

REPORT

유방암 진단 및 예후 분석

-

수업 : 2015-2 최적화 응용

담당 교수 : 최인찬 교수님

이름 : 2013170805 배지원

날짜 : 2015.12.07

-

1. 도입

1.1 유방암

유방암은 유방에 생긴 암세포로 이루어진 덩어리이며 일반적으로 유방의 유관과 소엽에서 발생한 암을 가리킨다. 여성의 사망원인 암 중에 2위를 차지한다. 유방은 선 조직, 유관, 지방 조직, 림프관, 혈관, 근막 등으로 이루어져 있다. 유방은 밖으로 돌출되어있으므로 암이 생겼을 경우 발견이 쉽지만 림프절이 매우 가깝게 분포되어 있어 암세포가 림프관을 타고 다른 장기로 전이 될 가능성이 매우 크다. 따라서 유방암을 조기에 발견하여 치료하거나 재발을 막는 것이 중요하다. 유방암에 영향을 주는 요인은 성별, 나이, 인구학적 특성, 유전적 요인, 호르몬 자극 등이다.

1.2. 유방암 진단 및 치료

유방암을 진단하는 방법은 유방 X선 촬영술, 디지털 유방 촬영술, 유방 초음파, 전산화 단층 촬영, 양성자 방출 단층촬영, 자기공명영상촬영이다. 조직 생체 검사를 하면 암을 정확히 진단할 수 있고 그 예는 미세 침 흡인 생체검사다. 가슴에 두꺼운 바늘을 넣으면 그 중간에 칼날이 지나가며 조직이 잘리는데 이것을 잘게 썰어 냉동한다. 암의 크기, 림프절 침범 여부, 전신 전이 여부 등에 따라서 병기를 매긴다. 종양의 크기, 림프절의 전이, 원격 전이 여부에 따라 병기를 나누고 이에 따라 치료를 결정한다. 유방암 병기는 0기부터 4기까지 나누어지며 치료 방법은 유방 절제술, 방사선 치료, 타목시펜 투여, 항암 호르몬요법, 항암 화학요법 등으로 다양하다.

1.3. 유방암 예후

전이는 림프절이 많은 부위에서 가장 잘 이루어지고 전이 호발 분위는 림프절(76%)>뼈(71%)>폐(69%)>간(65%) 순서이다. 예후 예측인자는 종양의 크기, 전이된 액와 림프절 수, 호르몬 수용체, 조직학적 소견, 나이다. 이중에서 가장 중요한 요인은 종양의 크기와 림프절 전이이다. 종양의 크기가 클수록, 전이된 림프절의 수가 많을수록 재발 위험이 크다.

1.4. 최적화를 이용한 유방암 진단 및 예후

병원에서 유방암을 확진하기 위해 미세 침 흡인 생체검사를 이용할 때는 의사의 주관과 경험을 통해 진단하였다. 논문 『Breast Cancer Diagnosis and Prognosis via Linear Programming』은 Xcyt 프로그램을 소개하여 컴퓨터가 의사 대신에 조직 분석을 하여 유방암 진단을 하는 기술 원리를 설명하였다. Xcyt 프로그램은 curve-fitting program을 이용하여 조직 사진에서 핵의 경계를 파악하여 30가지 변수 특성을 추출한다. 그리고 MSM(Multisurface Method)를 이용하여 분류 알고리즘을 생성한 후 조직 세포를 악성과 양성으로 분류한다. 569명의 환자를 대상으로 MSM 알고리즘을 검증하여 97.5%의 정확도를 얻었다. 유방암 예후 추정을 위해서 RSA(Recurrence Surface Approximation)을 이용하여 회귀 알고리즘을 생성한 후 조직세포의 재발 시기를 예측하였다.

2. 본론

『Breast Cancer Diagnosis and Prognosis via Linear Programming』에 소개된 MSM과 RSA가 중요한 이유는 최적화 기법을 이용하여 암의 영향 요인을 분석할 수 있기 때문이다.

분석한 자료는 암에 대한 이해를 높여주고 암을 정확히 진단하고 예후를 추정하는 데 도움이 될 것이다. 논문에 따르면 Xcyt 프로그램이 1994년 위스콘신 대학 병원에서 시범 사용되었고 진단 결과의 정확도가 암 전문센터의 정확도만큼 좋지만, 실제 의학 치료에 적용되기까지 현실적인 어려움이 많다고 한다.

