

# SVDD 기반 중요문서 변조 유출 탐지 알고리즘

길 지 호,<sup>1\*</sup> 남 기 효,<sup>2\*</sup> 강 형 석,<sup>1</sup> 김 성 인<sup>1</sup>  
<sup>1</sup>고려대학교 정보경영공학전문대학원, <sup>2</sup>(주)위너다임

## An Algorithm for Detecting Leak of Defaced Confidential Information Based on SVDD

Ji-Ho Ghil,<sup>1\*</sup> Kihyo Nam,<sup>2\*</sup> Hyung-Seok Kang,<sup>1</sup> Seong-in Kim<sup>1</sup>

<sup>1</sup>Graduate School of Information Management Engineering, Korea University,

<sup>2</sup>WinnerDigm Inc.

### 요 약

본 논문은 보호하고자 하는 중요문서의 다양한 변조를 통한 유출시도를 정확히 탐지하는 알고리즘을 제시한다. 중요문서는 내부자에 의해 다양한 방법으로 변조된 후 유출이 시도되고 있으나, 중요문서 유출탐지에 관한 기존 연구들은 유사도를 기반으로 함으로써 중요정보에 대한 다양한 변조 형태를 정확히 반영하지 못하여 탐지 정확도가 떨어지는 단점이 있다. 본 연구는 이를 해결하기 위해 SVDD(Support Vector Data Description)을 이용한 새로운 중요문서 유출 탐지 알고리즘인 v-SVDD 알고리즘을 제시한다. 본 연구에서 제시한 알고리즘 수행결과는 기존 연구결과와 비교할 때 변조 유출 탐지 측면에서 우수한 정확도를 보여준다.

### ABSTRACT

This paper proposes the algorithm which addresses the problem of detecting leak of defaced confidential documents from original confidential document. Generally, a confidential document is defaced into various forms by insiders and then they are trying to leak these defaced documents to outside. Traditional algorithms detecting leak of documents have low accuracy because they are based on similarity of two documents, which do not reflect various forms of defaced documents in detection. In order to overcome this problem, this paper proposes a novel v-SVDD algorithm which is based on SVDD, the novelty detection algorithm. The result of experiment shows that there is significant improvement in the accuracy of the v-SVDD in comparison with the traditional algorithms.

**Keywords:** SVDD, Data Loss Prevention, Contents Filtering

## 1. 서 론

최근 정보통신 기술의 급속한 발전과 통신 인프라의 확대에 내부의 중요문서가 통신망을 통해 외부로 유출되는 사고가 급증하고 있으며, 이에 따라 많은 피해 사례가 발생하고 있다. 특히, 중요문서를 불법적으로

로 유출하고자 하는 내부자는 중요문서를 원형대로 유출하기 보다는 다양한 형태로 변조하여 유출을 시도하고 있기 때문에 이 문제를 해결하기 위한 중요문서 유출탐지 알고리즘의 연구 필요성이 증가하고 있다[1].

최근 중요문서 유출 탐지를 위한 많은 연구가 진행되고 있는데, 이러한 연구들은 대부분 문서간의 유사도를 기반으로 하고 있으며, 이를 세부적 분류하면 인용기반 유사도 탐지기법, 문자기반 유사도 탐지기법, 인용-문자 결합기반 탐지기법 등으로 나눌 수 있다. 먼저 인용기반 유사도 탐지기법은 중요문서와 비교문

접수일(2009년 10월 21일), 수정일(2009년 12월 2일),  
게재확정일(2009년 12월 3일)

\* 주저자, cadoil@korea.ac.kr

# 교신저자, nkh@wdigm.co.kr

서 등 두 문서가 공통의 참조 특성치를 기반으로 결합 단위들을 공유하는 수준을 측정하여 문서간의 유사도를 분석하고 이를 바탕으로 유출여부를 탐지하는 기법으로 Vladutz 등에 의해 제안되었다[6]. 또한 문자 기반 유사도 탐지 기법은 문서 안에 포함된 문자 특성치 분석을 바탕으로 유사도를 분석하여 유출 여부를 탐지하는 기법으로 Janssens 등이 관련 연구를 수행하였다[4]. 한편, Janssens 등, Per 등은 기존의 인용기반 유사도 탐지 기법과 문자기반 유사도 탐지 기법을 결합한 인용-문자 결합형 유사도 탐지 기법을 제시하였다[2,4].

