

論文

J. of The Korean Society for Aeronautical and Space Sciences 43(1), 23-32(2015)

DOI:http://dx.doi.org/10.5139/JKSAS.2015.43.1.23

ISSN 1225-1348(print), 2287-6871(online)

dPCA-HMM을 이용한 전투기 조종사 모델링 연구

최예림*, 전승욱*, 박종현*, 신동민**

A Study on Modeling of Fighter Pilots Using a dPCA-HMM

Yerim Choi*, Sungwook Jeon*, Jonghun Park* and Dongmin Shin**

Department of Industrial Engineering, Seoul National University*

Department of Industrial and Management Engineering, Hanyang University**

ABSTRACT

Modeling of fighter pilots, which is a fundamental technology for war games using defense M&S (Modeling & Simulation) becomes one of the prominent research issues as the importance of defense M&S increases. Especially, the recent accumulation of combat logs makes it possible to adopt statistical learning methods to pilot modeling, and an HMM (Hidden Markov Model) which is able to utilize the sequential characteristic of combat logs is suitable for the modeling. However, since an HMM works only by using one type of features, discrete or continuous, to apply an HMM to heterogeneous features, type integration is required. Therefore, we propose a dPCA-HMM method, where dPCA (Discrete Principal Component Analysis) is combined with an HMM for the type integration. From experiments conducted on combat logs acquired from a simulator furnished by agency for defense development, the performance of the proposed model is evaluated and was satisfactory.

초 록

전투기 조종사 모델링은 국방 M&S(Modeling & Simulation)를 활용한 전쟁 모의 및 전투 실험의 기초 기술로 국방 M&S의 중요성이 대두됨에 따라 연구의 필요성이 높아지고 있다. 특히, 최근 전투 로그의 축적으로 통계적 학습 기법을 활용한 모델링의 적용이 가능해졌으며 전투 로그의 시계열적 특성을 반영할 수 있는 HMM(Hidden Markov Model)이 적합하다. 하지만 HMM은 이산형 혹은 연속형 중 한 형태의 변수만을 통해 학습되므로 이형 변수로 구성된 전투 로그에 적용을 위해서는 형변환 과정이 필요하다. 따라서 본 논문에서는 형변환을 위한 dPCA(Discrete Principal Component Analysis)와 HMM을 접목한 dPCA-HMM 기반 조종사 모델링 방법을 제안한다. 국방과학연구소 관급 시뮬레이터로부터 생성된 전투 로그를 이용한 비교 실험을 통해 제안하는 방법론의 성능을 평가하였으며, 만족스러운 성능을 나타내었다.

Key Words : Fighter Pilot Modeling(전투기 조종사 모델링), Combat Logs(전투 로그), Hidden Markov Model(HMM), Discrete Principal Component Analysis (dPCA)

† Received : April 20, 2014 Revised : November 3, 2014 Accepted : December 23, 2014

** Corresponding author, E-mail : dmshin@hanyang.ac.kr

1. 서론

국방 M&S(Modeling & Simulation)를 활용한 전쟁 모의나 전투 실험은 실제 수행이 불가능한 상황을 모의해 볼 수 있으며 비용을 최소화 할 수 있고 반복 수행을 통해 최적의 대안을 미리 예측 할 수 있다는 점에서 각광받고 있다[1]. 전투기 조종사 모델링은 국방 M&S의 구현을 위한 기초 기술 중 하나로, 실제 전투를 수행하는 개체의 행위 및 판단을 모사할 수 있는 모델을 만드는 것이다. 기존에는 개발 용이성으로 인해 퍼지회귀와 같은 규칙 기반 방법론이 많이 사용되었지만[2] 이러한 방법론은 규칙 간 충돌이 발생할 수 있고 사전에 규칙이 정의되어야 한다는 점에서 한계를 가진다.

최근 군사 기술의 발달로 실제 전투 및 전투 훈련에서의 상황과 행위가 기록된 로그의 축적이 이루어지고 있다. 이러한 데이터를 이용하여 통계적 학습 기법 기반 조종사 모델링이 가능할 것으로 기대된다. 통계적 학습 기법을 이용하면 주어진 전투 로그 내의 조종사의 상황에 따라 취하는 행동의 패턴을 학습하여 기존 로그와 유사하게 행동하는 모델을 만들 수 있다. 이를 이용하면 이미 발생하거나 알려진 상황 뿐 아니라 새로운 상황에 대해서도 조종사의 행위 예측이 가능해진다. 따라서 통계적 학습 기법을 이용한 조종사 모델링을 통해 기존 규칙 기반 방법론의 문제점을 해결 할 수 있을 것이다.

본 연구에서는 다양한 종류의 전투 로그 중 시간에 따른 교전 상황과 해당 시간의 전투기 조종사의 행위를 기록한 데이터를 대상으로 한다. 이때 교전 상황을 나타내는 변수들을 상태 변수, 해당 상황에서 조종사의 행위를 나타내는 변수를 행위 변수라고 명명한다. 전투 로그는 시간에 따라 발생한다는 점에서 시계열 데이터의 특성을 가지며, 상태 변수는 전투기의 속도 및 고도와 같이 실수값을 갖는 연속형 변수와 적기의 유무와 같이 이산값을 갖는 이산형 변수가 혼재되어 있다는 특징을 가진다. Fig. 1은 이산형 및 연속형 변수의 시간의 따른 변화 양상을 나타낸다. 실선으로 표시된 그래프는 적기 감지 여부를 나타내는 상태 변수로 0 또는 1의 이산값을 갖고, 점선으로 표시된 그래프는 정규화된 전투기의 속력으로 0부터 1사이의 실수값을 갖는다.