하지만 이 프로그램이 환자와 의사 모두에게 편의를 준다고 생각하여 이 프로그램을 발전시킬 방법과 활용 방안에 대하여 생각해보았다. 우선 데이터를 더욱 정확하게 분석하기 위하여 MSM 변형, 인공신경망, CART, 선형회귀분석을 적용하였다. 그다음으로 분석 결과를 이용하여 환자를 진단하고 예후를 추정하는 사이트를 만들었다. 3년 전에 이탈리아의 학생이 미세 침 흡인 생체검사와 인공신경망 코딩을 이용하여 유방암 진단 클라우드 서비스를 개발한 것에 아이디어를 얻었다.

2.1. 데이터 분석

데이터 분석을 위하여 요인분석, 논문의 MSM, 데이터마이닝을 이용하였다.

2.1.0. 요인분석

여러 가지 최적화 기법을 적용하기에 앞서 논문에서 사용한 입력 변수들 간의 관계를 분석하였다. 분석을 위해 SPSS를 이용하였다. 요인분석 결과 2가지 요인으로 진단과 관련한 변수의 91%를 설명할 수 있다. 요인1은 핵의 크기와 관련한 변수(반지름 극값, 반지름 평균값, 면적 평균값, 면적 극값, 면적 표준편차, 둘레 극값, 둘레 평균값)이다. 요인2는 핵의 구조와 관련한 변수(촉촉한 정도의 표준편차, 프랙털 차원의 표준편차, 오목한 정도의 표준편차)이다. 즉, 핵의 크기와 구조의 다양성이 유방암 진단과 관련 있는 중요한 요인이다. 3가지 요인으로 예후의 90%를 설명할 수 있다. 요인1은 핵의 구조와 관련한 변수(프랙털 차원의 평균과 극값, 평탄도의 평균), 요인2는 질감과 관련한 변수(질감의 평균과 극값), 요인3은 평탄도의 표준편차이다. 즉, 핵의 구조, 질감, 평탄도가 유방암 재발과 관련 있는 중요한 요인이다.

2.1.1. 목적식 수정

기존 논문에서는 아래의 식을 이용하여 진단 결과를 분류하였다.

$\min_{w,\gamma,y,z}$	$\frac{e^Ty}{m} + \frac{e^Tz}{k}$	$\min_{\gamma,e}$	$\frac{e^Ty}{m} + \frac{e^Tz}{k}$
subject to	$Aw + y \geq e\gamma + e$ $Bw - z \leq e\gamma - e$ $y, z \geq 0$		$y = \max(0, e\gamma + e - x^Tw)$ $z = \max(0, x^Tw - e\gamma + e)$
		subject to	$e \geq \exp^{-8}$ $w, r \geq 0$

표 2 논문 모델

표 1 논문 모델을 변형한 모델

y는 악성인데 양성으로 구분된 환자의 위배 정도를 의미하고 z는 양성인데 악성으로 구분된 환자의 위배 정도를 의미한다. 목적식은 악성 위배 평균과 양성 위배 평균을 합한 값이다. 논문에 따르면 유방암 악성 여부에 영향을 주는 특징은 질감의 평균, 면적의 최댓값, 평탄도의 최댓값이고 분류 평면은 $x^Tw: 0.2322(\text{mean texture}) + 0.01117(\text{extreme area}) + 68.37(\text{extreme smoothness})$ 이다. 정확도(accuracy)는 0.975이다.

RMSE는 근사모델의 정확도 평가 방법 중에 가장 많이 쓰이는 것이어서 이 방법을 적용하여

목적식을 위배 정도의 제곱 평균을 최소화하도록 만들면 오차범위가 더 줄어들 것이라 가정하였다. 논문의 모델과 목적식을 변형한 모델을 비교하기 위하여 엑셀 해 찾기 기능을 이용하였다. 하지만 결정변수를 y, z, x, r 로 설정할 경우 문제의 크기가 방대하여 해 찾기 기능이 되지 않아서 위 모델을 변형하여 1) 논문 모델을 변형한 모델과 2) 변형한 모델에서 목적식을 수정한 모델의 두 가지 모델을 비교하였다.

y 와 z 가 위배 값을 의미하므로 $y = \max(0, e\gamma + e - x^T w)$, $z = \max(0, x^T w - e\gamma + e)$ 으로 두었다. 기존 논문의 결과로 나온 w 를 원래 알고 있는 변수로 설정하고 r, e 를 결정변수로 두었다. 이 모델은 비선형 계획법이다.

