

논문 2009-3-4

강화학습에 기초한 지능형 웹 검색의 과잉적합 감소방안

Overfitting Reduction of Intelligence Web Search based on Enforcement Learning

한송이*, 정용규**

Song-Yi Han, Yong-Gyu Jung

요 약 강화학습을 통한 지능형시스템은 게임, 웹 검색 등 많은 분야에서 연구되고 있다. 좋은 훈련 모델은 훈련데이터에도 적합해야 하며 이전에 접해 보지 못한 레코드들도 정확하게 분류되어야 한다. 훈련 데이터에 잘 맞는 모델은 과잉적합 되어서 좋지 못한 일반화의 오류를 가질 수 있다. 어떤 분야에서도 이런 과잉적합은 피할 수 없는 문제이며 과잉적합을 방지하는 연구는 필요하다. 본 논문에서는 과잉적합을 감소시키기 위한 방법으로 자연계의 모델인 엔트로피와 돌연변이를 웹 검색에 적용하여 제시한다. 학습과정은 엔트로피의 변화량으로 설명될 수 있고, 자연계의 적자로 생존할 수 있는 돌연변이 현상은 데이터마이닝에서 엔트로피의 인위적 발생으로 설명될 수 있다. 즉, 최대 엔트로피를 주기적으로 발생시키는 방안을 본 논문에서 제시한다. 훈련데이터의 최대 엔트로피 모델은 지능형 웹 검색의 주기적 일반화 강화과정이라고 볼 수 있다.

Abstract Recent days intellectual systems using reinforcement learning are being researched at various fields of game and web searching applications. A good training models are called to be fitted with training data and also classified with new records accurately. A overfitted model with training data may possibly bring the unfavored fallacy of hasty generalization. But it would be unavoidable in actual world. The entropy and mutation model are suggested to reduce the overfitting problems on this paper. It explains variation of entropy and artificial development of entropy in datamining, which can tell development of mutation to survive in nature world. Periodical generation of maximum entropy are introduced in this paper to reduce overfitting. Maximum entropy model can be considered as a periodical generalization in intensified process of intellectual web searching.

Key Words : 강화학습, 웹 검색, 과잉적합, 엔트로피

I. 서 론

자연 현상은 관리를 제대로 하고자 하는 노력을 하지 않고 그대로 내버려 두면 모든 것이 잘 정돈되어 있던 질서상태에서 무질서한 상태로 진행된다. 이런 현상을 물리학적 용어로 엔트로피의 법칙이라고 부른다. 대부분 자연현상의 변화는 어떤 일정한 방향으로만 진행된다.

즉, 열이 낮은 온도에서 높은 온도로 흐르거나, 아무런 다른 변화 없이 열이 일로 바뀌는 것은 사람들의 경험을 통해 과거에 한 번도 일어난 일도 없고 앞으로도 일어나지 않을 것으로 사실상 확신할 수 있는 일들이다. 그러나 그것들이 이론적으로 불가능하다고 증명된 일은 없다. 따라서 일어날 수도 있지만 그 확률이 극히 낮아서 관측하기 불가능하다고 보는 것이 나올 것이다. 그렇다면 열역학 제2법칙이 제시하는 변화의 방향성, 예를 들어 열이 높은 온도에서 낮은 온도로 흐르는 것은 그 방향으로 변화가 일어날 확률이 지극히 높음을 이야기하는 것이다.

*정회원, 을지대학교 의료전산학전공

**정회원, 을지대학교 의료전산학전공(교신저자)

접수일자 2009.05.25, 수정완료.2009.06.05

또한 자연현상의 변화는 물질계의 엔트로피가 증가하는 방향으로 진행된다는 말은 엔트로피가 증가하는 변화가 일어날 확률이 지극히 높다는 것이 된다. 어떤 한 상태(상태1)에서 다른 한 상태(상태2)로의 변화가 항상 일어나기 위해서는 상태2가 상태1보다 확률이 지극히 높은 상태라야 한다. 결국 엔트로피는 이와 같이 확률과 연관되어 있다. 여러 분야에서 엔트로피를 이용한 알고리즘 및 모델이 계속적으로 연구되고 있다. 특히, 학습이나 선택의 문제에 있어서 엔트로피는 판단 기준이 되는 개념을 수치화하기에 좋은 성질을 가지고 있다. 즉 정보량(information content)의 정의를 세우고 주어진 정보의 량에 따라 엔트로피의 증감을 표현할 수 있다. 예를 들면, 불확실성의 증가, 선택의 자유 증가, 정보의 가치 증대를 엔트로피의 증가로 표현할 수 있다^[1]