이와 같은 특성을 갖는 전투 로그를 기반으로 조종사를 모델링하기 위해 시계열 데이터를 확률적으로 모델링하는 대표적인 기법으로 시계열 데이터 처리에서 좋은 성능을 보이는 것으로 알려

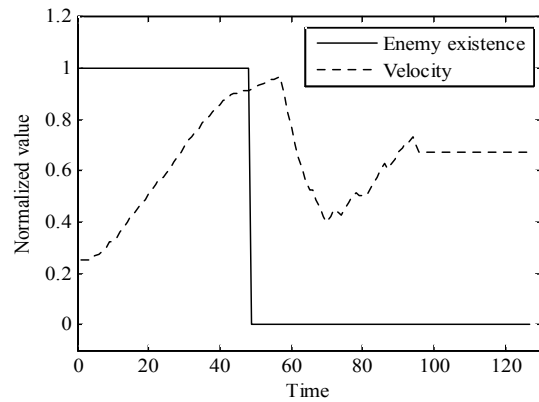


Fig. 1. Examples of discrete and continuous status variables of an aircraft

진[3] HMM(Hidden Markov Model)[4]을 사용한다. HMM은 이산형 혹은 연속형 변수를 기반으로 모형을 생성하는지에 따라 이산 HMM과 연속 HMM으로 나뉘어진다. 즉, 이산형 및 연속형 변수가 섞여있는 전투 로그의 경우 하나의 형태로 데이터를 통합한 후 HMM을 도입해야 한다. 연속형 변수를 이산화 한 후 이산 HMM을 사용하는 경우, 변환 과정에서 정보의 손실이 발생하여 성능이 저하되고, 이를 방지하기 위해 연속형 변수를 세밀한 범위로 나누어 이산화하면 모델에서 학습해야 하는 파라미터의 수가 증가하여 성능이 하락한다고 알려져 있다[5].

기존 HMM을 이용한 조종사 모델링 연구를 살펴보면 Table 1과 같다. HMM을 이용한 조종사 모델링 연구는 그 목적에 따라 조종사의 집중 상태를 예측하고자 하는 연구[6,7], 조종사의 기동을 모델링하는 연구[8,9], 조종사의 비행 목표 혹은 목적을 예측하는 모델[10,11] 등이 있었다. 조종사의 집중도를 예측하는 경우 눈의 위치를 추적한 데이터나 뇌파를 측정된 데이터를 사용하였는데, 눈의 위치 데이터의 경우 위치를 이산화하여 이산 HMM을 학습하였고[6], 뇌파의 경우 연속형 값을 그대로 연속 HMM을 학습하는데 사용하였다[7]. 조종사의 기동을 모델링하는 경우 연속형 값을 갖는 항공기의 움직임 데이터를 사용하므로 연속 HMM을 사용하였지만, 조종사의 목표를 예측하는 경우 이산형 및 연속형 데이터가 혼재되어 있어 이들을 이산화 한 후 이산 HMM을 이용하였다. 이와 같은 방식을 택할 경우 앞에서 언급한 바와 같이 이산화 과정에서 정보의 손실이 발생하여 모델의 성능이 하락한다는 단점이 있다.

HMM 뿐 아니라 대부분의 통계적 학습 기법을 사용하기 위해서는 변수 형태의 통합이 필요

하다[12]. 비교적 직관적이고 이해하기 쉬운 연속형 변수의 이산화 기법들이 많이 연구되었지만 [13,14], 이산형 변수의 연속화의 경우 연구가 미미한 상태이다. 이산형 변수 중 값 자체가 의미를 가지는 서수(ordinal number)는 연속형 변수 처럼 처리 될 수 있지만 특정한 항목을 나타내는 기명수(nominal number)는 값을 그대로 사용할 수 없다. 일반적으로, 기명수 값을 갖는 변수의 변환은 변수가 가질 수 있는 항목의 개수만큼 더미 변수 생성하고 각 항목에 속하는지 여부에 따라 더미 변수에 이진수 값을 부여하는 방식이 사용된다. 하지만 이는 값의 왜곡이 발생할 수 있다는 점[15], 과도한 차원 확장을 가져온다는 점에서 한계를 가진다. 이러한 문제점을 해결하기 위해 [16]에서는 이산형 변수들을 더미 변수를 이용하여 이진수로 표현한 후 차원 축소 기법을 도입하여 다시 의미 있는 적은 수의 변수로 표현하는 dPCA(discrete principal component analysis)방식이 사용되었다.

따라서 본 논문에서는 변수들을 연속형으로 통합한 후 이를 이용하여 전투기 조종사를 모델링하는 dPCA-HMM 기법을 제안한다. 제안하는 기법은 두 단계로 이루어지는데, 첫 번째 단계에서는 이산형 변수들의 연속화를 위해 dPCA가 사용된다. 구체적으로, 기명수 값을 갖는 변수들을 이진수로 표현한 뒤 PCA 기법을 이용하여 차원이 축소된 연속형 변수로 변환한다. 두 번째 단계에서는 연속화를 통해 변환된 변수들을 포함한 전체 변수들을 대상으로 연속 HMM을 학습시켜 조종사를 모델링한다.

본 논문은 다음과 같이 구성된다. 2장에서는 본 연구에서 제안하는 데이터 통합 기법과 전투기 조종사 모델링 기법에 대해 자세히 설명하고,

3장에서는 제안된 방법론의 성능 평가를 위한 실험 설정과 실험 결과를 보여준다. 마지막으로 4장에서는 본 연구를 요약하고 앞으로의 연구 방향을 제시한다.

II. 본 론

2.1 문제 정의

본 연구에서는 전투 로그를 학습하여 조종사의 행위를 모사하는 모델링 기법을 제안한다. 즉, 시간 t 에 따른 상태 변수들과 행위 변수를 기록한 전투 로그 $\{(\mathbf{x}_t, y_t) | t = 1, \dots, T\}$ 를 학습하여 새로운 시점 t' 에서의 상태 변수 벡터 $\mathbf{x}_{t'}$ 가 주어졌을 때 행위 변수 $y_{t'}$ 를 예측하는 모델을 개발한다. 여기서 T 는 전투의 종료 시점을 의미하며 상태 변수 벡터 \mathbf{x}_t 는 시점 t 의 상태 변수들의 값으로 구성된 벡터이다. y_t 는 행위 변수 집합 $B = \{b_k | k = 1, \dots, n_b\}$ 의 원소 중 시점 t 에 발생한 하나의 값이다. n_b 는 총 가능한 행위 변수의 수를 나타낸다.