엑셀 해 찾기 결과, $e=0.001091$ 일 때 분리평면은 $x^T w : 0.2322(\text{mean texture}) + 0.01117(\text{extreme area}) + 68.37(\text{extreme smoothness}) = 20517.34$ 이다. 정확도(precision)는 0.9717이다.

$$\begin{array}{ll} \min_{r,e} & \frac{e^T(y)^2}{m} + \frac{e^T(z)^2}{k} \\ & y = \max(0, e\gamma + e - x^T w) \\ & z = \max(0, x^T w - e\gamma + e) \\ \text{subject to} & e \geq \exp^{-8} \\ & w, r \geq 0 \end{array}$$

표 3 변형한 모델에서 목적식을 수정한 모델

엑셀 해 찾기 결과 e 가 $4.6e^{-0.7}$ 일 때 $x^T w : 0.2322(\text{mean texture}) + 0.01117(\text{extreme area}) + 68.37(\text{extreme smoothness}) = 47216465.06$, 정확도(precision)는 0.981132이다.

3) 결론 : 두 가지 모델을 비교한 결과 목적식이 위배 정도의 평균 합에서 위배 정도의 제곱 평균 합으로 바꾸었을 때 정확도(precision)는 1%가 개선되었지만, 민감도(sensitivity)는 3%가 감소하였다. 하지만 변형하여 세운 모델이 허점이 많고 논문을 잘 반영하지 않아서 개선해야 할 부분이 많다.

4) 새롭게 변형한 모델에서 유방암을 진단하는 기준은

if $4.6e^{-0.7}\gamma + 4.6e^{-0.7} - (0.2322 * \text{mean texture} + 0.01117 * \text{extreme area} + 68.37 * \text{extreme smoothness}) > 0$, then diagnosed as malignant
otherwise, diagnosed as benign

5) 결과 표 정리

	논문 변형	목적식 변형
precision	0.9716981	0.9811321
sensitivity	0.9493088	0.9122807
correction	0.9691877	0.9439776
specificity	0.9829545	0.9882698

표 4 엑셀 해찾기 : 논문 변형, 목적식 변형 비교

2.1.2. 진단 : 인공신경망 분류 (ANN classification)

데이터를 training set : validation set = 6:4로 나눈 후 엑셀마이너(xlminer)로 인공신경망 분류를 시행하였다. 숨은 층(hidden layer)1에서는 1개의 노드를 사용하고 숨은 층2에서는 25개의 노드를 사용하였다. R을 사용했을 때는 숨은 층 2개와 1개의 노드를 사용하였다.

진단			
Performance	인공신경망 분류		분류 트리
	XLminer	R	XLminer
Success Class	Malignant (악성)		
Precision	0.987805	0.9766082	0.932432
Recall (Sensitivity)	0.952941	0.9902913	0.811765
Specificity	0.993007	0.9558824	0.965035

```

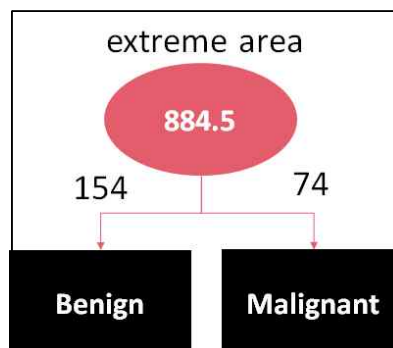
> cfmtest
      predicted
actual B  M
   B 102  3
   M  1 65
> summary(model.nnet)
a 30-2-1 network with 65 weights
options were - entropy fitting decay=5e-04
b->h1 i1->h1 i2->h1 i3->h1 i4->h1 i5->h1 i6->h1 i7->h1 i8->h1
-1.88 -0.84 -0.68 -0.87 -1.30  0.87 -5.46  3.52  3.24
i9->h1 i10->h1 i11->h1 i12->h1 i13->h1 i14->h1 i15->h1 i16->h1 i17->h1
 1.26  1.25  0.86 -1.21 -0.83  2.17  3.05 -4.53  0.62
i18->h1 i19->h1 i20->h1 i21->h1 i22->h1 i23->h1 i24->h1 i25->h1 i26->h1
-2.39 -0.54 -2.96  5.78  4.39  3.86  4.63 -0.96 -2.32
i27->h1 i28->h1 i29->h1 i30->h1
 1.54  5.33  0.28  6.48
b->h2 i1->h2 i2->h2 i3->h2 i4->h2 i5->h2 i6->h2 i7->h2 i8->h2
-2.68 -0.94  0.15 -0.72  0.15 -0.58  0.08  3.21  2.56
i9->h2 i10->h2 i11->h2 i12->h2 i13->h2 i14->h2 i15->h2 i16->h2 i17->h2
 1.29 -3.38  4.18 -1.29  2.49  3.50  0.18 -0.96 -0.68
i18->h2 i19->h2 i20->h2 i21->h2 i22->h2 i23->h2 i24->h2 i25->h2 i26->h2
-1.97  0.63 -4.91  2.38  3.77  1.34  2.96  0.14  2.68
i27->h2 i28->h2 i29->h2 i30->h2
 2.31  2.22  2.51 -1.14
b->o h1->o h2->o
-5.57 13.60 12.98
> accuracy
[1] 0.9766082
> sensitivity
[1] 0.9902913
> specificity
[1] 0.9558824

