

면 객체 매칭을 위한 판별모델의 성능 평가

Evaluation of Classifiers Performance for Areal Features Matching

김지영¹⁾ · 김정옥²⁾ · 유기윤³⁾ · 허 용⁴⁾
Kim, Jiyoung · Yu, Kiyun · Kim, Jung Ok · Huh, Yong

Abstract

In this paper, we proposed a good classifier to match different spatial data sets by applying evaluation of classifiers performance in data mining and biometrics. For this, we calculated distances between a pair of candidate features for matching criteria, and normalized the distances by Min-Max method and Tanh (TH) method. We defined classifiers that shape similarity is derived from fusion of these similarities by CRITIC method, Matcher Weighting method and Simple Sum (SS) method. As results of evaluation of classifiers performance by Precision-Recall (PR) curve and area under the PR curve (AUC-PR), we confirmed that value of AUC-PR in a classifier of TH normalization and SS method is 0.893 and the value is the highest. Therefore, to match different spatial data sets, we thought that it is appropriate to a classifier that distances of matching criteria are normalized by TH method and shape similarity is calculated by SS method.

Keywords : Matching, Normalization, Weight, Similarity, Precision-Recall Curve

초 록

데이터마이닝과 바이오인식 분야의 판별모델의 성능평가 방법을 이종의 공간 데이터 셋의 매칭에 적용함으로써 좋은 매칭결과를 보이는 판별모델을 도출하고자 한다. 이를 위하여 매칭 기준별 매칭 후보객체 쌍의 거리 값을 구하고, 이들 거리 값을 Min-Max 방법과 Tanh 방법으로 정규화하여 유사도를 산출한다. 산출된 유사도를 CRITIC 방법, Matcher Weighting 방법 그리고 Simple Sum 방법으로 결합하여 형상유사도를 도출하는 판별모델을 적용하였다. 각 판별모델을 PR곡선과 AUC-PR로 평가한 결과, Tanh 정규화와 Simple Sum 방법을 적용한 판별모델의 AUC-PR이 0.893으로 가장 높게 나타났다. 따라서 이종의 공간 데이터 셋의 매칭을 위해서는 Tanh 정규화를 이용하여 각 매칭기준별 유사도를 산출하고 Simple Sum 방법으로 형상유사도를 구하는 판별모델이 적합한 것으로 사료된다.

핵심어 : 매칭, 정규화, 가중치, 유사도, PR 곡선

1. 서론

데이터의 특징을 비교하여 동일한 현상 혹은 개체로부터 취득된 데이터 집합을 판별하는 것은 데이터마이닝 분야의 중요한 연구 주제이다. 일반적으로 분석자에 의하여 정의된 훈련 자료를 분석하여 판별모델을 정립하고 실험 자료에 적용

하여 그 성능을 평가한다(Han et al., 2011). 최근 매우 정확한 판별력이 요구되는 활용분야가 증가함에 따라 복수의 판별모델을 결합하는 연구가 수행되고 있다. 대표적으로 바이오인식 분야에서는 지문, 홍채, 안면 등의 정보를 종합하여 개개인을 판별할 수 있는 기법들이 연구되고 있다(Jain et al., 2005). 이를 위해서는 2개 이상의 생체 인식 시스템에서 도출된 매칭 값

1) 정회원 · 서울대학교 대학원 공과대학 건설환경공학부 박사과정(E-mail : soodaq@snu.ac.kr)
2) 정회원 · 서울대학교 공학연구소 선임연구원 공학박사(E-mail : geostar1@snu.ac.kr)
3) 정회원 · 서울대학교 공과대학 건설환경공학부 교수(E-mail : kiyun@snu.ac.kr)
4) 교신저자 · 정회원 · 서울대학교 공학연구소 선임연구원 공학박사(E-mail : huhyong78@gmail.com)

(matching score)을 정규화하는 과정(normalization), 정규화된 매칭 값을 단일 매칭 값으로 결합하는 과정(score fusion), 임계값 또는 분류기를 통하여 본인 여부를 결정하는 판별 과정을 거치게 된다. 이때, 정규화 방법과 결합 방법에 따라 본인 여부를 식별하는 즉, 판별모델의 성능이 결정된다(Snelick et al., 2005; Woo and Kim, 2006).

