Predicting antigenic changes of influenza viruses through data assimilation

Ito, Kimihito
北海道大学人獣共通感染症リサーチセンター

出版情報：MI lecture note series. 60, pp.80-84, 2014-11-28. 九州大学マス・フォア・インダストリ研究所
バージョン：
権利関係：

Predicting antigenic changes of influenza viruses through data assimilation

北海道大学　　伊藤公人　　Kimihito Ito

Abstract

Human influenza A viruses undergo antigenic changes with gradual accumulation of amino acid substitutions on the hemagglutinin molecule. Antigenic mismatch between vaccine and epidemic strains often requires the replacement of influenza vaccine strains. To establish a practical method enabling us to predict the future direction of the viral evolution, we are developing a new prediction method based on data assimilation. The timing when the dominant epidemic strains were replaced by other strains and the magnitude of outbreaks of seasonal influenza could be predicted by the method.

Introduction

The hemagglutinin (HA) molecule of influenza A viruses is the prime target of antibodies that neutralize viral infectivity. The strong immune pressure against HA in the human population selects a new variant every 2-5 years. Thus, influenza A viruses undergo antigenic changes with gradual accumulation of amino acid substitutions on HA. This antigenic change is one of the primary reasons why vaccination is not a perfect measure to control seasonal influenza. Influenza vaccine often requires replacement to avoid antigenic mismatch between vaccine and epidemic strains. The decision of vaccine replacement must be made several months before a minor strain become dominant strain. Therefore, the prediction of antigenic change of influenza A virus has been one of the major public health goals [1].

Antigenic Variation of Influenza A Viruses

Every year, the World Health Organization (WHO) makes a recommendation on the vaccine strains of influenza viruses, based-on the antigenic characteristic of influenza virus isolates collected by its reference laboratories from around the world [2-16]. Current trivalent inactivated vaccine contain antigens of H1N1 and H3N2 influenza A viruses and influenza B viruses. Table 1 shows vaccine strains of H3N2 virus recommended by WHO and dominant virus strains found in the north hemisphere. During the period from 1997 and 2011, eight out of fourteen influenza seasons had mismatches between vaccine strains and dominantly-circulating strains. As can be seen in Table 1, WHO have recommended a dominant strain of the previous season as a vaccine strain of the next season. Therefore, the vaccine replacement has been usually delayed one to two years. To improve the efficacy of influenza vaccines, prediction methods for the antigenic change of influenza A virus has been required.

Data Assimilation

Data assimilation is a statistical method by which actual observations are integrated into computer simulations. Ensemble Kalman filters [17, 18, 19] and particle filters [20] are

Table 1. Vaccine strains of H3N2 virus recommended by WHO and dominant virus strains.

Season	WHO Recommendation	Reported Dominant Strain
1997-1998	A/Wuhan/359/95-like	A/Sydney/5/97-like
1998-1999	A/Sydney/5/97-like	A/Sydney/5/97-like
1999-2000	A/Sydney/5/97-like	A/Moscow/10/99-like
2000-2001	A/Moscow/10/99-like	A/Moscow/10/99-like
2001-2002	A/Moscow/10/99-like	A/Moscow/10/99-like
2002-2003	A/Moscow/10/99-like	A/Moscow/10/99-like, A/Fujian/411/2002-like
2003-2004	A/Moscow/10/99-like	A/Fujian/411/2002-like
2004-2005	A/Fujian/411/2002-like	A/California/7/2004-like
2005-2006	A/California/7/2004-like	A/Wisconsin/67/2005-like
2006-2007	A/Wisconsin/67/2005-like	A/Wisconsin/67/2005-like
2007-2008	A/Wisconsin/67/2005-like	A/Brisbane/10/2007-like
2008-2009	A/Brisbane/10/2007-like	A/Brisbane/10/2007-like
2009-2010	A/Brisbane/10/2007-like	A/Perth/16/2009-like
2010-2011	A/Perth/16/2009-like	A/Perth/16/2009-like

innovative examples of data assimilation techniques. Sophisticated mathematical models and massively parallel computational resource are prerequisite for successful data assimilation. So far, data assimilation has been applied in meteorology, oceanography, engineering, and life science. Significant improvements in prediction accuracy have been reported, especially in weather forecast and hydrology [21].