```

그림 2 진단 인공신경망 분류(R)

2.1.3. 진단 : 분류 트리 (classification tree)

best-pruned tree는 아래와 같다. 변수 1개(핵의 면적의 극값)를 이용하여 test data에 검증해보았을 때 93.23%의 정확도가 나왔다.



2.1.4. 예후 : 분류

재발 데이터를 재발 여부에 따라 분류하기 위하여 XLMiner로 인공신경망, 회귀 트리, 로지스틱 회귀를 사용하였다. 인공신경망의 정확도는 0.429, 분류 트리의 정확도는 0.315, 정의되지 않았고, 로지스틱 회귀의 정확도는 0.23가 나왔다. 전체 데이터 수가 194개로 많지 않고, 이 중에 Non-recurrent로 분류될 확률 78%, Recurrent로 분류될 확률이 22%로 재발 환자의 데이터가 너무 적어서 모델의 training이 제대로 되지 않기 때문이다.

```
> summary(model.nnet)
a 32-2-1 network with 69 weights
options were - entropy fitting decay=5e-04
b->h1 i1->h1 i2->h1 i3->h1 i4->h1 i5->h1 i6->h1 i7->h1 i8->h1
-2.92 -0.07 -3.12 0.20 -2.93 -1.05 8.38 -3.32 1.76
i9->h1 i10->h1 i11->h1 i12->h1 i13->h1 i14->h1 i15->h1 i16->h1 i17->h1
-0.66 -6.17 7.26 -1.43 3.13 -9.24 0.83 1.79 -2.98
i18->h1 i19->h1 i20->h1 i21->h1 i22->h1 i23->h1 i24->h1 i25->h1 i26->h1
-2.17 1.65 0.92 7.57 3.31 -3.54 -4.65 -0.06 -4.54
i27->h1 i28->h1 i29->h1 i30->h1 i31->h1 i32->h1
3.09 -0.53 -1.81 1.94 -0.71 -0.39
b->h2 i1->h2 i2->h2 i3->h2 i4->h2 i5->h2 i6->h2 i7->h2 i8->h2
2.49 -3.12 1.88 -3.67 -3.59 -1.43 0.08 -2.21 -1.82
i9->h2 i10->h2 i11->h2 i12->h2 i13->h2 i14->h2 i15->h2 i16->h2 i17->h2
5.58 4.28 -0.93 0.66 -0.14 1.55 0.10 -0.98 1.65
i18->h2 i19->h2 i20->h2 i21->h2 i22->h2 i23->h2 i24->h2 i25->h2 i26->h2
6.24 -1.17 -2.81 0.17 1.13 1.30 1.01 -8.33 2.76
i27->h2 i28->h2 i29->h2 i30->h2 i31->h2 i32->h2
-5.54 4.77 1.38 -7.25 -1.55 -12.12
b->o h1->o h2->o
-0.05 16.16 -15.73
> cfmtest
      predicted
actual N R
N 43 19
R 12 4
> accuracy
[1] 0.6025641
> sensitivity
[1] 0.7818182
> specificity
[1] 0.173913
```

그림 4 예후 인공신경망 분류(R)

2.1.5. 예후 : 인공신경망 회귀 (ANN Regression)

데이터를 training set : validation set = 6:4로 나눈 후 재발 환자와 비재발 환자 각각 데이터에 대하여 엑셀 마이네로 인공신경망 회귀를 시행하였다.

