Title

Forecasting Israeli-Palestinian Conflict with Hidden Markov Models

Author

Major Robert Shearer

Center for Army Analysis

(703) 806 5356

robert.shearer@us.army.mil

Author Statement

The views expressed in this paper are those of the author and do not reflect the official policy or position of the United States Army, Department of Defense, or the United States Government. This paper is not a product of the Center for Army Analysis.

Abstract

This paper presents research into conflict analysis, utilizing Hidden Markov models to capture the patterns of escalation in a conflict and Markov chains to forecast future escalations. Hidden Markov models have an extensive history in a wide variety of pattern classification applications. In these models, an unobserved finite state Markov chain generates observed symbols whose distribution is conditioned on the current state of the chain. Training algorithms estimate model parameters based upon known patterns of symbols. Assignment rules classify unknown patterns according to the likelihood of known models generating the observed symbols. The research presented here utilized much of the Hidden Markov model methodology, but not for pattern classification, rather to identify the underlying finite state Markov chain for a symbol realization. Machine coded newswire story leads provided event data that served as the symbol realization for the Hidden Markov model. Fundamental matrices derived from the Markov chain led to forecasts that provide insight into the dynamic behavior of the conflict and describe potential futures of the conflict in probabilistic terms, to include the likelihood of conflict, the time to conflict, and the time in conflict.

Introduction

"A Markov chain model appears as a particularly appropriate format for analyzing conflict behavior. First, it operates naturally over time-it is a dynamic model. Second, it allows for flexibility in the definition of states of conflict, thereby broadening the class of models that may be considered. Third, various theoretical conceptualizations of conflict imply certain constraints on the form of the transition probabilities and hence can be empirically tested using statistical methods (Duncan and Siverson, 353)." Despite this cogent argument, made nearly twenty years ago, the conflict analysis community has only sparingly utilized Markov models. This paper attempts to reintroduce these models, which have not appeared in the literature in more than four years, to the community. First, I review the history of Markov models in the literature over the last two decades. Second, I discuss both Markov chains and Hidden Markov models, providing a general overview of their structures and applications. Third, I propose a new methodology for conflict analysis that utilizes both Hidden Markov models and Markov chains to forecast conflict. I then apply this methodology to the Israeli-Palestinian conflict, forecast the likelihood of conflict, the time to conflict, and the time in conflict for several periods of the conflict's history, and validate these forecasts.

Previous Applications

Conflict analysis with Markov chains has waxed and waned over the last 40 years. Wilkenfeld and Zinnes (1973) studied the effect of a country's foreign conflict behavior on its domestic conflict behavior with transition matrices. Duncan and Siverson (1975) utilized ergodic chains to model Sino-Indian relations from 1959-1964. Schrodt (1976) proved that one could derive the deterministic Richardson model as the expected value of an ergodic Markov chain with the Chapman-Kolmogorov equations. Ross (1978) suggested the use of Markov chains in modeling guerrilla warfare. Geller (1987) compared probability patterns of domestic political conflict in nations with different system structures. Geller (1990) claimed to use "Markov analysis" when researching the effect of nuclear capabilities of the antagonists in crisis escalation, referring to 2x3 contingency tables as "Markov Matrices" (Markov matrices are by definition square). Geller (1993) later tested several hypotheses regarding the conflictive behavior of a rival dyad with Markov chains. Schrodt (1998, 2000a, and 2000b) applied Hidden Markov models to event data collected from the open press to classify patterns of behavior between nation states that led to conflict. He applied the Hidden Markov methodology to: (1) measure similarities in crises in the Middle East (1998), (2) forecast conflicts in southern Lebanon (2000a), and (3) forecast conflicts in the Balkans (2000b).

Markov Chains

The stochastic process $X_t = \{X_t : t \in 0, 1, \dots \}$ with finite state space Σ is a Markov chain provided that the conditional distribution of any future state X_{t+1} is independent of all past states given the present state X_t, or equivalently

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i).$$

The Markov chain is time homogeneous, if $P(X_{t+1} = j \mid X_t = i)$ is independent of t, in which case define $p_{i,j} = P(X_{t+1} = j \mid X_t = i)$. These conditional probabilities represent the probability of a state transition from state i to state j. The transition matrix, $P=\{p_{i,j}\}$, represents the probability of moving from state i to state j in one transition. Multi-step transition probabilities, denoted by $p^k_{i,j}$, represent the probability of moving from state i to state j in k transitions. State-transition diagrams graphically portray the states and transition probabilities of a Markov chain. Circles represent the states that make up the state space of the model. Arcs represent the one-step transition probabilities, the $p_{i,j}$, from neighboring states. Figure 1 displays an n state Markov chain.