Data assimilation is well represented by a system space model. For each time step t, let x_t be a vector of random variables of a system state, and let y_t be a vector of random variables of observations from the system. A system model can be represented by $x_t=f(x_{t-1},v_t)$, where v_t is a process noise. Process noise is attributed from errors in the simulation model, and the v_t can absorb errors in the simulation model. An observation model can be represented by $y_t=h(x_t,w_t)$, where w_t is a measurement noise. Given a set of observations from time 1 to t, data assimilation infers the posterior distribution of both current system state $P(x_t \mid y_{1:t})$ and the one-step-ahead system state $P(x_{t+1} \mid y_{1:t})$. Figure 1 shows sequential Monte Carlo algorithm called particle filter, which is a data assimilation method proposed by Kitagawa [20].

1. For each j=1,…, N generate a state vector $x^{(j)}_0 \sim P(x_0)$
2. For each i=1,…, t do
3. For each j=1,…, N do
4. Generate a process noise $v^{(j)}_i \sim Q$
5. Calculate the next sate $x^{(j)}_i = f(x^{(j)}_{i-1}, v^{(j)}_i)$
6. Calculate the likelihood $\lambda^{(j)}_i = P(y_i \mid x^{(j)}_i)$
7. Resample $x^{(j)}_i$ with replacement according to their likelihood $\lambda^{(j)}_i$

Figure 1. The algorithm of particle filter

State Space Model for Predicting Antigenic Changes of Influenza Viruses

To establish a practical method enabling us to predict amino acid substitutions on the hemagglutinin molecule of influenza viruses, we consider a state space model of viral population, infections and herd immunity. Our aim here is to integrate actual observations of viral gene mutation into computer simulations, and infer current herd immunity and next mutations.

Let V_t be a matrix representing the proportion of amino acids in the HA of epidemic strains at time t. The globular head reason of HA proteins of A Hong Kong H3N2 viruses consists of 328 amino acids. Each amino acid position can potentially have any of 20 amino acids. From this numbers, V_t is a 328×20 matrix, and the element in the i-th row and j-th column of V_t represents the frequency of the i-th amino acids at the j-th position of HA proteins in the viral population. Let I_t be the number of infections at time t. Let H_t be a matrix representing the proportion of people having immunity against a particular HA of the virus. Then, H_t is a 328×20 matrix, and the element in the i-th row and j-th column of V_t represents the frequency of people who has immunity to the virus having the i-th amino acids at the j-th position of HA proteins.

Here we consider the dynamics among V_t , I_t and H_t. Assuming the number of viral strains at time t is equal to I_t, the multiplication of V_t and I_t represents the number of strains having a particular amino acid at a particular position. The increase of the herd immunity against a particular amino acid at a particular position would be correlated with the number of infections and viral population strains at time t. In other words, the increase of an element at (i, j) in H_t would be represented by a function of I_t and the element at (i, j) in V_t. The increase in the number of infections is attributed to the difference between amino acid distribution in herd

immunity against HA and that of currently circulating viral population. The increase of I_t would be represented by a function of I_t, V_t and H_t. The dynamics of V_t involves the evolutionary dynamics of viral fitness. A straightforward way of modeling would be an extension of the quasi-species equation formulated by Manfred Eigen and Peter Schuster [22].

Experiments

We have constructed a mathematical model of viral population, infection, and host immunity. Based on the developed model, actual viral evolution observed in past 42 years was analyzed by particle filters.

Nucleotide sequences for HA genes of H3N2 influenza A viruses isolated from humans were downloaded from the Influenza Virus Resource at the National Center for Biotechnology Information (NCBI). After eliminating sequences that contained ambiguous nucleotide codes, amino acid sequences of the HA1 domain were determined by translating the nucleotide sequences. All the amino acid sequences were 328 amino acids long.