일반적인 데이터 마이닝 분야에서 훈련데이터에 과잉적합된 학습결과는 새로운 문제에 대해서 좋은 결과를 얻지 못하는 경우가 많다. 웹 검색에서도 비슷한 문제가 야기되는데, 이는 웹 검색이 다양한 사전지식이 필요하며 현실적으로 이를 고려하지 못한 원인에 기인한다. 본 논문에서는 이런 과잉적합의 문제를 엔트로피를 이용하여 감소하는 방안을 제시한다. 훈련데이터에 주기적으로 최대 엔트로피를 강제로 줌으로써 복잡도를 인위적으로 높여 과잉적합을 방지하는 방안이다.

II. 관련연구

2.1 강화학습

강화 학습은 주어진 환경과의 상호 작용을 통하여 받는 보답의 합을 극대화하기 위한 각 상태에서의 최적행위 전략을 배우는 것이다. 강화 학습에서 사용되고 있는 agent 는 학습을 하는 개체로서, 행위(a)를 통해 환경과 상호 작용을 하며 환경으로부터 보답 신호를 받고, 환경(environment)은 agent 가 상호 작용하는 것으로 agent 에 임의로 바뀌지 않는다. Agent는 전략(policy)에 따라 행위를 취하게 되는데 이 전략 함수(π)는 임의의 상태(s)에서 행위들이 선택되어질 확률을 나타내고, 행위의 결과인 보답(reward)은 현재 및 지난 행위들을 평가해 준다.^[2]

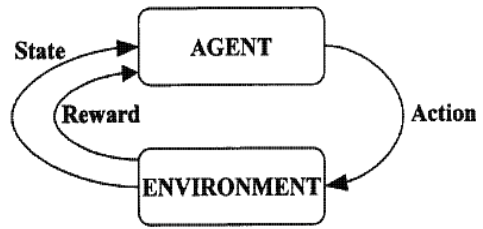


그림1. 강화학습의 상호작용
Fig. 1. Interactions in enforced learning

강화 학습이 적용될 수 있는 분야로는 로봇축구, 컴퓨터 게임, 전자 상거래, 웹 검색, 자가-관리 컴퓨터 시스템, 반-테러 응용 시스템, 실시간 이동 물체 추적^[3] 등 다양한 응용분야들이 존재한다.^[4]

2.2 강화학습을 이용한 웹 검색

웹 검색 문제는 문제의 특성상 강화 학습과 밀접한 관련이 있다. 웹 검색 에이전트는 여러 웹 페이지를 링크를 따라 이동하며 목적하는 문서를 찾는다. 즉 각 웹 페이지는 상태(state), 웹 페이지 내부의 링크는 행동(action)에 각각 대응되고 목적하는 웹 페이지에 도달했을 경우 보상(reward)을 받는 것에 대응되게 된다. 즉 보상을 얻기 위해서는 현재의 보상만이 아니라 미래의 보상을 고려해야 한다는 점과 올바른 답이 주어지는 것이 아니라 보상 형태로 목표가 정해진다는 점에서 웹 검색 문제는 강화 학습 문제에 적합하다.^[5] 이러한 방법으로 웹에서 존재하는 정보 중에서 특정 주제의 웹 페이지를 보다 더 정확히 찾게 되는 것이다.

일반적으로 강화학습을 이용한 정보검색은 많은 하이퍼링크 중에서 보상 값이 높은 링크를 선택하려는 경향을 보이게 된다. 이에 따라 많은 학습을 시킨 후에 나오는 결과는 자신에게 들어오는 많은 링크들을 갖고 있는 페이지인 전거나 자신에서 나가는 많은 링크들을 갖고 있는 허브 사이트를 찾게 되는 장점이 있다.^[6]