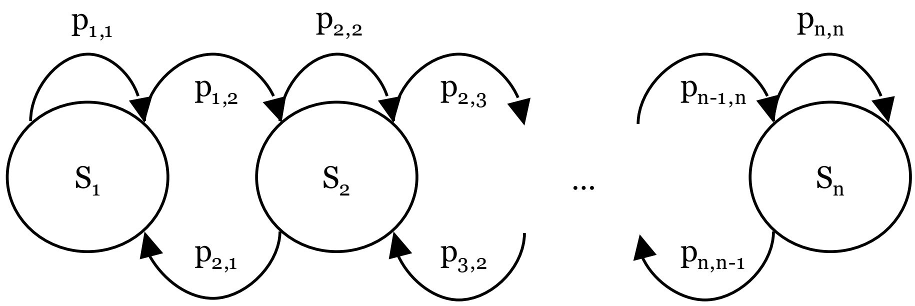

Figure 1

Mathematicians classify states in a Markov chain dependent upon whether it is possible to move from one state to another. Define T as the time until the next visit to state j. State j is *accessible* from state i if $p^k_{i,j} > 0$ for some k transitions in the future. States i and j *communicate* if both are accessible to one another. States that communicate are in the same if there is only one class among the states. Assume the process is currently in state j. State j is *recurrent* if $T < \infty$ with probability one, i.e. the process will eventually return to state j. State j is *transient* if the probability that $T = +\infty$ is nonzero, i.e. it is possible for the process to never again return to state j. State j is *absorbing* if $p_{j,j} = 1$, i.e. the probability of remaining in the state after entering is one. Time homogeneous Markov chains display predictable behavior in the long term, allowing for the forecasting of potential state futures. Forecasts of these behaviors include the distribution of the number of transitions into a state, the distribution of the number of transitions until absorption (assumes one absorbing state), the probability of absorption into an absorbing state (assumes two absorbing states), and the mean percentage of time in a state (assumes an irreducible chain). These forecasts readily avail themselves to likelihood of conflict, time to conflict, and time in conflict interpretations when the states model the spectrum of conflict.

Number of transitions into a state. Define n_j as the number of times that the process is in state j. Define $F_k(i,j)$ as the probability of reaching state j in k transitions, starting in state i, i.e.,

$$F_k(i,j) = \left\{ \begin{array}{ll} p_{i,j} & k = 1 \\ \sum_{b \in E - \{j\}} p_{i,b} F_{k-1}(b,j) & k \geq 2 \end{array} \right\}$$

Summing $F_k(i,j)$ over all k provides the probability of ever reaching state j, starting from state i, i.e.,

$$F(i,j) = \sum_{k=1}^{\infty} F_k(i,j)$$

The probability mass function of n_j is then

$$P_i(n_j = v) = F(i,j)F(j,j)^{v-1}(1-F(j,j))$$

(Çinlar, p. 121-123).

Number of transitions until absorption (time to conflict). Assume a chain with one absorbing state. Define d as the number of transitions (time) that the process passes through, given an initial position at state i, before absorption. The probability mass function for this discrete distribution is geometric, as d counts the number of transitions through transient states (failures) until absorption in the one absorbing state (success), i.e.,

$$P(d = d') = \left(1 - \sum_{t=0}^{\infty} p_{i,j}{}^{t}\right)^{d'-1} \left(\sum_{t=0}^{\infty} p_{i,j}{}^{t}\right)$$

(Mazzuchi, 2003). The interested reader will find derivations for the means and variances of the above motivated distributions in both Çinlar (1975) and Kemeny (1976).

Probability of absorption (likelihood of conflict). Assume a chain with two absorbing states. Define C as the set of absorbing states in the Markov chain with C_j as absorbing state j. Define D as the set of transient states. Define B as the probability of moving from transient state i to an absorbing state j.

$$B(i,j) = \sum_{k \in C_j} P(i,k), \quad i \in D$$

Define $Q(i,j)$ as the matrix obtained from $P(i,j)$ after removing all recurrent states, i.e. $Q(i,j)$ is the transition matrix amongst the transient states. Define B_n

as the probability of absorption into absorbing state *j* from transient state *i* within n transitions. B_n can then be written as a function of Q and B.

$$B_n = (I + Q + \ldots + Q^{n-1})B$$

Taking the limit of B_n as $n \to \infty$ yields the probability of absorption into absorbing state *j* from transient state *i*.

$$\lim_{n \to \infty} B_n = \left(\sum_{k=0}^{\infty} \left(Q^k\right)\right) B = (I - Q)^{-1} B$$

(Çinlar, p. 144-149).

Percentage of time in a state (time in conflict). Assume an irreducible Markov chain. The limiting probabilities of the transition matrix converge as $n \to \infty$ exist and are independent of the current state of the chain. The value of this limit for state j, Π_j, is the unique solution to the following system of linear equations.

$$\Pi_j = \sum_{i \in E} \Pi_i p_{i,j}, \quad j \in E$$
$$\sum_{j \in E} \Pi_j = 1.$$

(Ross, p. 173-174). Çinlar provides a more thorough proof and also proves the uniqueness of the solution (Çinlar, p. 152-153).