Considering the amino acid sequences from NCBI as observations, computer experiments of data assimilation was performed and posterior probability of I_t, V_t and H_t were inferred. Currently available results showed the data assimilation could have better potential to predict future amino acid substitutions on HA than our previous method [23]. In this presentation, I will introduce the newly developed data-assimilation-based system for the prediction of influenza virus evolution.

References

[1] Laver, W.G., Air, G.M. and Webster, R.G. (1981) Mechanism of antigenic drift in influenza virus : Amino acid sequence changes in an antigenically active region of Hong-Kong (H3N2) influenza virus hemagglutinin, *J Mol Biol*, 145, 339-361.
[2] WHO. (1997). Recommended composition of influenza virus vaccines for use in the 1997-1998 season. *Wkly Epidemiol Rec*, *72*(9), 57–61.
[3] WHO. (1998). Recommended composition of influenza virus vaccines for use in the 1998-1999 season. *Wkly Epidemiol Rec*, *73*(9), 56–61.
[4] WHO. (1999). Recommended composition of influenza virus vaccines for use in the 1999-2000 season. *Wkly Epidemiol Rec*, *74*(8), 57–61.
[5] WHO. (2000). Recommended composition of influenza virus vaccines for use in the 2000-2001 season. *Wkly Epidemiol Rec*, *75*(8), 61–68.
[6] WHO. (2001). Recommended composition of influenza virus vaccines for use in the 2001-2002 season. *Wkly Epidemiol Rec*, *76*(8), 58–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11236648
[7] WHO. (2002). Recommended composition of influenza virus vaccines for use in the 2002-2003 season. *Wkly Epidemiol Rec*, *77*(8), 62–6.

[8] WHO. (2003). Recommended composition of influenza virus vaccines for use in the 2003-2004 influenza season. *Wkly Epidemiol Rec*, *78*(9), 58–62.
[9] WHO. (2004). Recommended composition of influenza virus vaccines for use in the 2004-2005 influenza season. *Wkly Epidemiol Rec*, *79*(9), 88–92.
[10] WHO. (2005). Recommended composition of influenza virus vaccines for use in the 2005-2006 influenza season. *Wkly Epidemiol Rec*, *80*(8), 65–76.
[11] WHO. (2006). Recommended composition of influenza virus vaccines for use in the 2006-2007 influenza season. *Wkly Epidemiol Rec*, *81*(9), 82–86.
[12] WHO. (2007). Recommended composition of influenza virus vaccines for use in the 2007-2008 influenza season. *Wkly Epidemiol Rec*, *82*(9), 69–74.
[13] WHO. (2008). Recommended composition of infl uenza virus vaccines for use in the 2008–2009 infl uenza season. *Wkly Epidemiol Rec*, *83*(9), 77–88.
[14] WHO. (2009). Recommended composition of infl uenza virus vaccines for use in the 2009–2010 infl uenza season (northern hemisphere winter). *Wkly Epidemiol Rec*, *84*(9), 65–76.
[15] WHO. (2010). Recommended viruses for influenza vaccines for use in the 2010-2011 northern hemisphere influenza season. *Wkly Epidemiol Rec*, *85*(10), 81–92.
[16] WHO. (2011). Recommended composition of influenza virus vaccines for use in the 2011-2012 northern hemisphere influenza season. *Wkly Epidemiol Rec*, *86*(10), 86–90.
[17] Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10143.
[18] Service, A. E. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique, (1969), 796–812.
[19] Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D: Nonlinear Phenomena, 230(1-2), 112–126.
[20] Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. Journal of Computational and Graphical Statistics, 5(1), 1.
[21] J.L. Anderson (2001) An ensemble adjustment Kalman filter for data assimilation, *Mon Weather Rev*, 129, 2884-2093.
[22] Eigen, M., McCaskill, J., Schuster, P (1989) *The Molecular Quasi-species*. Advances in Chemical Physics 75: 149–263.
[23] Ito, K et.al (2011) Gnarled-Trunk Evolutionary Model of Influenza A Virus Hemagglutinin, *PLoS ONE*, 6:10, e25953.