inference and parameter estimation) can be derived as special cases of corresponding algorithms for Bayesian networks. Inference consists of computing all the conditional probabilities of hidden states, given the evidence entered into the observation nodes (i.e. the labels of the tree). Also most notably we have observed that HTMM has a very high capability of learning hierarchical structure in the data.

4

Conclusion and future works

In this Thesis, we have considered the problem of down-sampling a discrete time semi-Markov random process, superposition and combination of two independent discrete time semi-Markov processes. The resulting process, after those operations, is also semi-Markov processes or Markov renewal processes. We show that in this paper the statistics of the original sequence before the superposition operation of two semi Markov processes can be generally recovered. However the statistics of the original sequence cannot be recovered during the down-sampling operation, namely there are multiple solutions to both the sojourn time distribution and probability transition matrix for the original semi- Markov sequence, given the corresponding statistics of the down-sampled one. Our theoretical study as well as subsequent verification by simulation results demonstrate the pitfalls we could deal with when adopting semi-Markov models in characterizing nodes activity patterns in wireless networks.

The results in Chapter 4 indicate that the coupled Hidden Markov Model can characterize the influence of neighboring users for a particular node in a social network. The Hidden Markov Tree model, which uses rich features to capture the degree of association between words and semantic tags, plays an important role in discriminate text relatedness and similarity in computational linguistic.

References

[1] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Annals of Mathematical Statistics*, vol. 37, pp. 1554–1563, 1966.

[2] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989, errata at http://alumni.media.mit.edu/ rahimi/rabiner/rabiner-errata/rabiner-errata.html.

[3] M. J. Borran and R. D. Nowak, "Wavelet-based denoising using hidden markov models," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2001, pp. 3925–3928.

[4] V. Barbu and N. Limnios, *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications: Their Use in Reliability and DNA Analysis.* New York, NY, USA: Springer, 2008.

[5] W. W. T. E. Eliazov, V. Ramaswami and G. Latouche, "Performance of an atm switch: Simulation study," *Proceedings of the IEEE Infocom*, pp. 644–659, 1990.

[6] K. Elsayed, "On the superposition of discrete-time markov renewal processes and application to statistical multiplexing of bursty traffic sources," in *Global Telecommunications Conference, 1994. GLOBECOM '94. Communications: The Global Bridge., IEEE*, vol. 2, 1994, pp. 1113–1117 vol.2.

[7] D. Jung, T. Teixeira, A. Barton-sweeney, and A. Savvides, "Model-based design exploration of wireless sensor node lifetimes," in *In European conference on Wireless Sensor Networks (EWSN) 07*, 2007, pp. 29–31.

[8] M. Kadiyala, D. Shikha, R. Pendse, and N. Jaggi, "Semi-markov process based model for performance analysis of wireless lans," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, 2011, pp. 613–618.

[9] W. P. Cherry and R. L. Disney, "The superposition of two independent markov renewal processes," *Zastosowania Matematyki*, vol. 17, pp. 567–602, 1983.

[10] C. Y. T. Lam, "Stationary distributions for the superposition of markov renewal processes," *Technical Report, Deparment of Industrial and Operations Engineering, The University of Michigan, Ann Arbor*, vol. 90-23, no. 23, pp. 1–22, September 1990.

[11] C. fan Hsin and M. Liu, "Randomly duty-cycled wireless sensor networks: Dynamics of coverage," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 11, pp. 3182–3192, 2006.

[12] B. Carroll, "Learning and identification of wireless network internode dynamics using software defined radio," Master's thesis, LSU, May 2013.

[13] M. S. John Proakis, *Digital Communications, 4th edition*. Canada: McGraw Hill, 2000.

[14] P. D. Vu and S. Wei, "Effects of downsampling on statistics of discrete- time semi-markov processes," in *submitted to IEEE WCNC 2015*, July 2014.