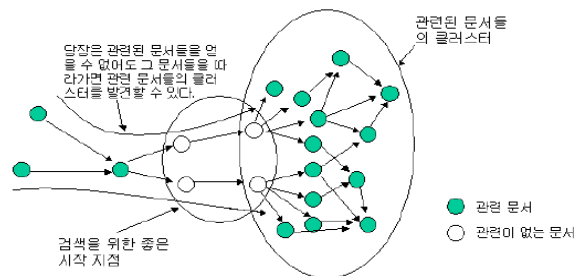


그림 2. 웹검색 에이전트의 탐색과정
Fig. 2. Process in web searching agent

2.3 과잉적합 문제

데이터 마이닝에서 데이터에 훈련 모델의 과잉적합은 새로운 데이터에 대해서는 적합하지 못한 예측의 결과를 가져오게 되는 경우가 있다.[7] 좋은 훈련 모델은 훈련 데이터에도 적합해야 하며 이전에 접해 보지 못한 레코드들도 정확하게 분류되어야 한다. 즉 훈련 데이터에 너무 잘 맞는 모델은 더 높은 훈련 오류를 갖는 모델보다 더 좋지 못한 일반화의 오류를 가질 수 있기 때문이다. 따라서 좋은 모델은 낮은 훈련 오류뿐만 아니라 낮은 일반화 오류도 가져야 한다. 이러한 상황을 모델의 과잉적합(Overfitting)이라고 한다.

과잉적합이 발생하는 이유로는 레이블에 잘못 부여된 훈련 레코드들에 적합하게 되어 결과적으로 시험 집합 레코드들의 분류 오류를 초래 하는 경우인 잡음으로 인한 과잉적합이 있으며, 적은 수의 훈련 레코드에 기초하여 분류를 결정하는 모델도 역시 과잉적합이 발생하게 되는 상황인 대표적인 견본으로 인한 과잉적합이 있다. 또한 다중 비교 절차(multiple comparison procedure)로 알려진 방법을 사용하는 학습 알고리즘에서 과잉적합이 발생 할 수 있다. 그 이유는 대안들의 수가 너무 많거나 훈련 레코드의 수가 너무 적은 것을 보상하지 못하게 되어 결국 모델의 과잉 적합이 발생하게 되는 것이다.

과잉적합의 주요 원인에 대해서는 아직 논의의 여지가 있지만, 모델의 복잡도가 모델 과잉적합에 영향을 준다는 것은 일반적이다.[8]

2.4 엔트로피

어떤 상태에 대한 엔트로피는 그 상태에 대한 확률의 척도일 뿐 아니라 ,또한 그 상태에 대한 무질서함(randomness)의 척도이기도 하다. 어떤 계의 무질서한 상태(예를 들어, 기체 상태)에 해당하는 분자들의 배열 방법 수는 매우 많고, 그대 대한 확률이 큰데 반해 질서가 있는 상태(예를 들어 고체 상태)에 대한 배열 방법 수와 그 확률은 극히 작기 때문이다. 52장의 카드놀이를 생각할 때 아무렇게나 섞여 있는 상태(무질서)와 다이아몬드 12장이 순서대로 끼어 있는 상태(질서 있는 상태)는 후자의 확률이 훨씬 작다는 것을 알 것이다. 카드를 아무렇게나 몇 번 치고 있으면 점차 확률이 큰 쪽으로 가게 된다. 10의 20제곱이 훨씬 넘는 분자들로 이루어진 기체 계의 분자배열에서는 그 정도가 얼마나 심하겠는가? 여기서 역행하는 확률이 거의 나타나지 않는다. 잘 정돈

된 진열장에서 유리컵이 떨어져 깨질 확률은 크지만 그 파편이 다시 모여서 온전한 유리컵이 될 확률은 거의 없다. 다시 말하면 엔트로피는 무엇보다도 물질계의 열적 상태로부터 정해진 양으로서, 통계역학의 입장에서는 열역학적인 확률을 나타내는 양으로 엔트로피 증가의 원리는 분자운동이 확률이 적은 질서 있는 상태에서부터 확률이 큰 무질서한 상태로 이동해 가는 자연적 현상으로 해석된다. 예를 들면, 마찰에 의한 열이 발생하는 것은 역학적 운동(분자의 질서 있는 운동)이 열운동(무질서한 분자운동)으로 변하는 과정이며 고온 물체에서 저온 물체로 열이 이동하는 것은 국부적으로 분자운동에 개별성이 있다고 하는 질서 있는 상태에서 분자운동이 균일해지는 무질서상태로 이동해 가는 것이다. 이들은 어느 경우나 엔트로피가 증가하지만 그 반대의 과정은 무질서에서 질서로 옮겨가는 것이며 이 과정은 자발적으로 일어나지 않는다. 일반적으로 열역학적 확률의 최대값은 온도가 균일한 열 평형상태에 대응하는 것으로서, 고찰하고 있는 물질계가 다른 에너지의 출입이 없는 고립계인 경우에는 늦던 빠르던 전체가 열 평형에 도달하여 모든 열 과정이 정지하는 것이라고 생각된다. 이로부터 엔트로피 증가의 법칙이 발표된 직후 만약 이 법칙이 정당하다면 우주는 어느 것이나 열 평형에 도달하여 모든 종류의 에너지가 분자의 불규칙적인 열 운동으로 변하여, 열의 종말, 즉 우주의 종말에 도달하게 될 것이라는 논쟁이 일어났다. 그러나 이는 우주를 고립된 유한한 계라고 가정했을 때의 결론이다.