상태 변수들은 가질 수 있는 값의 속성에 따라 이산형과 연속형으로 나누어진다. 구체적으로, i 번째 상태 변수에 대해 해당 상태 변수가 가질 수 있는 값의 집합, $E_i = \{e_{i,j} | j = 1, \dots, n_i\}$ 가 정의되고 $e_{i,j}$ 는 i 번째 상태 변수가 가질 수 있는 값 중 하나를, n_i 는 이러한 값의 총 개수를 의미한다. 즉, 이산형 상태 변수는 n_i 가 유한한 상태 변수를 의미하고, 연속형 상태 변수는 n_i 가 무한한 상태 변수를 의미한다. 각각 이산형 상태 변수의 집합을 $\{d_i | i = 1, \dots, n_d\}$ 로, 연속형 상태 변수의 집

Table 1. Summary of pilot modeling studies using HMM according to their purposes, methods, and data types

References	Purpose	Method	Data type
Hayasi, 2003	Attention detection	HMM	Discretized position of eye movement
Andersson and Pettersson, 2004	Aerial-mission recognition	HMM	Discretized speed, altitude, and maneuver activity
Mori and Suzuki, 2010	Landing approach control model	Combination of linear regression model and HMM	Continuous flight sensor data
Tervo, 2010	Intent recognition	HMM	Discretized flight data
Lowe, 2014	Navigation model	HMM with a modified likelihood function	Continuous trajectory data
Choi et al., 2014	Hypo-vigilance detection	HMM	Continuous brain signal data

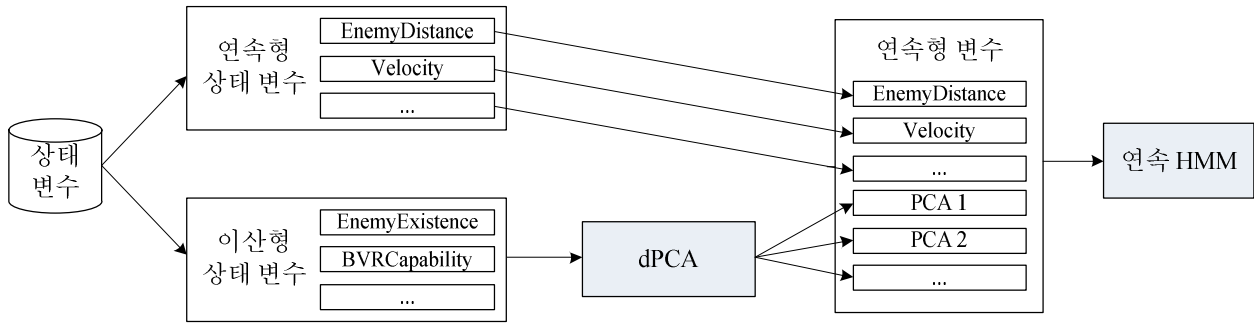


Fig. 2. Overview of the proposed dPCA-HMM method

합을 $\{c_j | j=1, \dots, n_c\}$ 로 나타낸다. n_d 와 n_c 는 각각 이산형 및 연속형 상태 변수의 총 개수를 나타낸다. 따라서 \mathbf{x}_t 는 이산형 및 연속형 변수로 구분된 두 서브벡터 $\mathbf{x}_t^d = \langle x_{1,t}^d, x_{2,t}^d, \dots, x_{n_d,t}^d \rangle$ 와 $\mathbf{x}_t^c = \langle x_{1,t}^c, x_{2,t}^c, \dots, x_{n_c,t}^c \rangle$ 로 나누어 질 수 있다.

2.2 dPCA-HMM 기반 조종사 모델링

제안하는 dPCA-HMM 기반 조종사 모델링은 Fig. 2에 나타난 절차로 진행된다. 우선, 이산형 상태 변수와 연속형 상태 변수를 분리 한 후 이산형 상태 변수를 dPCA 기법을 이용하여 연속형으로 변환한다. 이렇게 변환된 연속형 변수를 이용하여 시계열 데이터 모델링에 적합하다고 알려진 연속 HMM을 학습시킨다. 이때, 정보 손실을 최소화하고 과도한 파라미터 추정 과정을 방지하기 위해 이산형 변수들의 연속화를 통해 연속형으로 통합된 변수들을 연속 HMM을 이용하여 모델링한다. 아래에서는 각 단계 별 구체적인 방법론을 설명한다.

2.2.1 dPCA를 이용한 변수 형태 통합

dPCA 기법은 Filmer & Pritchett PCA라고도 불리며 이산형 변수에 PCA를 적용하기 위해 [16]에서 사용되었다. PCA는 주어진 다차원 데이터의 차원을 축소하기 위한 사용되는 방법론으로 변수 값들의 분산이 최대화되도록 기존 데이터를 새로운 변수로 선형 변환하는 방법론이다[17]. 고차원 데이터의 차원을 줄여 계산 복잡도를 낮출 수 있고 노이즈를 감소시켜 준다는 장점이 있어 항공기의 고장진단[18]과 같이 다양한 분야에서 사용되었다. 기본적으로 연속형 변수를 가정하므로 기명수 값을 갖는 이산형 변수에 적용하기 위해서는 해당 변수들을 더미 변수를 이용하여 이진수로 변환한 후 PCA를 도입하는 dPCA 방식이 사용된다[16]. 따라서 본 연구에서는 dPCA 도입하여 이산형 변수를 연속형으로 변환한다.

우선, 이산형 변수들을 더미 변수를 활용하여

이진법으로 나타낸다. 각 이산형 상태 변수 d_i 를 나타내기 위해 총 n_i 개의 더미 변수 $d_{i(1)}, \dots, d_{i(j)}, \dots, d_{i(n_i)}$ 들이 생성되고 t 시점의 이산형 변수들의 값으로 구성된 상태 변수 벡터 \mathbf{x}_t^d 가 주어지면 식 (1)과 같이 상태 변수 d_i 의 t 시점에서의 j 번째 더미 변수의 값 $x_{i(j),t}^d$ 이 부여된다.

$$x_{i(j),t}^d = \begin{cases} 1 & \text{if } x_{i,t}^d = e_{i,j} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

이와 같은 방식으로 총 n_d 의 원소로 구성된 벡터 \mathbf{x}_t^d 가 0 또는 1의 값만을 원소로 갖는 $\sum_i n_i$ 차원의 이진 벡터 \mathbf{x}_t^{db} 로 변환된다.

이렇게 생성된 벡터는 이산형 변수가 가질 수 있는 값의 개수만큼 차원을 확장시킨다는 점에서 이대로 사용할 경우 계산 복잡도가 증가하고, 실질적인 변수 값, 0 또는 1의 의미가 속하는 변수를 나타낼 뿐 행위 변수와의 관련성이 낮다는 문제가 있다. 따라서 \mathbf{x}_t^{db} 에 대해 추가적으로 PCA를 수행하여 n_p 개의 주요 변수로 구성된 벡터 $\mathbf{x}_t^p = \langle x_{1,t}^p, x_{2,t}^p, \dots, x_{n_p,t}^p \rangle$ 로 변환한다. 최종적으로 \mathbf{x}_t^c 와 \mathbf{x}_t^p 을 합쳐 연속형 값으로만 구성된 벡터 $\tilde{\mathbf{x}}_t$ 를 구한다.