[15] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, Feb. 2013. [Online]. Available: http://dl.acm.org/citation.cfm?id=2502581.2502602

[16] P. D. Vu and S. Wei, "Identification problems in restoring statistics of discrete time semi-markov processes under downsampling and superposition," *Technical report*, March 2014.

[17] M. V. Jelena Kovacevic, *Wavelets and Subband Coding.* Prentice Hall, USA: Springer, 1995.

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd ed*, 3rd ed. Baltimore, MD: Johns Hopkins, 1996.

[19] F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[20] S. Zhong and J. Ghosh, "Hmms and coupled hmms for multi-channel eeg classification," in *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on*, vol. 2, 2002, pp. 1154–1159.

[21] V. Raghavan, G. Ver Steeg, A. Galstyan, and A. Tartakovsky, "Modeling temporal activity patterns in dynamic social networks," *Computational Social Systems, IEEE Transactions on*, vol. 1, no. 1, pp. 89–107, March 2014.

[22] E. Grave, G. Obozinski, F. Bach *et al.*, "Hidden markov tree models for semantic class induction," in *CoNLL-Seventeenth Conference on Computational Natural Language Learning*, 2013.

[23] M. Diligenti, P. Frasconi, and M. Gori, "Hidden tree markov models for document image classification," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 4, pp. 519–523, 2003.

Appendix: Proofs of the main propositions in the thesis

The following are the proofs of the main propositions in Chapter 2.

4.1 Proof of Proposition 1

Proof. We want to show that the downsampled sequence of the semi Markov process is also a semi Markov process. For the finite alphabet set of the state space, we define again our dowmsampling as following: For the down-sample factor of m we keep the first letter and delete the next $m-1$ letters. One observation is that after the downsampling the state space of the resulting sequence is the same as the state space of the original sequence. From Section 2, since X is a homogeneous semi-Markov chain, $q^X_{ij}(k)$ does not depend on n, where from Equation ? we have: $q_{ij}(k) = P(X_{n+1} = j, T_{n+1} = k \mid X_n = i)$. Here the homogeneous property implies that $P(X_{n+1} = j, T_{n+1} = k \mid X_n = i, X_{n-1}, .., X_0, S_n, S_{n-1}, .., S_0) = P(X_{n+1} = j, T_{n+1} = k \mid X_n = i)$. We need to show that for the Y sequence, $P(Y_{n+1} = j, T^Y_{n+1} = k \mid Y_n = i, Y_{n-1}, .., Y_0, S^Y_n, S^Y_{n-1}, .., S^Y_0) = P(Y_{n+1} = j, T^Y_{n+1} = k \mid Y_n = i)$. Equivalently, we can show from our definition in section ? that $P(Y_{n+1} = j, T^Y_{n+1} = k \mid Y_n = i, Y_{n-1}, .., Y_0, S^Y_n, S^Y_{n-1}, .., S^Y_0) = P(Y_{S^Y_{n+1}} = j, S^Y_{n+1} - S^Y_n = k \mid Y_{S^Y_n} = i, Y_{S^Y_{n-1}}, ..., Y_{S^Y_0}, S^Y_n, S^Y_{n-1}, .., S^Y_0)$. Now, from our down-sampling result, $Y_k = X_{2k}$ we have: $P(Y_{S^Y_{n+1}} = j, S^Y_{n+1} - S^Y_n = k \mid Y_{S^Y_n} = i, Y_{S^Y_{n-1}}, ..., Y_{S^Y_0}, S^Y_n, S^Y_{n-1}, .., S^Y_0) = P(X_{2S^Y_{n+1}} = j, S^Y_{n+1} - S^Y_n = k \mid X_{2S^Y_n} = i, X_{2S^Y_{n-1}}, ..., X_{2S^Y_0}, S^Y_n, S^Y_{n-1}, .., S^Y_0)$. From X be the homogeneous semi Markov process and for steady-state result, the transition function $P^X_{ij}(k) = P(X_k = j \mid X_0 = i)$ so we can rewrite the above equation as: $P(Y_{n+1} = j, T^Y_{n+1} = k \mid Y_n = i, Y_{n-1}, .., Y_0, S^Y_n, S^Y_{n-1}, .., S^Y_0) = P(X_{2S^Y_{n+1}} = j, S^Y_{n+1} - S^Y_n = k \mid X_{2S^Y_n} = i = P(Y_{n+1} = j, T^Y_{n+1} = k \mid Y_n = i)$., this concludes the proof of Proposition 1 □