공간정보 분야에서도 이중의 공간 데이터 셋에서 동일한 면 객체를 탐지하기 위하여 다양한 기하학적 특징 값을 추출하고, 이들을 결합한 유사도를 이용하는 연구가 수행되고 있다(Fu and Wu, 2008; Huang et al., 2010; Samal et al., 2004; Wenjing et al., 2008). 이들 선행 연구에서도 데이터마이닝이나 바이오인식 분야와 유사한 과정으로 이중의 공간 데이터 셋의 매칭을 수행한다. 이중의 공간 데이터 셋을 중첩하여 교차하는 면 객체를 매칭 후보객체 쌍으로 추출한다. 이들 매칭 후보객체 쌍의 유사 정도를 측정하기 위하여 면 객체의 기하학적 성질이 반영된 매칭 기준으로부터 매칭 후보객체 쌍의 거리 값을 산출하고, 매칭 기준별로 산출된 거리 값은 범위가 상이하여 그 값의 최대값으로 정규화하여 유사도를 산출한다. 다음으로 각각의 유사도를 결합하여 통합 유사도, 즉 형상유사도를 생성하고, 일정 임계값 이상인 경우 매칭 후보객체 쌍은 서로 매칭으로 그렇지 않은 경우는 매칭이 아닌 것으로 판별된다. 즉 유사도를 산출하기 위하여 2개 이상의 매칭 기준별 거리 값을 정규화하는 과정, 유사도를 단일 척도인 형상유사도로 결합하는 과정, 임계값을 이용하여 매칭 여부를 판별하는 과정으로 이루어진다. 이중의 공간 데이터 셋의 매칭에서도 유사도를 산출하기 위한 정규화 방법과 유사도를 결합하는 방법에 따라 판별모델의 성능, 즉 매칭결과가 좌우될 수 있다.

따라서 본 연구에서는 이중의 공간 데이터 셋의 매칭 성능을 객관적이고 신뢰성 있게 평가하기 위하여 동일한 데이터 셋과 매칭 기준이 고려된 거리 값을 이용하여 상이한 정규화 방법과 결합 방법의 성능을 비교 분석하였다. 이를 통하여 면 객체를 관리하는 이중의 공간 데이터 셋의 매칭 성능을 향상시키기 위한 토대를 마련하고자 한다. 본 연구에서 대상으로 하는 이중의 공간 데이터 셋은 사용자의 요구가 많은 건물 정

보를 관리하는 수치지도 2.0과 도로명주소기본도이다.

2. 면 객체 매칭을 위한 판별모델의 성능 평가 방법

본 연구에서는 면 객체 매칭을 위한 판별모델의 성능 평가를 위하여 그림 1과 같은 방법을 적용하였다. 이중의 공간 데이터 셋의 기하학적 성질을 대표할 수 있는 매칭 기준을 선정하고, 매칭 기준별로 매칭 후보객체 쌍의 거리 값을 구한다. 거리 값에 정규화 방법을 적용하여 유사도를 산출하고, 이렇게 산출된 유사도를 결합 방법으로 정의되는 가중 선형 조합을 통하여 형상유사도를 구한다. 마지막으로 정보검색 분야에서 사용되는 지수(index)를 이용하여 어떤 판별모델이 이중의 면 객체 공간 데이터 셋에서 좋은 매칭결과를 보이는지를 평가한다.

