



となる。時刻 0 において内部状態を決定する初期ベクトルを  $\pi$  とすると、時刻  $t$  における期待累積感染報告数は、

$$E[R(t)] = \pi \int_0^t \exp\{(C + D)s\} ds De \quad (9)$$

で与えられる。ここで  $e$  は全ての要素が 1 である列ベクトルである。

### 3. パラメータ推定手法

前節で構築したモデルを用いてウィルスの増殖の予測をするためには、感染報告データに適合するようにモデルパラメータの推定を行う必要がある。

いま、 $K$  個のウィルス感染報告に関する時間間隔データ  $\mathcal{D} = \{x_1, x_2, \dots, x_K\}$  が与えられているもとで、パラメータの推定値を算出する手続きについて考える。特に、ここでは最尤推定値を算出する手続きとして EM アルゴリズム [8] を用いる。EM アルゴリズムは観測されていないデータ（不完全データ）に対して最尤推定値を求める手続きであり、期待対数尤度を算出する手続き（E ステップ）とそれを最大化する手続き（M ステップ）を持つ反復アルゴリズムである。

以下の記号を定義する。

$N_{ij}^{[u]}$ :  $u-1$  番目の感染報告から  $u$  番目の感染報告が発生するまでに内部状態が  $i$  から  $j$  へ推移する回数

$Y_{ij}^{[u]}$ : 内部状態  $i$  で  $u$  番目の感染報告が発生し、かつその直後に内部状態が  $j$  に推移する回数

$Z_{ij}^{[u]}$ :  $u-1$  番目の感染報告から  $u$  番目の感染報告が発生するまでに内部状態  $i$  に滞在する時間

感染報告データ  $\mathcal{D}$  が与えられたとき、それぞれの期待値  $E[N_{ij}^{[u]}|\mathcal{D}]$ ,  $E[Y_{ij}^{[u]}|\mathcal{D}]$ ,  $E[Z_{ij}^{[u]}|\mathcal{D}]$  を算出する手続きが E ステップに対応する。これは、文献 [9] にある一般的な MMPP に対する EM アルゴリズムにおける E ステップと同じであり、

$$E[N_{ij}^{[u]}|\mathcal{D}] = \frac{\mu_{i,j} c_j^u(i)}{\pi b_1}, \quad (10)$$

$$E[Y_{ij}^{[u]}|\mathcal{D}] = \frac{\xi_{i,j} [f_u]_i [b_{u+1}]_j}{\pi b_1}, \quad (11)$$

$$E[Z_{ij}^{[u]}|\mathcal{D}] = \frac{c_i^u(i)}{\pi b_1} \quad (12)$$

として算出される。ここで、 $\mu_{i,j}$  および  $\xi_{i,j}$  はそれぞれ行列  $C$  および  $D$  の  $(i,j)$  要素。また  $[\cdot]_i$  はベクトルの  $i$  要素を示す。さらに、これらの式において  $f_u, b_u$  は、

$$f_u = \pi \exp(Cx_1) D \cdots \exp(Cx_u), \quad (13)$$

$$b_u = \exp(Cx_u) D \cdots \exp(Cx_K) De, \quad (14)$$

であり、 $c_j^u(i)$  は

$$c_j^u(i) = \int_0^{x_u} [f_{u-1} D \exp(Cy)]_i \{\exp\{C(x_u - y)\} D b_{u+1}\}_j dy \quad (15)$$

で与えられる。

次に、E ステップで算出した期待値を用いて、M ステップは推定すべきパラメータ  $\beta, \delta, \phi$  に対して、

$$\hat{\beta} = \frac{\sum_{u=1}^K \sum_{i=1}^{N-1} E[N_{i,i+1}^{[u]}|\mathcal{D}]}{\sum_{u=1}^K \sum_{i=1}^N i(N-i) E[Z_i^{[u]}|\mathcal{D}]}, \quad (16)$$

$$\hat{\delta} = \frac{\sum_{u=1}^K \sum_{i=1}^N E[N_{i,i+1}^{[u]}|\mathcal{D}]}{\sum_{u=1}^K \sum_{i=1}^{N-1} i E[Z_i^{[u]}|\mathcal{D}]}, \quad (17)$$

$$\hat{\phi} = \frac{\sum_{u=1}^K \sum_{i=1}^N E[Y_{i,i}^{[u]}|\mathcal{D}]}{\sum_{u=1}^K \sum_{i=1}^N i E[Z_i^{[u]}|\mathcal{D}]} \quad (18)$$

となる。また、初期確率ベクトル  $\pi = (\pi_0, \pi_1, \dots, \pi_N)$  に対する M ステップにおける更新式は

$$\hat{\pi}_i = \frac{\pi_i [b_1]_i}{\pi b_1} \quad (19)$$

となる。

### 参考文献

- [1] Z. Chen, L. Gao and K. Kwiat, Modeling the spread of active worms, *Proceedings of IEEE INFOCOM 2003*, 2003.
- [2] S. Chen and Y. Tang, Slowing down Internet worms, *Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS 2004)*, pp. 312–319, 2004.
- [3] J. O. Kephart and S. R. White, Measuring and modeling computer virus prevalence, *Proceedings of the 1993 IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 2–15, 1993.
- [4] H. Toyozumi and A. Kara, Predators: good will codes combat against computer viruses, *ACM SIGSAC New Security Paradigms Workshop*, 2002.
- [5] W. H. Murray, The application of epidemiology to computer viruses, *Computers and Security*, Vol. 7, No. 2, pp. 139–145, 1988.
- [6] J. C. Wierman and D. J. Marchette, Modeling computer virus prevalence with a susceptible-infected-susceptible model with reintroduction, *Computational Statistics & Data Analysis*, Vol. 45, pp. 3–23, 2004.
- [7] 小林, 岡村, 土肥, 確率モデルによるコンピュータウィルスの特徴分析, *情報処理学会論文誌*, Vol 45, No. 5, pp. 1432–1441, 2004.
- [8] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, NY, 1997.
- [9] T. Rydén, An EM algorithm for estimation in Markov-modulated Poisson process, *Computational Statistics & Data Analysis*, Vol. 21, pp. 431–447, 1996.