4.2 Proof of Proposition 2

Proof. Here we provide the sketch of the proof. From Equation (2.8), we take the z-transform on both sides of that equation. In terms of matrix form, we have:

$$\begin{vmatrix} 1 & -Q_{12} & -Q_{13} \\ -Q_{21} & 1 & -Q_{23} \\ -Q_{31} & -Q_{32} & 1 \end{vmatrix} \bullet \begin{vmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{vmatrix} = \begin{vmatrix} \tilde{H}_1(z) & 0 & 0 \\ 0 & \tilde{H}_2(z) & 0 \\ 0 & 0 & \tilde{H}_3(z) \end{vmatrix} \tag{4.1}$$

where $\tilde{H}_i(z) = \frac{z}{z-1}(1 - \sum_{j \neq i} \frac{q_{ij}(z)}{q_{ij}(1)})$. So from the element (12) and (13) of the above equation we can solve the system of linear equations: $Q_{12}(z)P_{22}(z) + Q_{13}(z)P_{32}(z) = P_{12}(z)$

$Q_{12}(z)P_{23}(z)+Q_{13}(z)P_{33}(z) = P_{13}(z)$ and then we obtain the solutions as in Equation (12) and (13). Similarly we can apply the same technique and solve for $Q_{21}(z), Q_{23}(z), Q_{31}(z), Q_{32}(z)$. The z-transform relationships between the semi- Markov kernel and its transition function is given by

$$Q_{12}(z) = \begin{vmatrix} P_{12} & P_{13} \\ P_{32} & P_{33} \end{vmatrix} \; / \; \begin{vmatrix} P_{22} & P_{23} \\ P_{32} & P_{33} \end{vmatrix} \tag{4.2}$$

$$Q_{13}(z) = \begin{vmatrix} P_{12} & P_{13} \\ P_{22} & P_{23} \end{vmatrix} \; / \; \begin{vmatrix} P_{22} & P_{23} \\ P_{32} & P_{33} \end{vmatrix} \tag{4.3}$$

$$Q_{21}(z) = \begin{vmatrix} P_{21} & P_{23} \\ P_{31} & P_{33} \end{vmatrix} \; / \; \begin{vmatrix} P_{11} & P_{13} \\ P_{31} & P_{33} \end{vmatrix} \tag{4.4}$$

$$Q_{23}(z) = \begin{vmatrix} P_{21} & P_{23} \\ P_{11} & P_{13} \end{vmatrix} \; / \; \begin{vmatrix} P_{11} & P_{13} \\ P_{31} & P_{33} \end{vmatrix} \tag{4.5}$$

$$Q_{31}(z) = \begin{vmatrix} P_{31} & P_{32} \\ P_{21} & P_{22} \end{vmatrix} \; / \; \begin{vmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{vmatrix} \tag{4.6}$$

$$Q_{32}(z) = \begin{vmatrix} P_{31} & P_{32} \\ P_{11} & P_{12} \end{vmatrix} \; / \; \begin{vmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{vmatrix} \tag{4.7}$$