예를 들어, 이산형 상태 변수 중 하나인 EnemyExistence는 적군의 존재 여부를 나타낸다. 적군이 존재할 때는 1의 값을 존재하지 않는 경우는 2의 값을 갖는데, 이 상태 변수를 연속형 상태 변수와 동일하게 할 경우 2의 값이 1보다 크다는 성질이 반영되어 정보가 왜곡될 수 있다. 따라서 더미 변수 두 개를 이용하여 {1,2}로 나타내지던 값을 {(0,1), (1,0)}로 표현하여 수의 크기가 의미를 갖지 않도록 만든다. 해당 변수의 경우 가질 수 있는 값의 수가 두 개로 더미 변수로 표현하였을 때 문제가 발생하지 않지만, WeaponSelection과 같이 가질 수 있는 값의 수

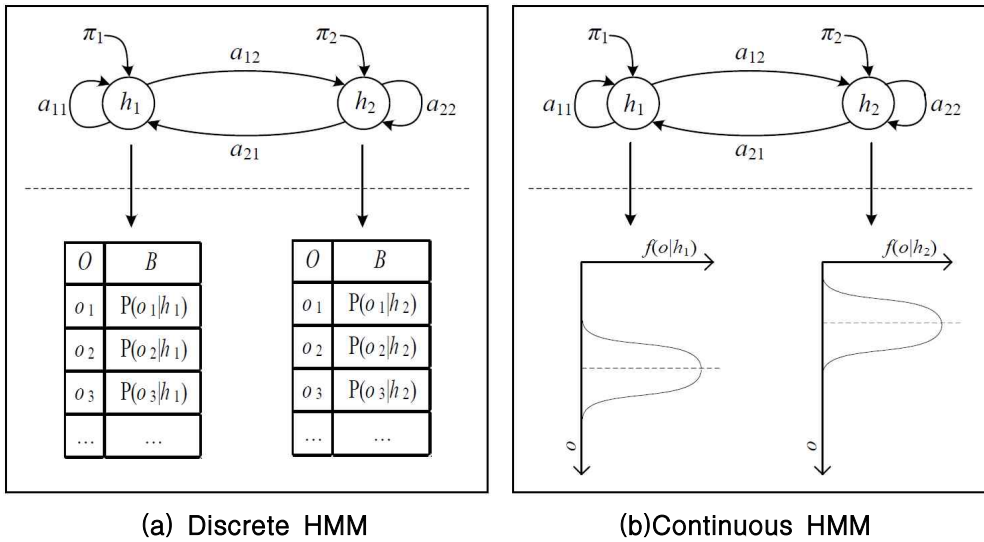


Fig. 3. Examples of discrete and continuous HMMs with two hidden states

가 4개 이상일 경우에는 필요한 더미 변수의 수가 많아져 변수의 차원이 과도하게 확장된다. 따라서 차원 축소 기법인 PCA를 적용하여 유의미한 변수로 압축함으로써 과도한 차원에 의해 발생하는 문제를 없애고, 성능을 향상시킬 수 있다.

2.2.2 연속 HMM을 이용한 행위 모델

모든 상태 변수가 연속형 변수로 통합된 후에는 연속 HMM을 이용하여 전투 로그를 모델링한다. HMM은 특정 관측의 발생 확률이 은닉상태에 따라 다르게 주어지고, 은닉상태 간 전이확률은 이전 은닉상태에만 영향을 받는다는 마코프 성질을 가정하는 모델이다[4]. 특히, HMM은 관측을 구성하는 값이 이산형 혹은 연속형인지에 따라 이산 HMM과 연속 HMM으로 나누어진다. Fig. 3은 각각 두 개의 은닉상태를 갖는 이산 HMM과 연속 HMM의 예시를 나타낸다.

본 연구에서는 연속 HMM을 사용하여 $t=1, \dots, T$ 에 대해 주어진 연속형 변수 벡터 $\tilde{\mathbf{x}}_t$ 와 이에 대응하는 행위 변수 y_t 로 구성된 전투로그를 학습한다. 구체적으로, 연속 HMM은 은닉상태집합 H , 관측집합 O , 전이확률행렬 A , 관측확률함수 f , 초기확률벡터 Π 의 다섯 요소로 구성되어 $\Lambda=(H, O, A, f, \Pi)$ 로 표기된다. 이때, 관측은 실제로 관찰되는 값이며 은닉상태는 관측이 불가능하지만 그 개수만 주어지면 데이터를 통해 전이확률행렬 A 가 학습된다.

A 는 특정 은닉상태에서 자신을 포함한 각 은닉상태로 변화할 확률로 구성된 행렬로 원소 a_{ij} 는 $i, j=1, \dots, n_h$ 에 대해 은닉상태 h_i 에서 h_j 로 변화할 확률을 나타내며 식 (2)와 같이 계산된다.

$$a_{ij} = P(q_{t+1} = h_j | q_t = h_i) \quad (2)$$

이때, q_t 는 t 시점에서의 은닉상태이고 n_h 는 총 은닉상태의 개수이다. t 는 은닉상태 h_i 에서 t 시점에서의 연속 변수 벡터 $\tilde{\mathbf{x}}_t$ 의 l 시점 이전부터 관측된 시퀀스 행렬 \tilde{X}_t 가 발생할 확률로 $f(\tilde{X}_t | q_t = h_i)$ 와 같이 정의된다. 제안하는 모델에서 사용하는 연속 HMM은 관측을 구성하는 값이 연속형인 HMM으로 관측 값들이 다변수 정규분포를 따른다고 가정한다. 마지막으로, Π 는 각 은닉상태가 초기상태일 확률로 원소 π_i 는 $i=1, \dots, n_h$ 에 대해 식 (3)과 같이 계산된다.

$$\pi_i = P(q_1 = s_i) \quad (3)$$

이렇게 정의된 HMM의 전이확률행렬 A , 관측 확률함수 f 가 나타내는 분포의 평균과 공분산, 그리고 초기확률벡터 Π 는 EM(Expectation and Maximization) 알고리즘을 이용하여 학습데이터로부터 추정된다[19]. 특히, 조종사의 모델링을 위해서 행위 변수 집합 B 의 원소 별로 HMM을 학습하고 주어진 연속형 변수 벡터 $\tilde{\mathbf{x}}_t$ 의 l 시점 이전부터의 시퀀스 행렬 \tilde{X}_t 가 발생할 확률이 가장 높은 HMM 선택하여 해당 HMM이 나타내는 행위 변수를 $\tilde{\mathbf{x}}_t$ 에 대응하는 y_t 로 도출한다.