2.1 매칭 기준별 거리 값 산출

본 연구에서 대상으로 하는 도로명주소기본도는 수치지도 2.0을 기반으로 제작이 되었으나 주소를 부여하는 본래의 목적과 각 지방자치단체의 담당자가 모화 입력하여 갱신하는 과정에서 위치정확도가 확보되지 않는 한계가 있다. 이와 같은 도로명주소기본도의 특성을 고려하여 이중의 공간 데이터 셋에서 중복되는 면 객체, 즉 매칭 후보객체 쌍(A, B) *i*번째에 해당하는 *m*번째 면 객체 A_m^i 와 *n*번째 면 객체 B_n^i 가 있다고 할 때 세 가지의 매칭 기준을 정의할 수 있다.

● 형상 지수(Shape Index, SI)

면 객체의 형상의 척도로 식 (1)로 산출되며, 면 객체의 형상이 원과 비슷할수록 그 값이 1이 된다. 각 매칭 후보객체 쌍의 형상 지수가 작을수록 해당 매칭 후보객체 쌍은 유사하다.

$$SI_{(A_m, B_n)^i} = \left| \frac{Perimeter(A_m^i)}{2 \times \sqrt{(\pi \times Area(A_m^i))}} - \frac{Perimeter(B_n^i)}{2 \times \sqrt{(\pi \times Area(B_n^i))}} \right| \quad (1)$$

● 형상 방향(Shape Direction, SD)

각 매칭 후보객체에서 너비가 동일한 최소사각형을 구하

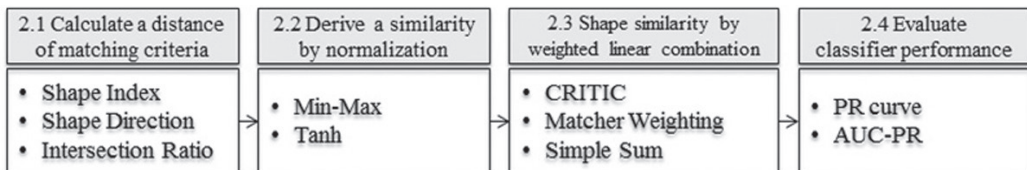


Figure 1. Methodology to evaluate classifiers performance for areal features matching

고, 이 최소사각형의 장축과 진복의 사이각($\theta_{A_m^i}, \theta_{B_n^i}$)을 해당 면 객체의 방향으로 정의한다. 식 (2)로 산출되며, 각 매칭 후보객체 쌍의 형상 방향 차가 작을수록 해당 매칭 후보객체 쌍은 유사하다.

$$SD_{(A_m, B_n)^i} = \left| \cos(\theta_{A_m^i} \times \frac{\pi}{180}) - \cos(\theta_{B_n^i} \times \frac{\pi}{180}) \right| \quad (2)$$

● 중복면적비(Intersection Ratio, IR)

식 (3)으로 산출되며, 매칭 후보객체 쌍이 서로 교차하는 면적이 클수록 중복면적비가 작아진다. 즉 매칭 후보객체 쌍의 중복면적비가 작을수록 해당 매칭 후보객체 쌍은 유사하다고 볼 수 있다.

$$IR_{(A_m, B_n)^i} = \frac{Area(A_m^i \cap B_n^i)}{Area(A_m^i \cup B_n^i)} \quad (3)$$

2.2 정규화를 통한 유사도 측정

매칭 기준으로부터 얻은 거리 값에서 유사도를 취득하기 위해서는 정규화 과정이 수행되어야 한다. Samal et al.(2004)은 매칭 후보객체 쌍(A, B)의 유사도 ($\Psi(A, B)$)를 거리 값 ($\Delta(A, B)$)의 1차 정규값을 1에서 차한 값으로 정의하였으며, 이때 정규화 지수(U)는 매칭 후보객체 쌍의 거리 값 중 최대값을 사용하였다(식 4).