이와 같은 기존 연구들은 모두 중요문서와 비교문서의 유사도를 분석하여 유사도가 일정 수준의 임계치 이상이 되는 경우를 유출시도로 탐지하기 때문에 유사도 분석방법의 정확도가 중요문서 유출탐지 정확도에 가장 큰 영향을 끼치는 특징이 있다. 특히 이 연구들은 모두 보호하고자 하는 중요문서와 비교문서의 일치성을 기반으로 유사도를 분석함으로써 중요문서를 변조하여 유출을 시도하는 경우에는 유사도가 낮아지고 이에 따라 탐지 정확도가 떨어지는 문제점이 있다.

본 연구에서는 이러한 기존 연구들의 문제점을 해결하기 위한 방법으로 이상 탐지를 위해 최근에 개발된 SVDD 기법을 적용한 새로운 중요문서 유출탐지 알고리즘으로써 v-SVDD 알고리즘을 제시한다. SVDD는 높은 정확도를 갖는 단일 클래스 분류 기법으로 최근 가장 많은 연구가 수행되고 있는 이상탐지 기법으로 여러 문서들의 학습을 바탕으로 문서들을 하나의 클래스로 판단하는 공통 영역을 결정하고 이를 바탕으로 새로운 문서의 이상 여부를 탐지하는 기법이다.

SVDD를 수행하기 위해서는 단일 클래스를 분류하기 위한 다수의 학습 문서가 필요하므로, 보호하고자 하는 하나의 중요문서에 대해서는 SVDD를 직접 적용할 수 없는 문제점이 있다. 또한 보호하고자 하는 여러 개의 중요문서들을 단일 클래스로 분류하여 중요문서 유출을 탐지하는 것은 개별적인 중요문서의 특성을 고려하지 못하여 탐지 정확도가 떨어지는 단점이 있다.

본 연구는 이러한 문제점을 해결하기 위해 보호하고자 하는 하나의 중요문서를 중심으로 가상벡터그룹을 형성하여 v-SVDD 알고리즘을 수행한다. 이에 따라 여러 개의 중요문서를 보호하고자 할 때에는 각 중요문서에 대해 개별적으로 SVDD 알고리즘을 수행함으로써 각 중요문서 별 독립적인 보호영역이 설정되고 이에 따라 각 중요문서 별 유출 탐지 정확도가 증가하

게 된다.

본 연구에서 제시한 알고리즘은 정확한 중요문서 유출탐지를 위해서 보호하고자 하는 중요문서를 벡터화하고, 이를 바탕으로 변형될 수 있는 다양한 변조 형태의 문서를 가상 변조벡터로 구성한다. 그리고 이 벡터들을 학습 데이터로 하여 SVDD 기법을 적용한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 본 논문과 관련된 기존 연구를 설명하고 3장에서는 SVDD 기법을 이용한 중요문서 유출 탐지 알고리즘을 제시한다. 또한 4장에서는 제시하는 중요문서 유출 탐지 알고리즘을 수행한 결과를 기존 연구의 수행결과와 비교한다. 마지막으로 5장에서는 결론 및 향후 연구 과제를 제시한다.

## II. 기존 연구

### 2.1 유사도 기반 탐지 기법

유사도 기반 탐지 기법은 두 개의 문서의 유사도를 기반으로 유출 여부를 탐지하는 기법이며, Vladutz 등의 인용기반 유사도 탐지방법[6], Janssens 등의 문자기반 유사도 탐지방법[4]와 Janssens 등, Per 등의 인용-문자 결합형 유사도 탐지방법[2,4] 등으로 다양하게 연구되어 중요문서 유출 탐지에 이용되고 있다.