III. 실험

3.1 전투 로그

제안한 전투기 조종사 모델링 기법의 성능을

평가하기 위해서는 전투기 조종사가 전투 및 훈련에서 생성한 로그를 사용해야 하지만, 실제 전투 및 훈련에서 발생한 데이터는 보안 문제가 존재하므로 시뮬레이터로부터 얻어진 전투 로그로 대체하였다. 구체적으로, 국방과학연구소 관급 공대공 전투 시뮬레이터로부터 실제 전장과 유사한 성질을 갖는 전투 로그를 추출하여 사용하였다. 해당 시뮬레이터의 경우 시뮬레이션 객체가 규칙 기반방법론을 이용하여 구축되어 있어 주어진 시나리오에 대해 확정적인(deterministic) 행위를 수행한다. 제안하는 모형이 주어진 로그를 성공적으로 모사한다면, 실제 전투 로그를 이용한 경우에도 동일한 수준의 성능을 나타낼 것으로 기대한다.

해당 시뮬레이터는 사용자가 객체 및 상황을 시나리오로 표현하여 정의된 전장 상황을 재현할 수 있는 기능을 제공한다. 따라서 다양한 상황을 묘사하는 로그를 생성하여 제안하는 모형을 학습시키기 위해 총 120 종류의 시나리오를 작성 및 수행하여 약 0.5GB의 로그를 추출하였다. 각 시나리오는 야군과 적기의 전투기 기종, 방어무장의 종류 및 탑재 여부, 무장의 종류 및 탑재 여부, 전투의 시작 위치 등에 따라 달라지며, 본 연구에서는 공대공, 일대일 전투로 상황으로 한정하여 수행하였다. Fig. 4는 수행한 시나리오 중 하나의 삼차원 화면으로 이를 통해 전투기의 움직임 및 궤적을 확인할 수 있다.

수집된 전투 로그는 전처리 과정을 거쳐 모델에 사용될 수 있는 시간에 따른 상태 변수와 행위 변수로 변환되었다. 추출된 상태 변수는 총 45개로 Table 2는 상태 변수의 예시를 보여준다. 상태 변수는 14개의 이산형 변수와 31개의 연속형 변수로 구성되며, 사용된 이산형 변수 모두 기명수 값을 갖는 것으로 연속화 과정을 필요로 한다. 행위 변수는 약 30가지가 추출되었으며 이중 유사한 의미를 갖는 변수를 집단화하여 최종적으로는 여섯 종류의 행위 변수를 대상으로 조

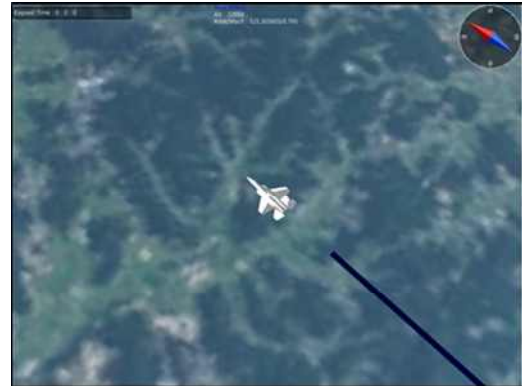


Fig. 4. Snapshot of a simulation

Table 3. Examples of behavior variables in terms of their names, classes, and meanings

행위 종류	행위 변수	의미
Velocity	Fast	속력 증가
	Slow	속력 감소
Fire	More	공격 가능성 증가
	Less	공격 가능성 감소
Turn	More	방향 전환 각도 증가
	Less	방향 전환 각도 감소

종사를 모델링하였다. 구체적으로 여섯 종류의 변수에는 Table 3과 같이 속력의 증가 및 감소, 공격 가능성의 증가 및 감소, 방향 전환 각도의 증가 및 감소가 있다.

3.2 실험 환경

다양한 실험 환경에서 제안하는 모형의 성능을 확인하기 위해 HMM의 은닉상태의 수 n_h 와 입력 벡터 시퀀스의 길이 l 에 따라 Table 4에 정리된 네 종류의 환경에서 실험을 수행하였다. 또한, dPCA에서 추출된 주요 요인의 수에 따른 성능 변화를 알아보기 위해 dPCA-HMM은 추출된 주요 요인의 수 n_p 가 각각 1, 3, 5인 경우에 대해

Table 2. Examples of status variables in terms of their names, types, and meanings

상태 변수	형태	의미
DestinationDistance	연속형	전투기의 목적지로부터의 거리
TargetAltitudeDifference	연속형	전투기와 타겟의 고도 차이
TargetDistance	연속형	전투기와 타겟의 거리
TargetExistence	이산형	전투 타겟의 존재 여부
Velocity	연속형	전투기의 속도
BVRCapability	이산형	전투기의 BVR(Beyond Visual Range) 전투 가능 여부
WeaponSelection	이산형	활성화된 무기의 종류(Gun, Short Range Missile, Midrange Missile, None)

Table 4. Four experiment settings according to the number of hidden states n_h and the length of input sequence l

	n_h	l
Setting 1	2	3
Setting 2	2	5
Setting 3	3	3
Setting 4	3	5

dPCA-HMM(1), dPCA-HMM(3), dPCA-HMM(5)의 세 종류의 모델을 실험하였다.

추가적으로, 제안하는 dPCA-HMM의 성능 비교를 위해 연속형 상태 변수만을 이용하여 학습한 연속 HMM, HMM(cf), 이산형 상태 변수만을 이용하여 학습한 이산 HMM, HMM(df), 연속형 변수를 이산화하여 이산형으로 통합된 전체 데이터를 이용하여 학습한 이산 HMM, df-HMM을 구현하였다. 이때, 연속형 변수의 이산화를 위해서는 각 변수 값의 범위를 동일한 너비로 나누는 기법과 최적의 너비를 찾는 기법을 사용하였다. 즉, 최적의 너비로 나누는 기법을 사용한 모델 df-HMM(opt)과 3개의 동일한 너비로 나누는 기법을 사용한 모델 df-HMM(3)을 실험하였다. 보다 유의미한 결과를 위해 총 120개의 시나리오 중 2개를 랜덤하게 선택하여 테스트에 사용하는 k-fold방식을 도입하였으며, 각 실험 환경 및 모델 별로 10회 반복 후 이들의 평균 정확도를 최종 정확도로 사용하였다.