Hidden Markov Models

A Hidden Markov model consists of two embedded stochastic processes, X_t and Y_t, with state space $\Sigma \times \Omega$. X_t is an n state, time homogeneous Markov chain with state space Σ and an unobservable (hidden) realization σ. The initial state probability distribution, π, describes the likelihood that X_1 is in one of the n states. The transition matrix, $P=\{p_{i,j}\}$, describes the likelihood that X_t moves from state i to state j. Y_t is an m state, discrete stochastic process conditionally distributed on X_t alone with state space Ω and an observable realization ω. The elements of ω are termed observations and assumed independent. The observation symbol matrix, $O=\{o_{i,j}\}$, describes the likelihood that Y_t generates symbol j given that X_t is in state i. Figure 2 displays an n state, m symbol Hidden Markov model.

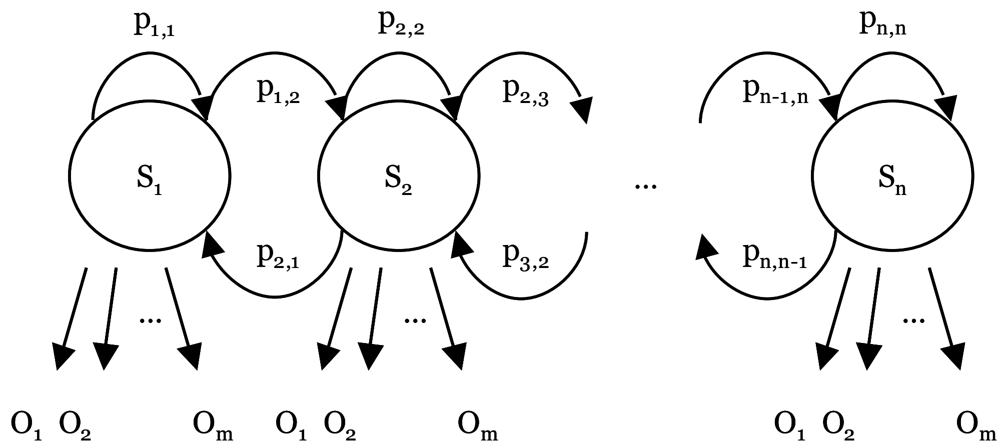

Figure 2

Due to the dependence between X_t and Y_t, the realization σ, although hidden, can be inferred from ω. Complete description of the model requires specification of the three distributions (π, P, and O). Parameter estimation is conducted with the Baum Welch algorithm, a recursive maximum likelihood estimation technique. Another algorithm, the Viterbi algorithm, estimates the most likely state sequence, given the model. In general pattern classification, several Hidden Markov models (referred to as templates) are trained with observation sequences from known classes, and future observation sequences are classified according to the template which was most likely to generate the observed sequence.

Likelihood of an Observation Sequence. The likelihood of a given observation sequence, $P(\omega|\lambda)$, is determined through the use of the forward variable, $\alpha_t(i)$. Define $\alpha_t(i)$ as the probability of observing the partial $(t<T)$ observed realization $\omega = (Y_1(\omega), Y_2(\omega), \ldots Y_t(\omega))$ and having $(X_t = i)$, given the model λ, i.e.,

$$\alpha_t(i) = P(Y_1(\omega), Y_2(\omega), \ldots, Y_t(\omega), X_t = i \mid \lambda) = P(\omega, X_t = i \mid \lambda)$$

To determine $\alpha_t(i)$ for all t, initially determine $\alpha_1(i)$, inductively solve for $\alpha_{t+1}(i)$ for each t, and finally calculate $P(\omega|\lambda)$ as a function of $\alpha_t(i)$. For $t=1$, calculate $\alpha_1(i)$ as the conditional probability of Y_1 generating $Y_1(\omega)$ and X_1 in state i given the model or equivalently as the product of the i, ω_1 element of the observation symbol matrix and the i^{th} element of the initial state probability distribution of X_t.

$$\alpha_1(i) = P(Y_1(\omega), X_1 = i \mid \lambda) = o_{i,Y_1(\omega)}\pi_i, \quad 1 \le i \le N$$

For $t>1$, calculate values of $\alpha_t(i)$ recursively as the conditional probability of observing all $Y_t(\omega)$ from $Y_1(\omega)$ to $Y_{t+1}(\omega)$ andX_{1+1} in state j given the model or equivalently as the product of the product of the sum of the probabilities of all n possible paths from $\alpha_t(i)$ to $\alpha_{t+1}(j)$ and the probability of X_t transitioning to state j from state i and the j, $Y_{t+1}(\omega)$ element of the observation symbol matrix.

$$\begin{aligned}\alpha_{t+1}(j) &= P(Y_1(\omega), Y_2(\omega), \ldots, Y_t(\omega), Y_{t+1}(\omega), X_{t+1} = j \mid \lambda) \\ &= \left[\sum_{i=1}^{n} \alpha_t(i) p_{ij}\right] o_{j, Y_{t+1}(\omega)}, \quad 1 \le i \le T-1, 1 \le j \le N\end{aligned}$$

Calculate the probability of the observed realization given the model as the sum of the final $\alpha_T(i)$ over all n states of X_t.

$$P(\omega \mid \lambda) = \sum_{i=1}^{n} \alpha_T(i)$$

(Rabiner, 262).