, this concludes the proof of Proposition 2 □

4.3 Proof of Proposition 5

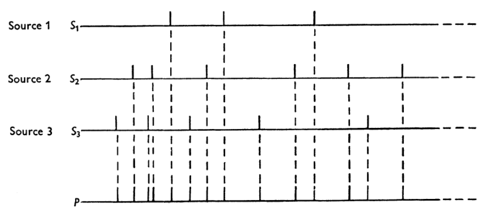

FIGURE 4.1. Arrival times of the superposition process

Proof. We want to show that the superposition of two independent semi Markov processes is also a semi Markov process. We need to show that for the superposition sequence, $P(J^p_{n+1} = j, T^p_{n+1} = k \mid J^p_n = j = i, J^p_{n-1}, .., J^p_0, S^p_n, S^p_{n-1}, .., S^p_0) = P(J^P_{n+1} = j, T^p_{n+1} = k \mid J^p_n = j)$. We have the function $q^S_{uv}(k)$ depends on the value of $t_i(v)$ for all processes $i = 1, 2$. So there are 3 cases: 1. Only process 1 changes state at v or $t_1(v) = 1, t_2(v) = 0$. The probability of this event to happen is equal to $P(J^p_{n+1} = j, T^p_{n+1} = k \mid J^p_n = j = i, J^p_{n-1}, .., J^p_0, S^p_n, S^p_{n-1}, .., S^p_0) = P(J^1_{n_1+1} = j_1, J^2_{n_2+1} = j_2, S^1_{n_1+1} - S^2_{n_2+1} = k \mid J^1_{n_1} = i_1, J^1_{n_1-1}, .., J^1_0, J^2_{n_2+1} = j_2, J^2_{n_2}, .., J^2_0, S^1_{n_1}, S^1_{n_1-1}, .., S^1_0, S^2_{n_2-1}, .., S^2_0$

$$\phi_{uv}(k)^i = \hat{q}^i_{x_i(u)x_i(v)}(k). \tag{4.8}$$

We have process j not changing state at v, the probability of this event to occur is equal to the probability that the sojourn time in state $x_j(u)$ is greater k can be given by

$$\phi_{uv}(k)^j = 1 - \tilde{q}^j_{x_i(u)x_i(v)}(k). \tag{4.9}$$

Due to the independence of these component processes,

$$q^S_{uv}(k) = \phi_i(u,v,k)\phi_j(u,v,k) = \hat{q}^i_{x_i(u)x_i(v)}(k)(1 - \tilde{q}^j_{x_j(u)x_j(v)}(k)) \tag{4.10}$$

2. Both processes i and j change state at v or $t_i(v) = t_j(v) = 1$. The probability of this event to happen is equal to the probability that the residual life-time in state $x_i(u)$ and $x_j(u)$ is equal k can be given by

$$q^S_{uv}(k) = \hat{q}^i_{x_i(u)x_i(v)}(k)\hat{q}^j_{x_j(u)x_j(v)}(k). \tag{4.11}$$