제안한 기법의 성능을 평가하기 위해서 정확도(accuracy)가 도입되었다. 정확도는 전체 경우 중 옳게 예측한 경우의 비율을 나타내며 통계적 학습 기법 분야에서 널리 사용된다. 예를 들어, 모델이 세 개의 시점에 대해 조종사의 행위를 {Velocity:Fast, Fire:More, Velocity:Slow}로 예측하였고 실제 조종사의 행위가 {Velocity:Fast, Turn:More, Fire:Less}라면 정확도는 0.33이 된다. 구체적으로 본 실험에서는 시나리오 별 정확도를 계산하고 전체 정확도를 시나리오 별 정확도의 평균을 구하였다. 즉, 시점 $t=1, \dots, T$ 로 구성된 시나리오 N 개에 대한 정확도는 식 (4)와 같이 계산된다.

$$Accuracy = \frac{1}{N} \sum_{n=1}^N \frac{1}{T} \sum_{t=1}^T \delta(\hat{y}_t = y_t) \quad (4)$$

이때, $\delta(\cdot)$ 는 조건이 만족하면 1, 만족하지 않으면 0을 반환하는 함수이다.

3.3 실험 결과

제안하는 조종사 모델의 성능을 평가하기에 앞서, HMM의 EM 알고리즘의 반복수와 로그 유사도(log likelihood)값의 변화를 관찰하여 학습이 올바르게 이루어지는지 살펴보았다. Table 5는 각 행위 변수 별 EM 알고리즘 반복수의 평균, 최소, 최댓값을 나타낸다. 전체적으로는, 평균 11번 정도의 반복이 이루어져있고, 최소 4~6번 최대 11~59번의 반복이 이루어졌다. Fig. 5는 예시로써, 시나리오 21과 8의 반복수에 따른 로그 유사도 값의 변화를 나타낸다. 반복 수가 약 10에 가까워지며 두 경우 모두 수렴하는 것을 알 수 있다. 이는 EM 알고리즘을 통해 HMM의 학습이 이루어졌음을 보여준다.

Figure 6는 총 네 가지 실험 환경에서 일곱 종류의 조종사 행위 모델을 비교한 결과를 나타낸다. 구체적으로 (a)는 은닉상태가 2개, 벡터 시퀀스의 길이가 3인 실험 환경 1, (b)는 은닉상태가 2개, 벡터 시퀀스의 길이가 5인 실험 환경 2, (c)는 은닉상태가 3개, 벡터 시퀀스의 길이가 3인 실험 환경 3, (d)는 은닉상태가 3개, 벡터 시퀀스의 길이가 5인 실험 환경 4에서 일곱 모델의 실험 결과를 보여준다. 전반적으로 연속 HMM을

Table 5. Average, minimum, and maximum number of iterations of EM algorithm

행위 변수		평균	최솟값	최댓값
Velocity	Fast	21.0	6.0	59.0
	Slow	12.8	5.0	27.0
Fire	More	10.0	4.0	36.0
	Less	7.0	4.0	20.0
Turn	More	6.5	5.0	11.0
	Less	8.8	5.0	25.0
평균		11.0	4.8	29.7

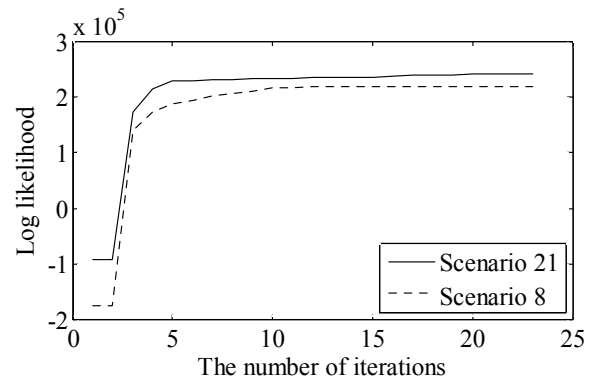


Fig. 5. Log likelihood of scenarios 21 and 8 according to the number of iterations

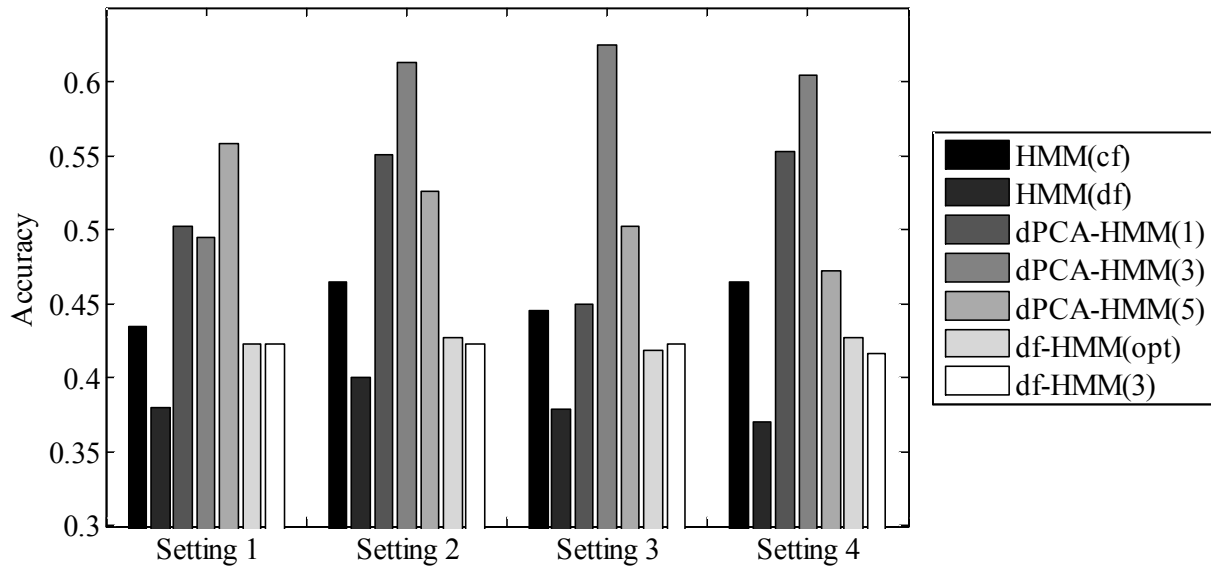


Fig. 6. Results of the performance comparison experiments

기반으로 한 모델들의 성능이 이산 HMM을 기반으로 한 모델에 비해 좋은 것을 알 수 있다. HMM(cf), dPCA-HMM 모델들의 평균 성능은 0.50, HMM(df), df-HMM 모델들의 평균 성능은 0.41로 10%의 성능 차이가 있음을 알 수 있다.