Baum-Welch algorithm. Define the backwards variable, $\beta_t(i)$, as the probability of observing the partial realization $\omega_\beta = (Y_{t+1}(\omega), Y_{t+2}(\omega),, \ldots Y_T(\omega))$ with ($X_t = i$), given the model λ.

$$\beta_t(i) = P(Y_{t+1}(\omega), Y_{t+2}(\omega), \ldots, Y_T(\omega), X_t = i \mid \lambda) = P(\omega, X_t = i \mid \lambda)$$

Arbitrarily assume $\beta_T(i)$ to equal one for all i and then inductively solve for $\beta_t(i)$.

$$\begin{aligned}&\beta_t(i) = 1, \quad 1 \le i \le N \\ &\beta_t(i) = \sum_{j=1}^{n} p_{i,j} o_{j, Y_{t+1}(\omega)} \beta_{t+1}(j)\end{aligned}$$

For $t<T$, calculate values of $\beta_t(i)$ recursively as the sum of the products of the probability of X_t transitioning to state j from state i, the probability of Y_t generating ω_{t+1} given X_t in state j, and $\beta_{t+1}(i)$ (Rabiner, 263).

Define $\xi_t(i,j)$ as the probability the process X_t is in state i at time t and state j at time $t+1$, given the complete observed realization $\omega = (Y_1(\omega), Y_2(\omega), \ldots Y_T(\omega))$ and the model λ, i.e.,

$$\xi_t(i,j) = P\left(X_t = i, X_{t+1} = j \mid Y_1(\omega), Y_2(\omega), \ldots, Y_T(\omega), \lambda\right)$$
$$\xi_t(i,j) = \frac{P\left(X_t = i, X_{t+1} = j, Y_1(\omega), Y_2(\omega), \ldots, Y_T(\omega) \mid \lambda\right)}{P\left(Y_1(\omega), Y_2(\omega), \ldots, Y_T(\omega) \mid \lambda\right)}$$
$$= \frac{\alpha_t(i)\, p_{ij}\, o_{j, Y_{t+1}(\omega)}\, \beta_{t+1}(j)}{P(\omega \mid \lambda)}$$

Define $\gamma_t(i)$ as the probability of the process X_t being in state i given the observation sequence, ω, and the model, λ, i.e.,

$$\gamma_t(i) = P\left(X_t = i \mid \omega, \lambda\right)$$
$$\gamma_t(i) = \frac{P\left(X_t = i, \omega \mid \lambda\right)}{P(\omega \mid \lambda)} = \frac{\alpha_t(i)\beta_t(i)}{P(\omega \mid \lambda)}$$
$$\gamma_t(i) = \sum_{j=1}^{n} \xi_t(i,j)$$

The sum of $\gamma_t(i)$ over time equates to the expected number of times that X_t visits state i, or equivalently, the expected number of transitions from state i (excluding $t = T$ from the summation). The sum of $\xi_t(i,j)$ over time equates to the expected number of times that state j is visited from j, or equivalently, the expected number of transitions from state i to state j (Rabiner, 263).

Given a fitted model λ = *(P, O, π)*, utilize the formulas above to re-estimate the parameters of the model, λ′. Baum et al (1966) showed that either λ′ defines a critical point in the response surface of the likelihood function, i.e., λ = λ′, or m model λ' is more likely than model λ, i.e., $P(\Omega|\lambda') > P(\Omega|\lambda)$.

$$\pi_i' = expected\ number\ of\ times\ (frequency)\ in\ state\ i\ at\ time\ (t=1) = \gamma_1(i)$$

$$p_{i,j}' = \frac{expected\ number\ of\ transitions\ from\ state\ i\ to\ state\ j}{expected\ number\ of\ transitions\ from\ state\ i}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$o_{j,k}' = \frac{expected\ number\ of\ times\ in\ state\ j\ and\ observing\ v_k}{expected\ number\ of\ times\ in\ state\ j}$$

$$= \frac{\sum_{t=1}^{T} \gamma_t(j) \quad s.t.\ \omega_t = v_k}{\sum_{t=1}^{T} \gamma_t(i)}$$

Iteratively use λ' in place of λ and repeat the re-estimation procedure, improving on $P(\Omega|\lambda)$ until some limiting point is reached. This final result is a maximum likelihood estimate for λ (Rabiner, 265).

Viterbi algorithm. The Viterbi algorithm determines the most likely state realization $\sigma = (X_1(\sigma), X_2(\sigma), \ldots, X_T(\sigma))$ given the model λ and an observed realization $\omega = (Y_1(\omega), Y_2(\omega), \ldots Y_T(\omega))$. Define $\delta_t(i)$ as the highest probability of an observed realization that ends in state i at time t, given the first t observations, i.e.,

$$\delta_t(i) = \max_{\omega_1,\omega_2,\ldots,\omega_{t-1}} P(X_1(\sigma), X_2(\sigma), \ldots, X_{t-1}(\sigma), X_t(\sigma), Y_1(\omega), Y_2(\omega), \ldots, Y_t(\omega) \mid \lambda)$$