#### 2.1.1 인용기반 유사도 탐지 알고리즘(CBA)

Vladutz 등이 제안한 인용기반 유사도 탐지 알고리즘(CBA: Citation Based Algorithm)은 사전화된 중요 단어 목록인  $n$  개의 참조 특성치로부터 중요문서와 비교문서에 포함된 참조 특성치의 포함여부를 판단하여 이를 바탕으로 각 문서를 벡터화한 후 유사도를 탐지한다. 인용기반 유사도 탐지를 위해서는 중요 단어 목록을 바탕으로  $n$  차원의 참조 특성치 벡터  $\beta$ 를 구성하고, 벡터  $\beta$ 를 구성하는 각각의 특성치  $r_m (m=1, \dots, n)$ 이 문서  $d_i$ 에 포함되었는지를 비교하여 문서  $d_i$ 의  $m$  번째 벡터값  $w_{m,i}$ 를 산출한다. 이 때, 문서  $d_i$ 의  $m$  번째 벡터값  $w_{m,i}$ 는 다음 식 (1)로 표현된다.

$$w_{m,i} = \begin{cases} 1 & \text{if } r_m \in d_i \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

이 벡터를 바탕으로 중요문서  $d_i$  와 비교문서  $d_j$  를 벡터로 표현하며, 두 문서의 유사도는 다음 식 (2)에 의해 산출한다.

$$sim(d_i, d_j) = \frac{\sum_{m=1}^n (w_{m,i} \times w_{m,j})}{\sqrt{\sum_{m=1}^n (w_{m,i})^2} \times \sqrt{\sum_{m=1}^n (w_{m,j})^2}} \quad (2)$$

이렇게 산출한 중요문서와 비교문서의 유사도가 임계치보다 큰 경우, 중요문서 유출로 판단한다.

### 2.1.2 문자기반 유사도 탐지 알고리즘(TBA)

Janssens 등이 제안한 문자기반 유사도 탐지 알고리즘(TBA: Text Based Algorithm)은 중요문서와 비교문서에 포함된 단어의 빈도를 바탕으로 각 문서를 벡터화하여 이를 기반으로 유사도를 탐지한다. 이 때, 각 문서가 갖는 벡터의 차원을  $m$ 이라 하면, 문서  $i$ 에 대한 각 벡터 값  $w_{m,i}$ 은 다음으로 표현된다.

$$w_{m,i} = freq_{m,i} \times \log\left(\frac{N}{n_m}\right) \quad (3)$$

이 벡터를 바탕으로 중요문서  $d_i$  와 비교문서  $d_j$  를 벡터로 표현하며, 두 문서의 유사도는 앞의 인용기반 유사도 탐지 알고리즘과 마찬가지로 식 (2)에 의해 산출한다. 이렇게 산출한 중요문서와 비교문서의 유사도가 임계치보다 큰 경우, 중요문서 유출로 판단한다.

### 2.1.3 인용-문자 결합형 유사도 탐지 알고리즘(CA)

Janssens 등, Per 등이 제안한 인용-문자 결합형 유사도 탐지 알고리즘(CA: Combination Algorithm)은 앞에서 제시한 인용기반 유사도 탐지 알고리즘과 문자기반 유사도 탐지 알고리즘을 결합한 것으로, 각 알고리즘의 수행결과를 결합하여 유사도를 산출하는 방법이다. 인용기반 유사도 탐지 알고리즘에 의한 유사도를  $sim_{cit}$ , 문자기반 유사도 탐지 알고리즘에 의한 유사도를  $sim_{text}$ 라 할 때, 인용-문자 결합형 유사도 탐지 알고리즘에 의한 유사도  $sim_{comb}$ 는 다음 식 (4)에 의해 산출된다.

$$sim_{comb} = w \times sim_{cit} + (1-w) \times sim_{text} \quad (4)$$

CA는 이 유사도를 산출한 후, 앞의 두 가지 알고리즘과 마찬가지로 중요문서와 비교문서의 유사도가 임계치보다 큰 경우, 중요문서 유출로 판단한다.

## 2.2 SVDD

SVDD는 특정 벡터 그룹에서 이상 벡터를 탐지하기 위해 David 등[8]이 처음으로 제시한 단일 클래스 분류기법이다. 이 기법은 비선형 변환을 통해 데이터를 고차원의 벡터공간으로 사영시킨 후 사영된 모든 데이터들을 포함하는 최소한의 반지름을 갖는 구형의 경계를 산출함으로써 동일한 특성을 갖는 단일 클래스를 분류한다.