또한, Table 6에 나타난 것과 같이 네 실험 환경에 대해 일곱 모델의 HMM(cf), HMM(df), dPCA-HMM(1), dPCA-HMM(3), dPCA-HMM(5), df-HMM(opt), df-HMM(3) 평균 성능은 0.45, 0.38, 0.52, 0.52, 0.50, 0.42, 0.42로 제안하는 dPCA-HMM 모델들이 가장 좋은 성능을 보여주었다. 네 가지 실험 환경에서 가장 좋은 성능을 갖는 모델은 주요 요인의 수에 따라 다르게 나타났다. 주요 요인의 수가 1개 혹은 3개인 모델이 좋은 성능을 보였다. 실험 환경에 따라 그 수가 달랐다. 따라서 모델을 만들 때 환경에 따른 최적의 주요 요인의 수를 도출할 필요가 있다는 것을 알 수 있다.

Table 6. Average accuracy across the four settings of the seven models

Model	Average accuracy
HMM(cf)	0.45
HMM(df)	0.38
dPCA-HMM(1)	0.51
dPCA-HMM(3)	0.58
dPCA-HMM(5)	0.51
df-HMM(opt)	0.42
df-HMM(3)	0.42

실험 결과 전반적으로 50% 내외의 정확도를 보였는데 이는 본 연구에서 해결하고자 하는 조종사 행위 모델링 문제는 기존 연구를 통해 어렵다고 알려진 인간 행동 예측의 하나이며[20], 분류 카테고리의 수가 여섯 개로 무작위 예측 시 기대되는 정확도는 약 17% 정도로 낮기 때문일 것이다. 특히, HMM을 이용하여 사용자의 의도를 예측하는 기존 연구에서도 세 개의 카테고리 로 예측하는 경우 평균 65%, 네 개의 카테고리의 경우 평균 42%의 정확도를 보였으며[21], 유사한 주제의 또 다른 연구에서도 일곱 개의 카테고리를 예측하는데 약 30%에서 50%의 정확도를 나타내었다[22]. 추후 연구를 통해 성능을 실제 적용 가능 할 수준으로 올릴 것으로 기대한다.

위의 결과에서 가장 좋은 성능을 보인 dPCA-HMM(3)을 이용하여 은닉상태의 수와 시퀀스 벡터의 길이가 성능에 미치는 영향을 알아보기 위해 실험을 수행하였다. 은닉상태의 수 2,3,4,5,6와 시퀀스 벡터 길이 3,5,10,15,10에 대한 성능을 측정하였으며, 성능을 색의 음영으로 표현하는 열지도(heatmap)를 이용하여 Fig. 6과 같이 나타내었다. 이때, 색이 짙을수록 높은 정확도를 의미한다. 은닉상태가 3개이고 시퀀스 벡터의 길이가 3일 때 가장 좋은 성능을 나타내었으며, 평균적으로는 시퀀스 벡터 길이가 5일 때와 은닉상태의 수가 3일 때 좋은 성능이 도출되었다. 시퀀스 벡터의 길이가 5 이상이 되면 모든 은닉상태의 수에서 성능은 하락하였으며 은닉상태의 경우 시퀀스 벡터 길이에 따라 좋은 성능을 보이는 개수가 달랐다.

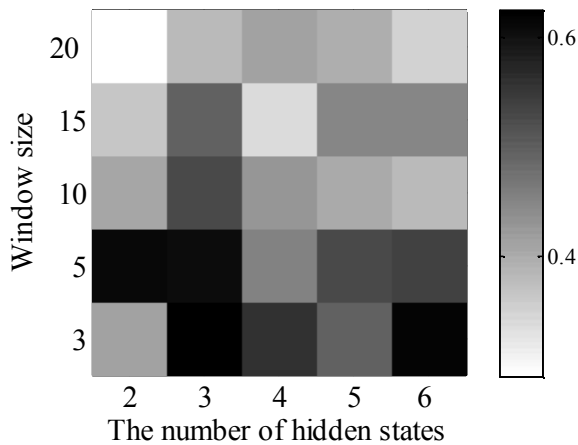


Fig. 7. Heatmap of accuracy of dPCA-HMM (3) according to the number of hidden states and window size

IV. 결 론

국방 M&S의 필요성이 대두 될수록, M&S를 가능하게 하는 필수요소인 조종사 모델링의 중요성이 부각되고 있다. 특히, 조종사의 행위를 모사할 수 있는 모델은 보다 실제적인 시뮬레이션을 가능하게 한다는 점에서 주목받는 연구주제이다. 따라서 본 논문에서는 기존에 축적된 데이터를 학습하여 조종사의 행위를 모사할 수 있는 통계적 학습 방법론 기반 조종사 모델링 기법을 제안한다. 이때, 학습에 필요한 데이터는 조종사 주변에 발생하는 상황정보를 나타내는 상태 변수와 그 시점에서 조종사가 수행한 행위를 나타내는 행위 변수로 구성된다. 특히 이 데이터는 시간에 따라 발생하는 시계열 데이터이며 상태 변수는 그 수가 많고 이산형 변수와 연속형 변수가 혼재되어 있다는 특징을 가진다.

이러한 특성에 따라 시계열 데이터를 모델링 하는데 효과적으로 알려져 있는 HMM을 조종사 모델링에 도입하였으며 이산형 변수와 연속형 변수의 통합을 위해 dPCA기법을 이용하였다. 구체적으로, 변수들의 형태 통합 과정에서 이산형 변수들은 더미 변수를 이용하여 이진법으로 표현되었고 이 변수들을 대상으로 PCA를 수행하여 주요 요인을 추출하였다. 이렇게 연속형 변수로 통합된 데이터를 이용하여 연속 HMM을 학습시켜 조종사 모델을 구축하였다.