Define $\psi_t(j)$ as the state *j* at time *t* that maximizes the product of the likelihood of the state realization up until time *t-1* and the probability of a state transition to state *j* at time *t*. Initially solve for $\delta_1(i)$ and $\psi_1(i)$.

$$\delta_1(i) = \pi_i o_{i,Y_1(\omega)}$$
$$\psi_1(i) = 0$$

The likelihood of a single observation realization from state *i* equals the product of the probability of starting in state *i* and the probability of generating the observation given state *i*. Recursively solve for subsequent values of *t*.

$$\delta_t(j) = \max_{1\le i\le n}[\delta_{t-1}(i)p_{i,j}]\cdot o_{j,Y_t(\omega)} \quad 2\le t\le T, 1\le j\le n$$

$$\psi_t(j) = \arg\max_{1\le i\le n}[\delta_{t-1}(i)p_{i,j}] \quad 2\le t\le T, 1\le j\le n$$

The likelihood of an observation realization up to time *t* equals the product of the maximum likelihood the realization up to time *t-1*, with a state transition from state *i* to state *j*, and the probability of the observation at time *t* from state *j*. Define r^* as the probability of the state realization with maximum likelihood and s^*_T as the state *i* that maximizes the likelihood of the state realization.

$$r^* = \max_i[\delta_T(i)]$$
$$s^*_T = \arg\max_i[\delta_T(i)]$$

Recursively determine the most likely state realization backwards from s^*_T to s^*_1 with $\psi_t(j)$

$$s^*_t = \psi_{t+1}\left(s^*_{t+1}\right)$$

(Rabiner, 264). Rabiner (1989) provides an excellent tutorial on the application of Hidden Markov models to real processes. See McDonald and Zucchini (1997), Duda (2001), and Theodoridis and Koutroumbas (2003) for more detailed examinations of these algorithms and Hidden Markov models in pattern classification.

Proposed Methodology

Concept. I utilize Hidden Markov models with event data, but not in the classical pattern classification approach taken by Schrodt. Rather, I make use of the maximum likelihood techniques motivated by Baum et al (1966, 1970) for Hidden Markov model parameter estimation in order to estimate the parameters of the underlying finite state Markov chain. I then make inferences from the chain to describe potential futures in probabilistic terms. I execute this methodology in six phases.

1. Obtain event data for a specific conflict.
2. Hypothesize an n-state, m-symbol Hidden Markov model, based upon a spectrum of conflict model.
3. Obtain initial parameter estimates for the model, required for the Baum-Welch estimation algorithm, from a subject matter expert.

4. Estimate the parameter values with the Baum-Welch algorithm and the most likely state sequence with the Viterbi algorithm.
5. Forecast the likelihood of conflict, the likelihood of conflict, and the time in conflict from the underlying finite state Markov chain.
6. Validate the model by comparing time to conflict forecasts with the Viterbi sequence times to conflict.

Data. I selected the Israeli-Palestinian conflict for analysis with the new methodology. The Israeli-Palestinian conflict dates back to initial Arab protests of Jewish setters in 1891 and continues to the present time. The data set for the dyad was obtained from the KEDS project Levant data set. The data set contains 27,679 events between the Israelis and Palestinians from 1979 to 2003, coded daily according to the WEIS protocol. Three distinct periods of conflictive activity appear in the WEIS scores: termed pre-Intifada (1979-1987), 1st Intifada (1988-1998) period, and 2nd Intifada (2001-2003). Training data sets and testing data sets were identified within each of these three periods. The training set for the pre-Intifada ran from April 1979 to August 1983. The test set for the Pre-Intifada ran from August 1983 to November 1987. Both data sets contained 1572 data points. The training set for the 1st Intifada ran from January 1988 to July 1993. The test set for the 1st Intifada ran from July 1993 to December 1998. Both data sets contained 1998 data points. The training set for the 2nd Intifada ran from August 2001 to August 2002. The test set for the 2nd Intifada ran from August 2002 to July 2003. Both data sets contained 357 data points.