SVDD는 단일 클래스의 영역을 정하기 위해 학습을 수행하는데, 학습을 하기 위한 단일 문서의 집합을  $\{d_i\}$  (단,  $i=1, \dots, n$ ) 라 하면, 모든 문서들을 포함하는 최소의 구  $S$ 는 다음 최적화 문제로 모형화할 수 있다.

$$Min \quad F(R, a) = R^2 \quad (5)$$

$$S.T. \quad \|d_i - a\|^2 \leq R^2 \quad (6)$$

SVDD에서는 이 최적화문제를 보다 유연성 있는 문제로 현실화하기 위해서 여유변수(slack variable)를 추가함으로써 수정된 최적화문제를 구성한다.

$$Min \quad F(R, a) = R^2 + C \sum_i \xi_i \quad (7)$$

$$S.T. \quad \|d_i - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad (8)$$

이 모형을 라그랑지안 완화법과 편미분을 통해 다음의 쌍대(dual) 문제로 모형화한다.

$$Max \quad L = \sum_i \alpha_i K(d_i, d_i) - \sum_{i,j} \alpha_i \alpha_j K(d_i, d_j) \quad (9)$$

$$S.T. \quad 0 \leq \alpha_i \leq C, \quad \sum_i \alpha_i = 1, \quad j = 1, 2, \dots, l \quad (10)$$

여기서  $K(d_i, d_j) = \Phi(d_i) \cdot \Phi(d_j)$ 는 머시 조건을 만족하는 커널 함수이다. 이 최적화문제의 해를 구하면,  $\alpha_i > 0$  값을 갖는  $d_i$ 들은 모두 분류하고자 하는 클래스의 경계에 위치하는 서포트 벡터가 되며, 이를 이용하여 보호영역에 대한 반지름  $R$ 과 영역의 중심  $a$ 를 산출할 수 있다. 따라서 비교벡터  $z$ 가 아래 식 (11)을 만족하면 특정 클래스에 포함되고, 만족하지 않으면 이상 벡터라고 판단한다.

$$\|z - a\|^2 = K(z, z) - 2 \sum_i \alpha_i K(z, d_i) + \sum_{i,j} \alpha_i \alpha_j K(d_i, d_j) \leq R^2 \quad (11)$$

### III. 중요문서 유출탐지를 위한 v-SVDD 알고리즘

본 절에서는 중요문서 유출탐지를 위해 본 연구에서 제시하는 v-SVDD 알고리즘의 세부 절차를 기술한다. SVDD 기법에서 단일 클래스를 분류하기 위한 학습을 위해서는 많은 학습데이터가 필요한데, 이를 중요문서 유출탐지에 적용하기 위해서도 보호하고자 하는 하나의 중요문서의 보호영역을 설정하기 위한 다수의 학습데이터가 요구된다. 본 논문에서 제시하는 알고리즘에서는 단일 중요문서에 대한 SVDD 학습을 위해서 중요문서 및 관련된 다양한 변조 형태들을 가상 변조벡터 그룹으로 하여 학습을 수행한다. 특히 본 논문에서 제시하는 알고리즘에서 가상 변조벡터는 중요문서를 벡터화 한 후, 그 벡터의 요소들 중 일부분을 선택적으로 활용한 벡터들을 가상 변조벡터로 생성함으로써 구성된다. 이 가상 변조벡터를 이용하여 SVDD 기반의 보호영역을 설정하고, 이를 바탕으로 비교문서의 중요문서 유출 여부를 판단한다. 중요문서 유출 탐지를 위한 v-SVDD 알고리즘의 세부 수행절차는 다음과 같다.

#### 3.1 가상 변조벡터 생성

(단계 1-1) 중요문서의 특성치 추출

보호하고자 하는 중요문서  $d_0$ 로부터  $n$ 개의 특성치를 추출하여 집합  $F$ 를 산출한다.

$$F = \{f_1, \dots, f_n\}$$

(단계 1-2) (단계 1)의 결과를 이용하여 중요문서  $d_0$ 를 다음과 같이 벡터  $\mathbb{X}$ 로 변환한다.

$$d_0 \rightarrow \mathbb{X} = (x_1, x_2, \dots, x_n)$$

(단계 1-3) 다음 절차에 따라 가상 변조벡터 후보들을 생성한다.