제안하는 모델의 성능을 평가하기 위해서는 국방과학연구소 판급 시뮬레이터로부터 추출된 전투 로그를 사용하였다. 변수의 형태, 연속

HMM 혹은 이산 HMM의 종류에 따라 총 일곱 종류의 모델의 성능을 비교하였으며, 제안하는 dPCA-HMM이 10% 가량 향상된 정확도를 나타내었다. 이를 통해, 전반적으로 연속 HMM의 성능이 이산 HMM에 비해 좋다는 것을 확인 할 수 있었으며, 연속형 변수를 이산형 변수를 변환하는 것보다는 이산형 변수를 연속형으로 변환하여 사용하는 것이 성능 손실을 줄일 수 있다는 것을 알 수 있었다.

본 연구에서 사용한 dPCA기법은 이진화로 변환한 변수들간 존재하는 음의 상관관계를 반영하지 못한다는 한계가 있다. 따라서, 추후 연구를 통해 보다 정교화 된 연속형 통합 기법을 개발한다면 더욱 향상된 성능을 얻을 수 있을 것으로 기대된다. 또한, 동일한 상황에서 조종사가 취한 행위는 다수 개 일 수 있으며 실제로 실험 데이터의 전체 조종사의 행위 중 약 10% 는 동일 상황에서 이루어진 다수의 행위 중 하나이다. 이는 동일 상황에서 하나의 행위만을 예측하는 본 모델의 성능을 하락시키는 주원인으로 추후 다수 개의 행위를 효과적으로 예측하는 모델 개발을 통해 성능 향상이 가능할 것이다.

후 기

본 연구는 국방과학연구소의 연구비 지원 하에 인간 판단 모의 기법 기술 과제의 일부로서 수행되었습니다.

References

- 1) Jeong, S., "Required technology for implementing defense M&S," *Industrial engineering magazine*, vol. 20, no. 4, 2014, pp. 35~41.
- 2) Jang, D., Cho, S., Tahk, M., Koo, H., and Kim, J., "Fuzzy Logic Based Collision Avoidance for UAVs," *Journal of the Korean Society for Aeronautical and Space Sciences*, vol. 34, no. 7, 2006, pp. 55~62.
- 3) Won, D., Shim, S., Kim, K., Tahk, M., Seong, K., and Kim, E., "Track-Before-Detect Algorithm for Multiple Target Detection," *Journal of the Korean Society for Aeronautical and Space Sciences*, vol. 39, no. 9, 2011, pp. 848~857.
- 4) Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech

recognition," *Proceedings of the IEEE*, vol. 77, no. 2, 1989, pp. 257~286.

5) Lai, C., Lu, S. L., and Zhao, Q., "Performance analysis of speech recognition software," *Proceedings of the Workshop on Computer Architecture Evaluation using Commercial Workloads*, 2002.

6) Hayashi, M., "Hidden Markov Models to identify pilot instrument scanning and attention patterns," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2003, vol. 3, pp. 2889~2896.

7) Choi, Y., Kwon, N., Lee, S., Shin Y., Ryo, C., Park, J., Shin, D., "Hypo-vigilance Detection for UCAV Operators Based on a Hidden Markov Model," *Computational and Mathematical Methods in Medicine*, 2014.

8) Mori, R., Suzuki, S., "Modeling of Pilot Landing Approach Control Using Stochastic Switched Linear Regression Model," *Journal of Aircraft*, vol. 47, no. 5, 2010, pp. 1554~1558.

9) Lowe, C. D., "Predicting pilot intent and aircraft trajectory in uncontrolled airspace," Massachusetts Institute of Technology, 2014.

10) Andersson, M., Petterson, G., "Improving situation awareness using aerial-mission recognition and temporal information," *Proceedings of the International Conference on Information Fusion*, Stockholm, Sweden, 2004.

11) Trevo, K., "Human adaptive mechatronics methods for mobile working machines," Aalto-yliopiston teknillinen korkeakoulu, 2010.

12) Quinlan, J. R., "Learning decision tree classifiers," *ACM Computing Surveys*, vol. 28, no. 1, 1996, pp. 71~72.

13) Baek, J., Kim, C., and Kim, S., "Multi-Interval Discretization of Continuous Valued Attributes for Constructing Incremental Decision Tree," *Journal of the Korean Institute of Industrial Engineers*, vol. 27, no. 4., 2001, pp. 394.

14) Schenk, J., Schwarzler, S., Ruske, G., and Rigoll, G., "Novel VQ designs for discrete HMM on-line handwritten whiteboard note recognition," *Lecture Notes in Computer Science*, vol. 5096, 2008, pp. 234~243.

15) Krzanowski, W. J., "Discrimination and classification using both binary and continuous variables," *Journal of the American Statistical Association*, vol. 70, no. 352, 1975, pp. 782~790.

16) Filmer, D. and Pritchett, L. H., "Estimating wealth effects without expenditure Data-Or tears: An application to educational enrollments in states of india," *Demography*, vol. 38, no. 1, 2001, pp. 115~132.

17) Jolliffe, I., *Principal component analysis*, John Wiley & Sons, Ltd, 2005.

18) Kim, J., Goo, Y., and Lee, H., "Signal-based Fault Diagnosis Algorithm of Control Surfaces of Small Fixed-wing Aircraft," *Journal of the Korean Society for Aeronautical and Space Sciences*, vol. 40, no. 12, 2012, pp. 1040~1047.

19) Dempster, A. P., Laird, N. M., and Rubin, D. B., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal statistical Society*, vol. 39, no. 1, 1977, pp. 1~38.

20) Kim, E., Helal, S., Cook, D., "Human Activity Recognition and Pattern," *IEEE Transactions on Pervasive Computing*, vol. 9, no. 1, 2010, pp. 48~53.

21) Antwarg, L., Rokach, L., Shapira, B., "Attribute-Driven Hidden Markov Model Trees for Intention Prediction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 6, 2012, pp. 1103~1119.

22) Kiseleva, J., Lam, H. T., Pechenizkiy, M., Calders, T., "Predicting Current User Intent with Contextual Markov Models," *Proceedings of the IEEE International Conference on Data Mining Workshops*, 2013, pp. 391~398.