Model. I selected a five-state, five-symbol model as an appropriate multi-state conflict model for the dyad. I derived the states from Michael Lund's Spectrum of Conflict model (Lund, p. 39). Lund hypothesized that the spectrum of conflict contains five states: enduring peace – a state that involves a high level of reciprocity and cooperation with a virtual absence of self-defense measures among parties; stable peace - a state of wary communication and limited cooperation; unstable peace - a state of tension and suspicion among the parties with at most sporadic violence; crisis – a state of confrontation between forces that are mobilized and ready to fight and may engage in threats and occasional low-level skirmishes but have not exerted any significant amount of force; and war – a state of sustained fighting between organized armed forces. These states differ according to the level of animosity and the level of violence within the dyad. I made three modifications to Lund's model. First, I dropped the state of enduring peace, as the Israeli-Palestinian dyad has clearly never entered this state. Second, I added an additional state: threatened peace – a state with at most sporadic violence but where both sides perceive the other as enemies. I added this state to more fully capture the dyad's states of conflict. I placed this new state between unstable peace and crisis. Third, I renamed the fifth state war, to conflict, to avoid misunderstandings as to what the methodology forecasts. The five-states of the model then take the following form: 0 - stable peace; 1 - unstable peace; 2 - threatened peace; 3 - crisis; and 4 - conflict. I took the five symbols from conflict pattern classification work done by Schrodt (2000b). In his early efforts in identifying patterns of conflict, Schrodt based the Hidden Markov models symbols upon the 22 WEIS scores and added a 23rd symbol for non-events. He later found that he incurred little increase in classification error when he consolidated the 22 WEIS symbols into four symbols: physical cooperation, verbal cooperation, verbal confrontation, and physical confrontation. Occam's razor argues for the more parsimonious model since *entia non sunt multiplicanda praeter necessitatem* – "entities are not to be multiplied without necessity". The five-symbols of the model then take the following form: 0 – no event; 1 – physical cooperation; 2 – verbal cooperation; 3 – verbal conflict; and 4 – physical conflict. The use of the five-symbol model necessitated a transformation of the data set. First, the event codes from the 23 symbol WEIS scores were mapped into the appropriate five symbol scores, termed modified WEIS in this research. Second, since a large number of days had multiple events, the events were aggregated (mean score rounded up) daily in order to allow the forecasts made from the model to have a temporal interpretation. The resulting data set had 8,872 events.

Forecasts. Three models were fit, one for each period. The training data set for each was used to estimate the model parameters. The precision required for stopping the recursion was set at 0.01, i.e. the algorithm continued to re-estimate the parameters until the sum of the twenty-five differences between corresponding elements of the last and next transition matrices failed to exceed 0.01. After estimating the Hidden Markov model's parameters, the underlying five-state Markov chain was analyzed in order to forecast the dyad's potential futures. Figure 3 displays the underlying Markov chain.

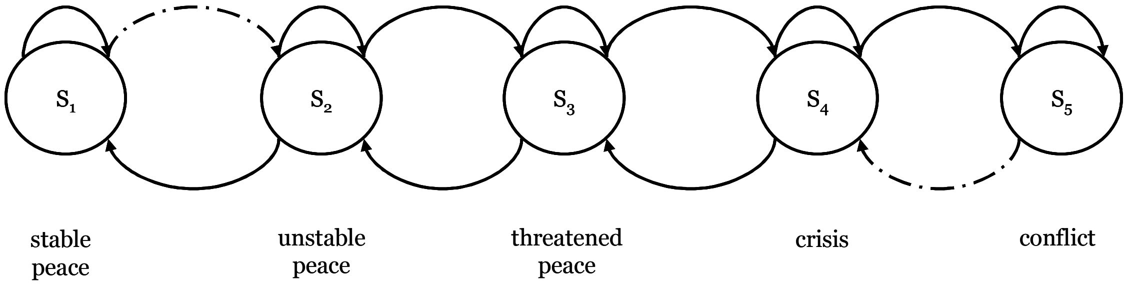

Figure 3

Determination of the likelihood of conflict required the assumption that the first state (stable peace) and the last state (conflict) in the final transition matrix were absorbing. Under this assumption, the process is eventually absorbed into one of these extreme states, as the other three are transient. The likelihood of the process being absorbed into the first state equates to the likelihood of the dyad entering the state of stable peace before conflict. The likelihood of the process being absorbed into the last state equates to the likelihood of the dyad entering the state of conflict before stable peace. Therefore, the probabilities of absorption into each of the two extreme states, given an initial position in one of the transient states, were calculated. Determination of the time to conflict required the assumption that the last state (conflict) in the final transition matrix was absorbing. Under this assumption, the process is eventually absorbed into the last state, as the other four are transient. The number of transitions in a realization of the process resulting in absorption into the last state equates to the time taken by the dyad to move into a state of conflict. Determination of the time in conflict required the assumption that the transition matrix will remain constant for the foreseeable future. Under this assumption, $P^n_{i,j}$ converges to limiting probabilities for each j as $n \rightarrow \infty$. These limiting probabilities equal the proportion of the time that the dyad will spend in each state *j*. Therefore, the limiting probabilities for each *j* were calculated. Figure 4 displays the three forecasts for each time period.

	Intifada Period		
Forecasts	**Pre**	**1st**	**2nd**
Likelihood of Conflict [%] [1]			
Unstable Peace	3	8	33
Threatened Peace	9	37	63
Crisis	35	91	85
Time to Conflict [days] [2]			
Stable Peace	165	53	18
Unstable Peace	164	52	17
Threatened Peace	155	37	13
Crisis	112	7	7
Time in Conflict [%] [3]			
Stable Peace	24	18	7
Unstable Peace	48	32	20
Threatened Peace	16	8	23
Crisis	3	25	34
Conflict	9	17	16

[1] Likelihood that the process will enter the state of conflict before stable peace from the given state for the given period.

[2] Mean number of days that the process will take before entering conflict from the given state for the given period.

[3] Percentage of the time that the process will spend in each of the five states for the given period.