- ① 특성치집합  $F$ 의 멱집합  $2^{|F|}$ 를 구한 후,  $\emptyset$ 을 제외한 집합  $V = (2^{|F|} - \emptyset)$ 를 구한다.
- ②  $V = \{v_1, v_2, \dots, v_{|V|}\}$ 로 놓는다.
- ③  $|V|$ 개의 가상 변조벡터 후보들을  $y^k = (y_1^k, y_2^k, \dots, y_n^k)$  ( $k = 1, \dots, |V|$ )로 놓고, 다음 식을 이용하

여 산출한다.

$$y_i^k = \begin{cases} x_i & \text{for } f_i \in v_k \\ 0 & \text{otherwise} \end{cases}, \text{ 단, } f_i \in v_k$$

(단계 1-4) 가상 변조벡터 후보  $y^k = (y_1^k, y_2^k, \dots, y_n^k)$  ( $k = 1, \dots, |V|$ ) 중 0인 요소의 개수가  $\theta$ 이하인 것을 모두 가상변조벡터로 선택하기 위해 다음 절차를 따른다.

① 허용할 수 있는 변조 특성치의 수  $\theta$ 를 정한다.

②  $s_j = y_j^k$ , if  $\sum I(y_i^k) \geq n - \theta$ , for all  $k$ ,

$$\text{여기서, } I(k) = \begin{cases} 1 & \text{if } k > 0 \\ 0 & \text{otherwise} \end{cases}$$

#### 3.2 보호영역설정을 위한 학습

(단계 2-1) SVDD 최적해 산출

보호하고자 하는 중요문서 벡터  $\mathbb{X}$ 와 가상 변조벡터  $s_j$ 를 식 (9), (10)의 최적화 모형에 적용하여 최적해를 산출한다.

(단계 2-2) 보호영역의 결정

"(단계2-1)"에서 구한 최적해를 바탕으로  $\alpha_i > 0$  값을 갖는 서포트 벡터를 찾고, 이를 이용하여 보호영역에 대한 반지름  $R$ 과 중점  $a$ 를 산출한다.

#### 3.3 중요문서 유출 탐지

(단계 3-1) 비교문서의 벡터화

보호하고자 하는 중요문서 벡터집합  $F$ 를 이용하여 비교문서를 벡터  $z$ 로 변환한다.

(단계 3-2) 비교문서의 중요문서 변조유출 탐지

비교벡터  $z$ 가 식 (11)을 만족하면 중요문서 유출이라고 판단하고, 만족하지 않으면 중요문서가 아니라고 판단한다.

## IV. v-SVDD 실험 및 결과

### 4.1 실험 개요

본 논문에서 제시한 v-SVDD 알고리즘의 정확도를 평가하기 위해 문서분류 연구에서 알고리즘 정확도 검증을 위해 널리 이용되는 Australian Credit 데이터를 이용하였다. 이 데이터는 14개의 특성치를 가

진 벡터로 구성되어 있는데, 본 연구에서는 같은 그룹에 속한 469개의 문서를 중요문서로 선택하여 유출탐지 실험을 수행하였다. 특히 본 실험에서는 각각의 중요문서를 최대 30%까지 변조하여 각각의 중요문서마다 1,471개의 학습데이터를 생성한 후, 실험에 활용하였다. 또한 정상적인 문서반출을 중요문서의 유출로 잘못 탐지하는 과탐율을 판단하기 위해 중요문서와 다른 그룹에 속한 324개의 비교문서를 선택하여 실험에 이용하였다. 한편, 본 연구에서 제시한 v-SVDD 알고리즘의 정확도를 판단하기 위해 수행결과를 TBA, CBA, CA 등 기존 연구들의 수행결과와 비교하였다.