Follow the results of superposition process,
$q(0,1,k) = \frac{1-\sum_{l=1}^k h^0_A(k)}{\sum_{k=0}^\infty kh^0_A(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^0_B(k)}{\sum_{k=0}^\infty kh^0_B(k)})$
$q(0,2,k) = \frac{1-\sum_{l=1}^k h^0_B(k)}{\sum_{k=0}^\infty kh^0_B(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^0_A(k)}{\sum_{k=0}^\infty kh^0_A(k)})$
$q(0,3,k) = \frac{1-\sum_{l=1}^k h^0_A(k)}{\sum_{k=0}^\infty kh^0_A(k)} \frac{1-\sum_{l=1}^k h^0_B(k)}{\sum_{k=0}^\infty kh^0_B(k)}$
$q(1,0,k) = \frac{1-\sum_{l=1}^k h^1_A(k)}{\sum_{k=0}^\infty kh^1_A(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^0_B(k)}{\sum_{k=0}^\infty kh^0_B(k)})$
$q(1,2,k) = \frac{1-\sum_{l=1}^k h^1_A(k)}{\sum_{k=0}^\infty kh^1_A(k)} \frac{1-\sum_{l=1}^k h^0_B(k)}{\sum_{k=0}^\infty kh^0_B(k)}$
$q(1,3,k) = \frac{1-\sum_{l=1}^k h^1_A(k)}{\sum_{k=0}^\infty kh^1_A(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^0_B(k)}{\sum_{k=0}^\infty kh^0_B(k)})$
$q(2,0,k) = \frac{1-\sum_{l=1}^k h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^0_A(k)}{\sum_{k=0}^\infty kh^0_A(k)})$
$q(2,1,k) = \frac{1-\sum_{l=1}^k h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)} \frac{1-\sum_{l=1}^k h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)}$
$q(2,3,k) = \frac{1-\sum_{l=1}^k h^0_A(k)}{\sum_{k=0}^\infty kh^0_A(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)})$
$q(3,0,k) = \frac{1-\sum_{l=1}^k h^1_A(k)}{\sum_{k=0}^\infty kh^1_A(k)} \frac{1-\sum_{l=1}^k h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)}$
$q(3,1,k) = \frac{1-\sum_{l=1}^k h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^1_A(k)}{\sum_{k=0}^\infty kh^1_A(k)})$
$q(3,2,k) = \frac{1-\sum_{l=1}^k h^1_A(k)}{\sum_{k=0}^\infty kh^1_A(k)}(1 - \sum_{l=1}^k \frac{1-\sum_{m=1}^{l-1} h^1_B(k)}{\sum_{k=0}^\infty kh^1_B(k)})$, this concludes the proof of Proposition 5 □

4.4 Proof of Proposition 9

Proof. We can define the element of the kernel $\mathbf{q}^{\mathbf{S,M}}$ for the mislabelling process as

$$q^{S,M}_{ij}(k) = P(X^{S,M}_{n+1} = j, T^{S,M}_{n+1} = k \mid X^{S,M}_n = i). \tag{4.12}$$

Now, we have the straightforward derivations for $q^{S,M}_{01}(k)$, $q^{S,M}_{05}(k)$, $q^{S,M}_{10}(k)$ and $q^{S,M}_{15}(k)$:

$$q^{S,M}_{01}(k) = P(X^{S,M}_{n+1} = 1, T^{S,M}_{n+1} = k \mid X^{S,M}_n = 0) = P(X^S_{n+1} = 1, T^{S,M}_{n+1} = k \mid X^S_n = 0) = q^S_{01}(k). \tag{4.13}$$

Next, we showed that:

$$q^{S,M}_{50}(k) = P(X^{S,M}_{n+1} = 0, T^{S,M}_{n+1} = k \mid X^{S,M}_n = 5) = P(X^S_{n+1} = 0, T^{S,M}_{n+1} = k \mid (X^S_n = 2 or X^S_n = 3)) = \frac{P(X^S_{n+1}=0, T^{S,M}_{n+1}=k, P(X^S_n=2 or X^S_n=3))}{P(X^S_n=2 or X^S_n=3)}. \tag{4.14}$$

Then we can write:

$$q^{S,M}_{50}(k) = \frac{(P(X^S_{n+1}=0, T^{S,M}_{n+1}=k, X^S_n=2) + P(X^S_{n+1}=0, T^{S,M}_{n+1}=k, X^S_n=3))}{P(X^S_n=2 or X^S_n=3)} \tag{4.15}$$

hence,

$$q_{50}^{S,M}(k) = \frac{(P(X_{n+1}^S=0,T_{n+1}^{S,M}=k|X_n^S=2)P(X_n^S=2)+P(X_{n+1}^S=0,T_{n+1}^{S,M}=k|X_n^S=3)P(X_n^S=3))}{P(X_n^S=2 or X_n^S=3)} \tag{4.16}$$

So finally we can get: $q_{50}^{S,M}(k) = \frac{q_{20}^S(k)\pi_2+q_{30}^S(k)\pi_3}{\pi_2+\pi_3}$

A similar proof can be written for: $q_{51}^{S,M}(k) = \frac{q_{21}^S(k)\pi_2+q_{31}^S(k)\pi_3}{\pi_2+\pi_3}$, this concludes the proof of Proposition 9 □