Figure 4

Validation

General. The validation of a statistical forecasting model typically consists of comparing a model's forecasts with actual data. This approach proved impossible with the model presented here in that one cannot compare forecasted states with actual states that are "hidden." Instead, I compared the forecasts against the most likely state sequence, the Viterbi sequence. This approach created its own set of difficulties: (1) it favorably biases the validation as the same model that produced the forecasts also produced the Viterbi sequence; and (2) the validity of the Viterbi sequence remains questionable. I could not remove the bias in the validation and one should view the validation of the forecasts with this fact in mind. I could address the validity of the Viterbi sequence and did so with a face validation of the sequence against known historical trends.

Viterbi Validation. How then to validate the Viterbi sequence? Statisticians working with linear regression models typically graph the estimated regression line over the data as a diagnostic tool to visually validate the linear model. A similar approach is available for the Viterbi sequence and real events. Although the past states are hidden, subject matter experts would typically agree on whether a conflict experienced escalation or de-escalation during significant historical periods. Several significant events in the Israeli-Palestinian conflict were identified and then plotted on the Viterbi sequence for comparison. Figure 5 displays these significant events plotted on the Viterbi sequence (The year markers designate the middle of each year. The state sequence scores were exponentially smoothed to display the trends in the sequence).

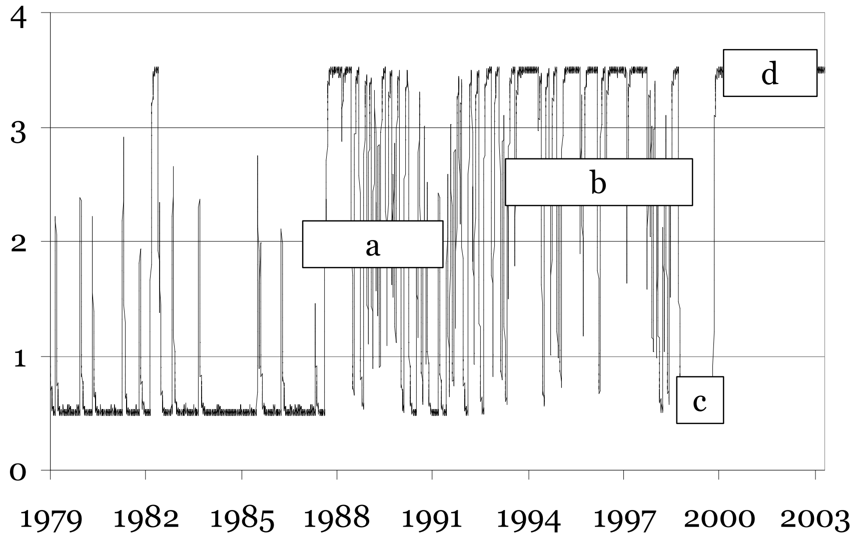

Figure 5

Marker 'a' denotes the First Intifada, beginning in December 1987 and slowly fading in intensity some time in 1991. The uprising began with demonstrations, strikes, and other forms of civil disobedience. Violence increased as the Intifada continued. Escalating conflict brought barricades and stone throwing into the Intifada. Israel arrested many of the leaders of the various Palestinian resistance groups, ending the cohesive force of the Intifada by 1991. The Viterbi sequence captured these trends. The sequence corresponds with the early rise in conflict that accompanied the Intifada, the sustained level of conflict during the end of the 1980s, and the decline in conflict leading to the end of the Intifada in 1991. Marker 'b' denotes the various efforts in the peace process from 1993-1999 (Oslo I - 1993, Oslo II - 1995, Wye River I - 1998, and Wye River II – 1999). Following the Gulf War the Israelis and Palestinians entered a series talks with the goal of settling the conflict. The region experienced sporadic outbursts of violence during this period as extremists on both sides attempted to derail the peace process. The Viterbi sequence corresponds with the volatile nature of the Israeli-Palestinian relationship during this period. The sequence oscillates between unstable peace and conflict, with longer stays in the latter. Marker 'c' denotes the Camp David II talks that occurred in July 2000. Both populations experienced initial optimism for the talks as news leaked of the Israeli's softening of several positions. This optimism led to a decrease in violence in the region. The failure of the two sides to reach a compromise after two weeks of talks again halted the peace process. The Viterbi sequence corresponds with both the de-escalation of conflict during the lead up to and during the talks and the escalation of conflict following the end of the talks. Marker 'd' denotes the Second, or Al-Aqsa, Intifada. This second Intifada began shortly after the collapse of the Camp David II talks. Both sides argue as to what caused the return to violence, but the violence itself is not debated. Both the Israelis and the Palestinians quickly employed greater force than in the first Intifada and have continued to maintain a high level of violence through 2003 (end of the data set). The Viterbi sequence also corresponds with this increased and sustained level of violence.

Model Validation. Assuming that the Viterbi sequence represented the true state sequence, I then compared the forecasted mean times to conflict (arguably the most interesting of the forecasts) from stable peace, unstable peace, threatened peace, and crisis, to the actual mean times from each of these states in the Viterbi sequence. Figure 6 displays the comparisons across the three periods.