### 4.2 평가척도

본 논문에서 제시한 v-SVDD 알고리즘의 수행결과를 비교하기 위한 평가척도는 Ray 등(3)이 사용한 탐지정확도(A)와 변조유출탐지 정확도(SR)를 이용하였다. 본 논문에서 사용한 두 개의 평가척도를 산출하는 식은 식 (12), (13)과 같다.

$$A = \frac{\text{정확하게 판단한 데이터수}}{\text{전체 실험 데이터수}} \quad (12)$$

$$SR = \frac{\text{변조유출로 판단한 데이터수}}{\text{전체 변조유출 데이터수}} \quad (13)$$

### 4.3 실험결과

본 논문에서는 v-SVDD의 수행결과와 정확도를 판단하기 위해 CBA, TBA, CA 등 3가지 기존 알고리즘의 수행결과와 비교하였다. 특히 여기서는 기존 CBA, TBA, CA 알고리즘 수행에 대해 5%의 오차를 허용하여 유연하게 탐지할 수 있도록 95% 임계치를 적용하였으며, 특히 결합기반 유출탐지 알고리즘인 CA의 결합율은 50%로 하여 실험하였다.

[표 1] 변조율 변화에 따른 알고리즘 수행결과 비교

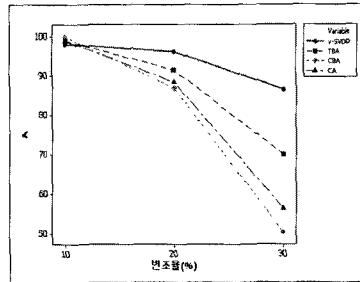
평가 척도 \ 알고리즘	A			SR		
	10%	20%	30%	10%	20%	30%
TBA	99.1	91.3	69.9	76.9	60.4	49.1
CBA	100.0	86.7	50.3	100.0	39.6	15.8
CA	99.1	88.2	56.3	76.9	46.2	26.1
v-SVDD	98.2	96.2	86.4	93.0	88.0	78.0

본 논문에서 제시한 v-SVDD 알고리즘의 수행결과와 CBA, TBA, CA 등 3가지 알고리즘의 수행결과를 변조율 변화에 따라 비교하면 [표 1]과 같다.

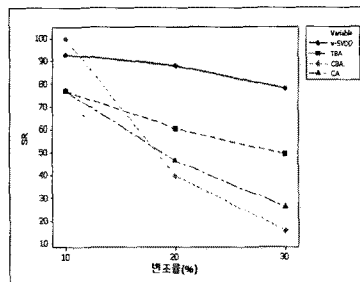
이 표를 통해서 변조율이 증가하더라도 v-SVDD 알고리즘은 두 가지 평가척도에 대해 모두 78% 이상의 정확도를 보임을 알 수 있다.

이 결과를 각각의 평가척도별로 도식화하면 (그림 1), (그림 2)와 같다.

(그림 1)은 각 알고리즘의 수행결과를 식 (12)의 평가척도인 탐지 정확도 A를 기반으로 비교한 것이며, (그림 2)는 각 알고리즘의 수행결과를 식 (13)의 평가척도인 변조유출탐지정확도 SR을 기반으로 비교한 것이다. 이 그림들에서도 알 수 있는 바와 같이 중요문서의 변조율이 적은 경우에 본 논문에서 제시한 v-SVDD 알고리즘은 기존 연구들과 비슷한 수준의 성능을 보이지만, 변조율이 증가함에 따라 기존 연구들에 비해 우수한 성능을 보임을 알 수 있다.



(그림 1) 평가척도 A에 대한 알고리즘 성능 비교



(그림 2) 평가척도 SR에 대한 알고리즘 성능 비교

## V. 결론 및 향후연구과제

본 논문은 최근 내부자에 의해 다양하게 변조되어 유출이 시도되는 중요문서 유출탐지문제를 해결하기

위한 방법으로 v-SVDD 알고리즘을 제시하였다. 본 논문에서 제시하는 v-SVDD 알고리즘은 문서분류의 이상탐지 기법으로 많은 연구가 수행되고 있는 SVDD를 기반으로 하고 있으며, 특히 가상 변조벡터를 생성하여 활용함으로써 기존의 SVDD를 중요문서 유출탐지에 직접적으로 적용하기 어려운 구조를 개선한 것이 특징이다. 또한 v-SVDD 알고리즘은 보호하고자 하는 중요문서 자체 뿐 아니라 다양한 변조 형태까지 고려하여 탐지함으로써 기존의 중요문서 유출탐지 기법보다 우수한 유출탐지 성능을 보여준다.