$$\Psi(A, B) = 1 - \frac{\Delta(A, B)}{U} \quad (4)$$

본 연구에서는 식 (4)를 이용하여 유사도를 측정하기 위하여 정규화 지수에 가장 일반적으로 사용되고 있는 Min-Max(MM) 정규화와 Tanh(TH) 정규화를 적용하고자 한다(표 1). 여기서, 모든 매칭 후보객체 쌍(S)에서 매칭 기준으로 산출된 거리 값(s)의 정규화된 값(n)을 구한다고 가정한다. MM 정규화는 이상치에도 민감하고, 데이터의 분포가 정규화 후에 상이한 분포로 바뀔 수 있는 한계가 있다. 그러나 매칭 기준별 거리 값의 최대값(max(S))과 최소값(min(S))을 알고 있는 경우에 가장 적절한 방법으로 [0, 1]범위로 정규화 된다. 다음으로 강건(robustness)한 정규화 방법 중에서 사용자

가 설정하는 매개변수에 따라 정규화의 효율(efficiency)이 영향 받지 않는 TH 정규화를 선택하였다. TH 정규화는 거리 값이 (0, 1)범위로 정규화 된다.

2.3 가중 선형 조합을 통한 형상유사도

마지막으로 정규화 방법을 적용하여 거리 값을 유사도로 변환한 후, 유사도를 결합하여 단일 유사도인 형상유사도(Shape similarity)를 산출한다. 다중 인체 시스템이나 이종의 공간 데이터 셋의 융합 그리고 공간정보 분야에서는 매칭 기준의 결합 방법으로 가중 선형 조합이 일반적으로 사용되고 있다. 식 (5)와 같이 가중치(w_m)와 매칭 기준($m, m = 1, 2, \dots, M$)으로부터 산출된 거리 값을 정규화한 값, 즉 유사도(S_m)의 가중 선형 조합으로 형상유사도(S_{sim})가 산출된다. 이때, 가중치는 $0 \leq w_m \leq 1, \forall m, \sum_{m=1}^M w_m = 1$ 이다.

$$S_{sim} = w_1 S_1 + w_2 S_2 + \dots + w_m S_m \quad (5)$$

본 연구에서는 가중 선형 조합에 사용될 가중치를 산출하기 위하여 바이오인식 분야의 선형연구(Snelick et al., 2005; Zhang et al., 2011)에서 좋은 성능을 보인 Simple Sum(SS) 방법과 Matcher Weighting(MW) 방법, 그리고 결합된 유사도를 이용하여 도로망의 패턴을 분석하는데 적용된 CRITIC Importance Through Intercriteria correlation (CRITIC) 방법(Yang et al., 2010)을 비교하고자 한다(표 2). SS 방법은 가중 선형 조합에서 모든 가중치가 동일한 경우로 매칭 기준별 유사도의 합으로 산출된다. MW 방법은 동일 오류율(Equal Error Rate, EER)을 기반으로 각 매칭 기준별 가중치가 할당된다. EER($e_m, m = 1, 2, \dots, M$)은 [0, 1]범위의 임계값에 대하여 오매칭률과 오비매칭률이 일치하는 시점을 의미한다(Bengio et al., 2005). MW 방법으로 산출된 가중치는 대응되는 오류확률의 역수와 같으므로, 더 정확한 매칭 기준이 덜 정확한 매칭 기준보다 더 크게 부여된다. 마지막으로, Diakoulaki et al.(1995)이 제안한 CRITIC 방법은 각 기준의 표준편차뿐만 아니라 기준간의 상관관계를 고려함으로써 가중치를 결정하는 방법이다. CRITIC 방법은 1) 매개 변수를 정규화하고 2) 다변량에서 나타나는 충돌과 대조강도를 평가하기 위하여 매칭 변수간의 상관계수($r_{mn}, n = 1, 2, \dots, N$)를 도출한 후 3) 도출된 상관계수와 각 매칭 변수별 표준편차(σ_m)를 이용하여 각 매칭 변수가 포함하는 정보량(C_m)을 산출하고 4) 매칭 변수별 가중치를 구하는 단계로 계산된다. 이때, CRITIC 방법의 1) 매개 변수의 정규화 단계는 이미 매칭 기

Table 1. Normalization methods

| Method | Equation |
|--------|---|
| MM | $n = \frac{s - \min(S)}{\max(S) - \min(S)}$ |
| TH | $n = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s - \text{mean}(S))}{\text{std}(S)} \right) + 1 \right]$ |

준별 거리 값을 정규화 하여 산출된 유사도를 이용하기 때문에 그것으로 대체된다.