4.5 Proof of Proposition 7

Proof. We want to interpret those conditions and whether they are the only conditions for $h^i(k), i = 0,1,2,3$ or $\tilde{p}_{ij}$ to satisfy. From the 12 equations of the semi Markov kernel, denote the new terms as following:

$$\begin{aligned} A_1 &= \frac{1-\sum_{l=1}^k h^0(k)}{\sum_{k=0}^\infty kh^0(k)}; \\ A_2 &= 1-\sum_{l=1}^k A_1; \\ B_1 &= \frac{1-\sum_{l=1}^k h^1(k)}{\sum_{k=0}^\infty kh^1(k)}; \\ B_2 &= 1-\sum_{l=1}^k B_1; \\ C_1 &= \frac{1-\sum_{l=1}^k h^2(k)}{\sum_{k=0}^\infty kh^2(k)}; \\ C_2 &= 1-\sum_{l=1}^k C_1; \\ D_1 &= \frac{1-\sum_{l=1}^k h^3(k)}{\sum_{k=0}^\infty kh^3(k)}; \\ D_2 &= 1-\sum_{l=1}^k D_1 \end{aligned} \tag{4.17}$$

So we have:

$$\begin{aligned} q_{01}^S(k) &= A_1C_2; q_{02}^S(k) = A_2C_1; \\ q_{03}^S(k) &= A_1C_1; q_{10}^S(k) = B_1C_2; \\ q_{12}^S(k) &= B_1C_1; q_{13}^S(k) = B_2C_1; \\ q_{20}^S(k) &= D_1A_2; q_{21}^S(k) = D_1A_1; \\ q_{23}^S(k) &= D_2A_1; q_{30}^S(k) = B_1D_1; \\ q_{31}^S(k) &= B_2D_1; q_{32}^S(k) = B_1D_2 \end{aligned} \tag{4.18}$$

So we have $C_1 = \frac{q_{02}^S(k)}{A_2}$, $C_2 = \frac{q_{01}^S(k)}{A_1}$, $B_1 = \frac{q_{10}^S(k)A_1}{q_{01}^S(k)}$ $B_2 = \frac{q_{13}^S(k)A_1}{q_{03}^S(k)}$, $D_1 = \frac{q_{20}^S(k)}{A_2}$, $D_2 = \frac{q_{23}^S(k)}{A_1}$
In other word, $A_1, A_2, B_1, B_2, C_1, C_2, D_1, D_2$ can be written as functions of $q_{ij}^S(k)$ and

A_1, A_2. From that we can have the following relationship: $\frac{q^S_{32}(k)}{q^S_{23}(k)} = \frac{q^S_{10}(k)}{q^S_{01}(k)}$ and

$$\frac{A_1}{A_2} = \frac{q^S_{03}(k)}{q^S_{02}(k)} = \frac{q^S_{21}(k)}{q^S_{20}(k)} = \frac{q^S_{12}(k)q^S_{01}(k)}{q^S_{10}(k)q^S_{02}(k)} = \frac{q^S_{30}(k)q^S_{01}(k)}{q^S_{20}(k)q^S_{10}(k)} = \frac{q^S_{31}(k)q^S_{03}(k)}{q^S_{13}(k)q^S_{20}(k)} \tag{4.19}$$

, this concludes the proof of Proposition 7. □

Vita

Phuoc Doan Vu was born in 1989, in Hanoi, Vietnam. He finished his undergraduate studies at Louisiana Tech University in May 2012. He earned a double major degree: bachelor of science degree in mathematics and B.S. in Electrical Engineering from Louisiana Tech University summa cum laude in May 2012. In August 2012 he came to Louisiana State University to pursue graduate studies in Electrical Engineering under the advisory of Dr. Shuangqing Wei. He is a member of the honor society of Phi Kappa Phi and inducted member of Electrical Engineering honor society Eta Kappa Nu.