2.5 엔트로피 적용

물질계의 열적 상태를 나타내는 물리량의 하나이다. 1865년 클라우지우스가 변화를 뜻하는 그리스어에서 유래했다. 이론적으로는 물질계가 흡수하는 열량 dQ 와 절대온도 T 와의 비 $dS=dQ/T$ 로 정의한다. 여기서 dS 는 물질계가 열을 흡수하는 동안의 엔트로피 변화량이다. 열기관의 효율을 이론적으로 계산하는 이상기관에서는 모든 과정이 가역과정이므로 엔트로피는 일정하게 유지되지만 일반적으로 현상이 비가역 과정인 자연적 과정을 따르게 될 때에는 이 양은 증가하고 자연적 과정에 역행할 때에는 감소하는 성질이 있다. 즉 자연현상의 변화가 자연적 방향을 따라 발생하는가를 나타내는 척도로 다루어진다. Shannon은 정보이론(information theory)에 관한 논문에서 정보량과 엔트로피를 다음과 같이 정의하였

다.[9] 먼저 n 개의 사건으로 구성된 다음과 같은 확률 분포가 존재할 때,

$$P_i \geq 0 \quad (i = 1, 2, \dots, n), \quad \sum_{i=1}^n P_i = 1 \quad (1)$$

이때 정보량 I 와 엔트로피 H 는

$$I = - \sum_{i=1}^n \log_2 P_i \quad (2)$$

$$H = - \sum_{i=1}^n P_i \log_2 P_i \quad (3)$$

여기서 엔트로피는 확률변수의 불확실성(uncertainty)에 대한 수학적 척도로서, 이상 확률 변수 X 의 엔트로피는 다음과 같이 정의된다.[10]

$$H(X) = - \sum_{\forall x} p(x) \log p(x) \quad (4)$$

일반적으로 이산 확률 변수 X 가 균일 분포를 따르는 경우 그 엔트로피 값은 최대가 된다. 따라서 최대 엔트로피 모델이 미지의 사실에 대해서 균일 분포를 갖는다는 것은 미지의 내용에 대해서 모델의 엔트로피가 최대가 된다는 것을 의미한다.

III. 과잉적합 감소방법

3.1 일반화오류 추정방법

강화학습을 이용한 웹 검색에서 발생하는 과잉적합 해결을 위한 방법중의 하나는 일반화 오류에 대한 추정이다. 이 방법은 어떤 모델의 복잡도가 모델의 과잉적합에 영향을 준다는 일반적인 견해에서 시작한다. 문제는 어떻게 올바른 모델 복잡도를 추정할 것인가이다. [11][8] 지금까지의 연구된 추정방법들은 다음과 같이 제시될 수 있다.

첫째, 훈련오류에 대한 통계적 보정(statistical correction)으로 추정될 수 있다. 보통 일반화 오류가 훈련 오류보다 더 크므로, 일반적으로 통계적 보정은 특정한 단달 노드에 도달한 훈련 레코드 수를 고려한, 훈련

오류에 대한 상한(upper bound)으로 계산된다. 그러나 일반화 오류를 계산하기 위해서는 관찰된 훈련 오류에 대한 상한값을 정하여야 한다. 따라서 요구되는 계산들이 낭비적일 수 있다.

둘째, 검증 집합(validation set)을 이용하는 방법이다. 이 방법은 훈련 집합을 사용하여 일반화 오류를 추정하는 대신, 원래 훈련 집합이 두 개의 작은 부분집합으로 나누어진다. 부분 집합 중 하나는 훈련을 위해 사용되고, 다른 하나는 검증 집합으로써 일반화 오류를 추정하기 위하여 사용된다. 이 방법은 복잡도를 갖는 모델들을 구하기 위해 매개변수화 될 수 있는 분류 기법에 일반적으로 사용된다. 복잡도가 가장 좋은 모델은 학습 알고리즘에 의하여 생성되는 실험적 모델이 검증 집합에 대하여 가장 낮은 오류율에 도달할 때까지 학습 알고리즘의 매개변수를 조정함으로써 추정할 수 있다. 그러나 이 방법은 이전에 접하지 않은 레코드들에 대하여 모델이 얼마나 잘 수행하는지를 추정하기 위한 더 나은 방법은 제시하지만, 훈련을 위한 데이터가 부족한 단점이 있다.