Table 2. Fusion algorithms applied to calculate a shape similarity

| Method | Equation |
|--------|---|
| SS | $w_m = \sum_{m=1}^M S_m$ |
| MW | $w_m = \frac{1}{e_m} \left(\frac{1}{\sum_{m=1}^M \frac{1}{e_m}} \right)$ |
| CRITIC | $w_m = \frac{C_m}{\sum_{m=1}^M C_m}, C_m = \sigma_m \times \sum_{m=1}^M (1 - r_{mn})$ |

2.4 판별모델의 성능 평가

일반적으로 이진 판별모델의 성능은 Receiver Operator Characteristic(ROC) 곡선을 이용하여 평가한다. 그러나 ROC 곡선의 경우 데이터 분포가 왼쪽이나 오른쪽으로 몰려 있어 한쪽으로 꼬리가 긴 분포에서는 판별모델이 과 최적화 되어 나타날 수 있다. 따라서 정보검색 분야에서는 한쪽으로 꼬리가 긴 분포의 경우는 ROC 곡선의 대안으로 Precision-Recall (PR) 곡선을 사용한다(Davis and Goadrich, 2006).

PR 곡선은 임계값이 달라짐에 따라 산출되는 재현율 (Recall)과 정확률(Precision)을 각각 x 축과 y 축으로 하는 그래프로 표현한 것이다. 이때, 실제와 예측 간의 관계를 나타내는 표 3과 같은 혼동행렬(confusion matrix)로 이진 클래스의 판별모델의 결과를 표현할 수 있으며, 정확률과 재현율은 식 (6)과 식 (7)로 구해진다. PR 곡선에서 판별모델의 정확도는 PR 곡선의 아래의 면적(area under the PR curve, AUC-PR)에 의해 측정될 수 있으며, 그 값이 클수록 판별모델의 성능이 좋다. 즉 그려진 PR 곡선이 우측 위쪽 모서리에 가까울수록 좀 더 정확한 판별모델이라고 해석할 수 있다.

Table 3. Confusion matrix

| Predicted \ Actual | Positive | Negative |
|--------------------|---------------------|--------------------|
| | Positive | True Positive (TP) |
| Negative | False Positive (FP) | True Negative (TN) |