	Intifada Period		
Validation	**Pre**	**1st**	**2nd**
Forecasted Time to Conflict [days] [1]			
Stable Peace	165	53	18
Unstable Peace	164	52	17
Threatened Peace	155	37	13
Crisis	112	7	7
Viterbi Time to Conflict [days] [2]			
Stable Peace	221	21	n/a
Unstable Peace	220	20	n/a
Threatened Peace	140	20	n/a
Crisis	66	3	3

[1] Mean number of days that the process will take before entering conflict from the given state for the given period.

[2] Mean number of days that the Viterbi sequence took before entering conflict from a given state for the given period.

Figure 6

The Pre-Intifada time to conflict forecasts closely approximated the Viterbi sequence values. The forecasts underestimated the times to conflict from stable and unstable peace, while overestimating the times from threatened peace and crisis. The 1st Intifada time to conflict forecasts overestimated the Viterbi sequence values. The Viterbi sequence never entered stable, unstable, or stable peace during the 2nd Intifada period. Accordingly, I could only compare the forecasted time to conflict from the state of crisis to the Viterbi sequence. The forecast overestimated the time to conflict. Despite the differences between the forecasts and observed mean number of days to conflict, all of the observed times fell within one standard deviation of the forecasted means.

Conclusion

This paper has presented a novel methodology for modeling conflict escalation: utilizing Hidden Markov models with event data to model a conflict, identifying the underlying finite state Markov chain, and then making forecasts regarding the conflict from this Markov chain. Application of this methodology to the Israeli-Palestinian conflict produced favorable results. The forecasts used to describe the conflict's potential futures, when generated from training sets and compared to test sets, appear valid. The Israeli-Palestinian conflict has undergone three increasingly conflictive phases since 1979. The model captured these three with forecasts that reflect this increased level of conflict. The likelihood of conflict and the time in conflict forecasts increased across the three periods. The time to conflict forecasts decreased across these periods. The Israeli-Palestinian conflict has also displayed great variability in the level of violence experienced during each period. The model captured this variability in the forecasts. The hidden nature of the actual states created difficulties in validating the model. The Viterbi sequence provided a realization of most likely future states in the test sets for validation of the forecasts, but lacked validity in that the sequence is a function of the model. A comparison of the sequence with known events and general trends from the conflict's history provided face validation for the sequence. The time to conflict forecasts appear valid as all of the Viterbi sequence mean times to conflict fell within one standard deviation of the forecasted mean times to conflict.

References

Baum, L.E., and T. Petrie. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains, *Annals of Mathematical Statistics*, No 37, 1554-1563.

Baum, L.E., T. Petrie, G. Soules, and N. Weiss. 1970. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *Annals of Mathematical Statistics*, No 41, 164-171.

Çinlar, E. 1975. *Introduction to Stochastic Processes*. Prentice Hall.

Duda, Richard O., Peter E. Hart, and David G. Stork. 2001. *Pattern Classification*. Wiley-Interscience.

Duncan, G. and R. Siverson. 1975. Markov Chain Models for Conflict Analysis: Results from Sino-Indian Relations, 1959-1964, *International Studies Quarterly*, No 19, 344-374.

Geller, Daniel S. 1987. The Impact of Political System Structure on Probability Patterns of Internal Disorder, *American Journal of Political Science*, No 31, 217-235.

---. 1990. Nuclear Weapons, Deterrence, and Crisis Escalation, *Journal of Conflict Resolution*, No 34, 291-310.

---. 1993. Power Differentials and War in Rival Dyads, *International Studies Quarterly*, No 37, 173-193.

Kemeny, J. and L. Snell. 1976. *Finite Markov Chains*. Van Nostrand.

Lund, M. 1996. *Preventing Violent Conflicts*. Institute of Peace Press.

MacDonald, Iain L. and Walter Zucchini. 1997. *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman and Hall.

Rabiner, Lawrence R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, No 77, 257-286.

Ross, D.S. 1978. Elementary Process Models: Their Place in International Relations Forecasting, in *Forecasting in International Relations: Theory, Methods, Problems, and Prospects*. Eds. N. Choucri and T.W. Robinson. W.H. Freeman.

Schrodt, P.A. 1976. Richardson's Model as a Markov Process, in *Mathematical Models in International Relations*. Eds. D. Zinnes and J. Gillespie. Praeger.

---. 1998. Pattern Recognition of International Crises Using Hidden Markov Models, in *Non-linear Models and Methods in Political Science*. Ed. D. Richards. University of Michigan Press.

---. 2000a. Early Warning of Conflict in Southern Lebanon using Hidden Markov Models, in *Political Complexity: Nonlinear Models of Politics*. Ed. D. Richards. University of Michigan Press.

---. 2000b. Forecasting Conflict in the Balkans using Hidden Markov Models. Presented at the American Political Science Association meetings, August 2000. Washington, DC.

Theodoridis, Sergios and Konstantinos Koutroumbas. 2003. *Pattern Recognition*. Academic Press.

Wilkenfield, Jonathan and Dina A. Zinnes. 1973. A Linkage Model of Domestic Conflict Behavior, in *Conflict Behavior and Linkage Politics*. Ed. Jonathan Wilkenfield. David McKay.