3.2 주기적 최대 엔트로피 강화방법

본 논문에서 제시되는 훈련데이터의 최대 엔트로피 모델은 지능형 웹 검색의 주기적 최대 엔트로피 강화방법으로 그 수행과정을 표현하면 다음과 같다.

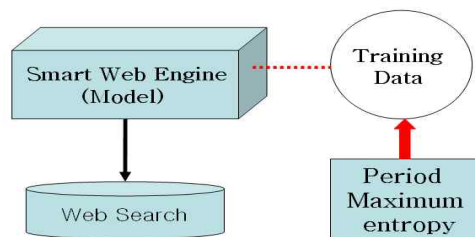


그림 3. 주기적 최대 엔트로피 강화방법 수행과정
Fig. 3. Periodical generation of maximum entropy

엔트로피를 이용하여 지능형 웹 검색에서 바이어스 문제로 나타나는 과잉적합의 문제를 해결하기 위해서는 결국 모델 복잡도(model complexity)를 증가시킴으로써 모델의 훈련 오류를 감소시킬 수 있다. 즉 모델이 복잡할수록 과잉적합은 더욱 증가 한다.^[5] 강화학습을 이용한 웹 검색에서 지나친 바이어스의 문제로 새로운 검색에 능동적이지 못하다는 것은 훈련 오류율은 높지만 전체적인 시험 오류율은 낮다는 것이다. 따라서 엔트로피의 개

념을 적용하여 훈련 데이터의 복잡도를 증가시켜 훈련 오류를 높이고, 결국 모델 복잡도의 증가를 막게 된다. 훈련데이터에 주기적으로 최대 엔트로피를 강제로 줌으로써 복잡도를 인위적으로 높여 과잉적합을 방지한다. 이는 자연현상에서 돌연변이와 비슷한 이치로 볼 수 있다. 주기적으로 돌연변이를 인위적으로 발생시킴으로써 적자가 된 군집(Population) 들은 진화한다는 이치와 비슷하다고 볼 수 있다. 학습대상의 모델이 복잡도가 증가하면, 학습의 결과는 로컬영역(Local Area)에 도달되어 과잉적합될 가능성이 높다. 따라서 모델의 복잡도를 감소시키기 위해 주기적, 인위적으로 훈련데이터의 복잡도를 증가할 필요가 있다. 결국 시험 오류율이 감소하게 되고, 과잉적합은 해결될 수 있다.

훈련 데이터의 복잡도를 인위적으로 최대가 되기 하기 위해서는 이산 확률 변수를 균일분포에 따르게 하면 된다. 따라서 최대 엔트로피 모델은 다음과 같은 형태를 지닌다.

$$P(x,y) = \frac{\prod_i \mu_i^{f_i(x,y)}}{Z} \quad (5)$$

여기서, μ_i 는

$$\mu_i = \exp(\lambda_i) \quad (6)$$

이고, λ_i 는 실수 파라미터이다. 또한, Z 는 정규화 상수이고 $f(x,y)$ 들은 한 예제 (x,y) 의 자질들이다. 각 자질 $f(x,y)$ 에 대해서, 학습 집합 $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 에 대한 분포 p 하의 기댓값은 특정 값 K_i 로 규정된다.

$$E_p[f_i] = \sum_{j=1}^N p(x_j, y_j) f(x_j, y_j) = K_i \quad (7)$$

K_i 의 값은 또한 S 의 경험적 분포 \tilde{p} 하의 기댓값으로도 설정된다. 따라서 제약은 다음과 같다.

$$E_{\tilde{p}}[f_i] = \sum_{j=1}^N \tilde{p}(x_j, y_j) f_i(x_j, y_j) = E_p[f_i] \quad (8)$$

또한 \tilde{p} 하의 f_i 의 기댓값을 K_i 와 비교한 후, μ_i 를 다음 식에 의해 변경한다.