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

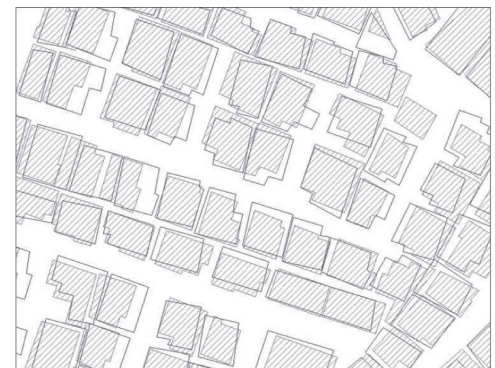
3. 실험 및 평가

3.1 실험 자료

실험에 사용된 이중의 공간 데이터 셋은 축척이 1:5,000인 수치지도2.0 37612006도엽과 2012년 9월에 갱신된 도로명주소기본도의 건물 레이어를 사용하였다(그림 2). 이들 데이터 셋에 실험을 진행하기 전에 파일형식은 SHAPE 파일, 좌표계는 세계측지계에 기준한 TM좌표로 통일하였으며, 기준점을 추출하여 어핀변환(affine transform)을 수행하여 계통오차를 제거하였다. 그림 2(b)와 같이 어핀변환 수행 후 실험 자료의 계통오차가 제거되었으나 여전히 면 객체의 중복만으로 매칭유무를 판별하는데 한계가 있다. 계통오차가 제거된 각각의 데이터 셋에서 동일한 영역을 실험 지역으로 선정하였는데, 수치지도 2.0은 701개의 면 객체, 도로명주소기본도는 757개의 면 객체가 포함되었다. 여기에서 매칭인 면 객체(매칭) 384쌍과 매칭이 아닌 면 객체(비 매칭) 58쌍을 수동으로 추출하였으며, 이를 판별모델의 성능 평가를 위한 실험에 사용하였다.



(a) Before affine transform



(b) After affine transform

Figure 2. Test data

3.2 실험 결과

수동으로 추출된 매칭과 비 매칭인 면 객체 쌍에 대하여 매칭 기준별 거리 값을 산출하고, MM 정규화와 TH 정규화를 적용하여 유사도를 산출하였다. 산출된 매칭 기준별 유사도를 가중 선형 조합하여 형상유사도를 구하기 위하여 표 4와 같은 가중치가 적용되었다. 이때, CRITIC 결합의 경우 MM 정규화와 TH 정규화에서 동일하게 산출되었다. 이는 본 연구에서 적용된 정규화가 원 데이터의 분포는 그대로 유지되고 값의 범위만 바뀌기 때문에 MM 정규화와 TH 정규화에서 동일한 가중치가 산출된 것으로 보인다.

Table 4. Weight applied in weighted linear combination

| Fusion Matching criteria | CRITIC | | MW | |
|-----------------------------|--------|-------|-------|-------|
| | MM | TH | MM | TH |
| SI | 0.302 | 0.302 | 0.352 | 0.333 |
| SD | 0.409 | 0.409 | 0.429 | 0.335 |
| IR | 0.289 | 0.289 | 0.219 | 0.332 |

결과적으로 각각의 정규화 방법과 결합 방법을 적용하여 산출된 형상유사도 분포는 그림 3과 같다. MM 정규화의 정

우 MW 결합은 비 매칭 객체 쌍에서 그 값이 [0, 1]범위에서 넓게 퍼져서 분포되었으며, CRITIC 결합과 SS 결합은 매칭 객체 쌍과 비 매칭 객체 쌍이 일정 값을 기준으로 밀집되어 있음을 알 수 있다. TH 정규화의 경우 결합 방법에 따라 분포의 차이가 크게 나타나지 않았다. 그러나 MM 정규화보다 일정 값을 기준으로 매칭 객체 쌍과 비 매칭 객체 쌍이 더 밀집되어 나타나는 것을 볼 수 있다.

3.3 판별모델의 성능 평가

이종의 면 객체 공간 데이터 셋의 매칭에 적용된 판별모델의 성능을 평가하기 위하여 그려진 PR 곡선과 AUC-PR(범례 내 수치)은 그림 4와 같다. 거의 비슷한 AUC-PR을 보이나, TH-MW 판별모델과 TH-SS 판별모델의 AUC-PR이 0.893으로 가장 높게 나타났다. 그러나 TH-MW 판별모델은 가중 선형 조합으로 형상유사도를 산출할 때 매칭 기준별 적용된 가중치가 표 4에서와 같이 0.333, 0.335, 0.332로 비슷하여, 결과적으로 TH-SS 판별모델과 동일한 가중치가 형상유사도를 구하는데 적용된 것으로 판단할 수 있다. 이에 이종의 면 객체 공간 데이터 셋의 매칭을 위해서는 TH-SS 판별모델이 가장 좋은 성능을 나타내는 것으로 사료된다.