한편, v-SVDD 알고리즘은 변조유출탐지 성능에서 기존 연구보다 우수한 탐지 정확도를 보이지만, 중요문서의 변조율이 증가하는 경우 학습 시간이 증가하고 정확도가 감소하므로 이를 개선하기 위한 연구가 필요하다.

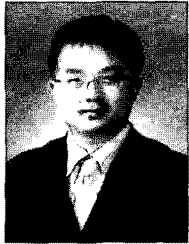
### 참 고 문 헌

- [1] 남기효, 강형석, 길지호, 김성인, "내부정보 유출방지(DLP) 기술동향," 정보통신산업진흥원, 주간 기술동향, 제1413호, pp. 1-9, 2009년 9월.
- [2] A. Per and C. Cristian, "Document-document similarity approaches and science mapping: Experimental comparison of five approaches," *Journal of Informetrics*, vol. 3, no. 1, pp. 49-63, Jan. 2009.
- [3] C. James and H. Ray, "Tightening the net: A review of current and next generation spam filtering tools," *Computers & Security*, vol. 25, no. 8, pp. 566-578, Nov. 2006.
- [4] F. Janssens, J. Leta, W. Glänzel, and B.D. Moore, "Towards mapping library and information science," *Information Processing & Management*, vol. 42, no. 6, pp. 1614-1642, Dec. 2006.
- [5] F. Janssens, W. Glänzel, and B.D. Moore, "Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analysis," In *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 360-369, Aug. 2007.
- [6] G. Vladutz and J. Cook, "Bibliographic coupling and subject relatedness," In *Proceedings of the 47th ASIS annual meeting*, pp. 204-207, Dec. 1984.
- [7] P. Glenisson, W. Glänzel, F. Janssens, and B.D. Moore, "Combining full text and bibliometric information in mapping scientific disciplines," *Information Processing & Management*, vol. 41, no. 6, pp. 1548-1572, Dec. 2005.
- [8] T. David and D. Robert, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45-66, Jan. 2004.
- [9] Z. Yong, C.Z. Xian, and L.K. Qiu, "Fuzzy multi-class classifier based on Support vector data description and improved PCM," *Expert System with Applications*, vol. 36, no. 5, pp. 8714-8718, July 2009.

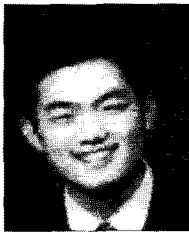
〈著者紹介〉



길 지 호 (Ji-Ho Ghil) 학생회원  
 2008년 2월: 고려대학교 산업시스템정보공학과 학사  
 2008년 3월~현재: 고려대학교 정보경영공학전문대학원 재학  
 <관심분야> 인공지능, 웹 보안, 콘텐츠 보호



남 기 효 (Kihyo Nam) 종신회원  
 1993년 2월: 고려대학교 산업공학과 학사  
 1995년 2월: 고려대학교 산업공학과 석사  
 1999년 2월: 고려대학교 산업공학과 박사  
 1999년~2002년: 고려대학교 정보통신기술공동연구소 선임연구원  
 2002년~2005년: 프롬투정보통신(주) 기술기획팀장  
 2005년~현재: (주)위너다임 이사  
 <관심분야> 개인정보보호, 웹 보안, 콘텐츠보호



강 형 석 (Hyung-Seok Kang) 학생회원  
 2007년 2월: 고려대학교 산업시스템정보공학과 학사  
 2007년 3월~현재: 고려대학교 정보경영공학전문대학원 재학  
 <관심분야> 인공지능, 웹 보안, 콘텐츠 보호



김 성 인 (Seong-in Kim) 정회원  
 1970년: 서울대학교 경제학과 졸업  
 1973년: 서울대학교 응용수학과 졸업  
 1975년: KAIST 산업공학과 석사  
 1979년: KAIST 산업공학과 박사  
 1979년~현재 : 고려대학교 정보 경영공학부 교수  
 <관심분야> 인공지능, 응용통계, 미술치료