$$\mu_i = \mu_i \cdot \frac{K_i}{E_{\tilde{p}}[f_i]} \quad (9)$$

그 다음, 경험적 분포 \tilde{p} 는 새로운 μ_i 를 가지고 변경된다.

$$\tilde{P}(x,y) = \frac{\prod_i \mu_i^{f_i(x,y)}}{Z} \quad (10)$$

식(10)을 이용하여 최대 엔트로피 모델을 훈련데이터 주기적으로 줌으로써 복잡도를 증가시킨다. 다시 말해 강화학습을 이용한 웹 검색의 시험 오류율은 낮아지게 되는 주기적 최대 엔트로피 강화방법을 실험 및 검증하고 있으며 이를 본 논문에서 제안한다.

IV. 결 론

웹 검색에서 지나친 바이어스의 문제로 새로운 검색에 능동적이지 못하다는 것은 훈련 오류율은 높지만 전체적인 시험 오류율은 낮다는 것이다. 훈련 데이터에 너무 잘 맞는 모델은 더 높은 훈련 오류를 갖는 모델보다 더 좋지 못한 일반화의 오류를 가질 수 있다는 뜻이다. 이런 문제를 해결하기 위해 자연계에서 사용하는 물리량인 엔트로피를 적용한다. 지능형 웹 검색에서 바이어스 문제로 나타나는 과잉적합의 문제를 해결하기 위해서는 결국 모델 복잡도를 증가시킴으로써 모델의 훈련 오류를 감소시킬 수 있다. 엔트로피의 개념을 적용하여 훈련 데이터의 복잡도를 증가시켜 훈련 오류를 높이고, 결국 모델 복잡도의 증가를 막게 된다. 훈련데이터에 주기적으로 최대 엔트로피를 강제로 줌으로써 복잡도를 인위적으로 높여 과잉적합을 방지한다. 향후 실험을 통해 제시된 방법의 효율성을 검증할 계획이다.

참 고 문 헌

- [1] 김정인, 엔트로피를 이용한 다개체 로봇 시스템의 자율 그룹핑 방법에 관한 연구, 건국대학교 석사 논문,

- 2007
- [2] 차중환, 연속 상태 및 행위 공간에 대한 영역 Q-학습 기법 연구, 석사학위논문, 2001
- [3] 김상현, 정재영, 윤학수, 박민욱, 김과형 “강화학습을 사용한 실시간 이동 물체 추적에 관한 연구” 한국퍼지 및 지능 시스템학회 추계학술대회 학술발표논문집, 2003
- [4] 권기덕, 효율적인 멀티 에이전트 강화학습을 위한 나이브베이지안 기반 상대 정책 모델, 박사학위논문, 2008
- [5] 이승준, 장병탁 “강화학습을 사용한 개인화된 웹 검색” 한국퍼지 및 지능시스템학회 학술발표논문집, 2001.
- [6] 정태진, 강화학습을 이용한 웹 정보 검색, 서울대학교 석사 논문, 2002
- [7] Ron Kohavi, Dan Sommerfeldt “Feature Subset Selection using Wrapper Method: Overfitting and Dynamic Search Space Topology, AAAI, 1995
- [8] Ian H. Witten, Frank Eibe, “Data Mining: Practical Machine Learning Tools and Techniques”, Morgan Kaufmann, 2000.
- [9] C. E. Shannon, “A Mathematical Theory of Communications,” Bell System Technical Journal, 1948, pp.623-656
- [10] 장정호, 장병탁, 김영택, “최대 엔트로피 기반 문서 분류기 학습” 한국정보과학회 학술 발표논문집, 1999
- [11] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, “Introduction to Datamining”, Addison-Wesley, 2005
- [12] 박성배, 장병탁 “최대 엔트로피 모델을 이용한 텍스트 단위화 학습” 한글 및 한국어정보처리학술대회, 2001

저자 소개

한 송 이(정회원)



• 2007년~현재 을지대학교 의료전산학
진공 재학중
<주관심분야: 데이터마이닝>

정 용 규(정회원)



• 1981 서울대학교 (이학사)
• 1994년 연세대학교 (공학석사)
• 2003년 경기대학교 (이학박사)
• 1999년~현재 을지대학교 교수
• 1994~현재 UN/Cefact/TBG3멤버
• 2001~현재 ISO/TC154K 위원장
• 2005~현재 산업표준(KS)심의위원
<주관심분야: 임상데이터마이닝, 의료정보시스템, e-비즈니스 국제표준>