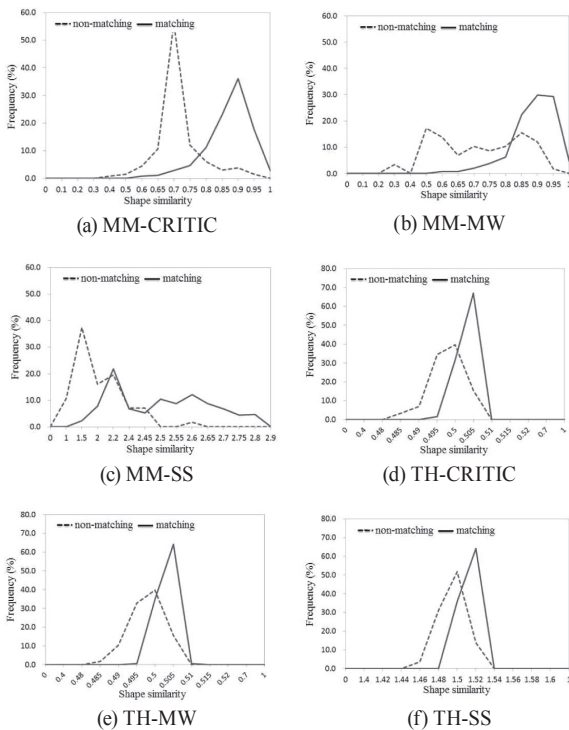


Figure 3. Histogram of a shape similarity in each classifier

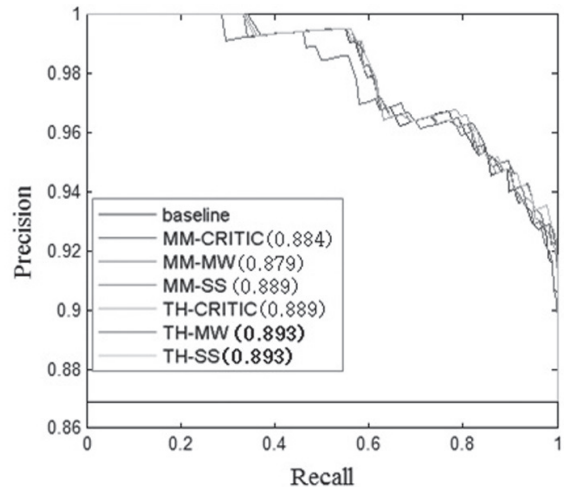


Figure 4. PR curve and AUC-PR in each classifier

따라서 본 연구에서 실험을 통하여 나타난 결과를 종합해 보면 다음의 판별모델이 이종의 면 객체 공간 데이터 셋에서 정확한 매칭 성능을 보인다:

- 매칭 기준별 거리 값을 이용한 유사도 산출을 위한 정규화 방법: TH 정규화

- 가중 선형 조합을 이용한 형상유사도 산출을 위한 결합 방법: SS 결합.

4. 결론

데이터마이닝과 바이오인식 분야에서 널리 이용되는 판별 모델의 성능 평가 방법을 이종의 공간 데이터 셋의 매칭에 적용하여 매칭 성능을 향상시키는 판별모델을 도출하고자 하였다. 이를 위하여 매칭 기준별로 거리 값을 정의하고, 정의된 거리 값에 MM 정규화와 TH 정규화를 적용하여 매칭 기준별 유사도를 산출하였다. 매칭 기준별 유사도를 CRITIC 결합, MW 결합, 그리고 SS 결합으로 계산된 가중치로 가중 선형 조합하여 형상유사도를 구하였다. 마지막으로 정보검색 분야에서 사용되는 PR 곡선과 AUC-PR을 이용하여 각 판별모델의 성능을 평가하였다. 그 결과 TH 정규화와 MW 결합이나 SS 결합을 이용한 판별모델의 AUC-PR이 0.893으로 가장 높게 나타났다. 그러나 MW 결합으로 산출된 가중치가 매칭 기준별로 거의 비슷하여, 결과적으로 SS 결합과 동일한 가중치가 형상유사도를 구하는데 적용되었다. 따라서 이종의 공간 데이터 셋의 매칭 시 TH 정규화와 SS 결합을 적용한 판별모델이 좋은 매칭결과를 보이는 것으로 판단된다.

본 연구를 통하여 면 객체를 관리하는 이종의 공간 데이터 셋의 매칭 성능을 향상시키기 위한 토대를 마련하였다는 데 그 의의가 있다. 그러나 보다 정확하고 객관적인 이종의 공간 데이터 셋의 판별모델을 개발하기 위해서는 다양한 이종의 공간 데이터 셋을 대상으로 지속적으로 연구가 진행되어야 할 것이다.

감사의 글

본 연구는 중소기업청에서 지원하는 2011년도 산학연공동 기술개발사업 (No.00045395)의 연구수행으로 인한 결과물임을 밝힙니다. 또한 본 연구는 서울대학교 공학연구소의 연구비 지원으로 수행되었습니다.

References

Bengio, S., Maréthoz, J. and Keller, M. (2005), The expected performance curve, *Proceedings of the ICML'05 workshop on ROC analysis in machine learning*, Germany, pp. 43–50.

Davis, J. and Goadrich, M. (2006), The relationship

between precision-recall and ROC curves, *Proceedings of the 23rd International Conference on Machine Learning*, USA, pp. 233–240.

Diakoulaki, D., Mavrotas, G. and Papayannakis, L. (1995), Determining objective weights in multiple criteria problems: the CRITIC method, *Computers & Operational Research*, Elsevier, Vol. 22, No. 7, pp. 763–770.

Fu, Z. and Wu, J. (2008), Entity matching in vector spatial data, *Proceedings of the XXIIth ISPRS Congress*, ISPRS, China, pp. 1467–1472.

Han, J., Kamber, M. and Pei, J. (2011), *Data Mining: Concepts and Techniques, Third Edition*, Morgan Kaufmann, USA, pp. 370.

Huang, L., Wang, S., Ye, Y., Wang, B. and Wu, L. (2010), Feature matching in cadastral map integration with a case study of Beijing, *Proceedings of 2010 18th International Conference on Geoinformatics*, IEEE, China, pp. 1–4.

Jain, A., Nandakumar, K. and Ross, A. (2005), Score normalization in multimodal biometric systems, *Pattern Recognition*, Elsevier, Vol. 38, No. 12, pp. 2270–2285.

Samal, A., Seth, S. and Cueto, K. (2004), A feature-based approach to conflation of geospatial sources, *International Journal of Geographical Information science*, Taylor & Francis, Vol. 18, No. 5, pp. 459–489.

Snelick, R., Uludag, U., Mink, A., Indovina, M. and Jain, A. (2005), Large scale evaluation of multimodal biometric authentication using state-of-the-art systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp. 450–455.

Wenjing, T., Yanling, H., Yuxin, Z. and Ning, L. (2008), Research on areal feature matching algorithm based on spatial similarity, *Proceedings of Control and Decision Conference (CCDC 2008)*, China, pp. 3326–3330.

Woo, N. and Kim, H. (2006), Performance evaluation of various normalization methods and score-level fusion algorithms for multiple-biometric system, *Journal of the Korea Institute of Information Security and Cryptology*, Vol.16, No. 3, pp. 37–46.

Yang, B., Luan, X. and Li, Q. (2010), An adaptive method for identifying the spatial patterns in road networks,

Computers, Environment and Urban Systems, Elsevier,
Vol. 34, No. 1, pp. 40–48.

Zhang, L., Zhang, L., Zhang, D. and Shu, H. (2011),
Ensemble of local and global information for finger-
knuckle-print recognition, *Pattern Recognition*,
Elsevier, Vol. 44, No. 9, pp. 1990–1998.

(접수일 2013. 01. 22, 심사일 2013. 02. 06, 심사완료일 2013. 02. 15)