hidden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	bias
1	-0.05	0.13	-0.32	0.25	0.66	0.18	0.00	0.26	0.16	0.35	-0.16	0.02	0.09	0.00	-0.36	-0.15	-0.29	0.11	-0.24	-0.14	0.09	0.26	-0.17	0.16	0.28	-0.15	0.11
2	0.18	-0.04	0.08	0.02	-0.24	-0.22	-0.18	0.18	-0.30	-0.21	0.24	-0.12	-0.11	-0.30	-0.02	-0.11	-0.31	0.33	0.03	0.31	0.15	0.17	-0.03	-0.31	-0.34	0.28	0.10
3	-0.19	0.03	0.00	-0.17	0.43	-0.01	0.09	-0.29	0.34	-0.18	-0.11	0.54	0.11	0.12	0.04	0.27	0.27	-0.14	0.10	-0.26	-0.15	-0.23	0.28	0.05	-0.32	0.10	0.27
4	-0.04	-0.23	0.27	0.34	0.06	-0.47	0.37	-0.33	-0.03	0.28	0.16	0.33	-0.37	-0.28	0.10	-0.28	-0.15	-0.19	-0.34	-0.25	-0.13	-0.20	-0.35	-0.30	-0.33	0.09	0.08
5	0.23	-0.04	0.25	0.14	-0.30	0.15	0.18	0.21	-0.08	0.07	0.05	-0.23	-0.28	0.28	0.45	0.15	-0.01	0.49	0.35	0.45	0.00	0.36	0.35	0.03	-0.16	-0.06	0.10
6	0.02	0.13	0.30	0.30	-0.34	0.07	-0.18	0.01	-0.70	0.05	-0.35	-0.31	-0.45	-0.06	0.24	-0.44	-0.43	-0.42	0.13	-0.14	0.35	-0.03	0.28	0.10	0.56	0.28	0.16
7	0.20	0.32	0.33	-0.05	-0.25	0.06	-0.11	0.49	0.05	-0.06	-0.30	-0.38	-0.14	-0.21	0.11	-0.16	-0.21	0.15	0.29	0.22	0.07	0.26	-0.11	0.32	0.08	0.12	0.10
8	0.21	-0.21	0.13	0.13	0.08	-0.24	0.19	-0.24	0.05	0.22	-0.14	-0.06	0.08	-0.10	-0.28	0.02	0.03	0.12	0.16	0.12	0.30	0.21	-0.26	-0.12	0.15	-0.08	-0.27
9	0.21	0.17	-0.18	-0.11	0.09	0.15	0.19	0.08	-0.15	-0.26	0.24	0.15	0.20	-0.02	-0.17	-0.17	-0.02	0.05	-0.24	-0.13	0.24	0.07	0.06	-0.21	-0.25	-0.12	-0.28
output	1	2	3	4	5	6	7	8	9	bias																	
1	0.99	-0.21	0.75	1.02	-1.25	-1.42	-1.00	0.17	0.52	-0.89																	

그림 5 예후 재발 인공신경망 회귀 가중치

hidden	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	bias
1	0.12	0.52	-0.3	0.58	0.43	0.65	0.18	0.57	-0.2	0.42	-0.4	0.17	0.13	-0.1	-0.3	0.06	0.03	0.66	0.03	-0.4	-0	0.31	-0.3	0.03	0.1	-0.2	0.08
2	0.21	0.26	0.13	-0	-0.3	-0	0.13	0.24	-0.2	-0.4	0.29	0.36	-0.2	-0.6	0.2	0.17	0.2	0.94	0.44	0.43	-0.3	0.2	-0.4	-1.1	-0.9	0.37	0.22
3	-0.7	-0.3	-0.4	-0.7	0.77	0.15	-0.1	-0.2	0.44	0.03	-0	0.15	0.23	0.21	0.72	0.53	0.26	0.1	0.72	-0.4	-0.3	-1	0.31	0.01	-0.2	0.09	0.05
output	1	2	3	bias																							
1	-0.8	-1.4	1.21	-0.1																							

그림 6 예후 비재발 인공지능망 회귀 가중치

2.1.6. 다중 선형 회귀 분석 (MLR)

-재발 환자의 재발 시기(TTR) = $82.04068 + (-3.0808938) \times \text{mean radius}$

-비재발 환자의 무질병 기간(DFS) = $(-118.54294) \times \text{extreme concavity} + 1310.2919 \times \text{extreme fractal}$

-재발 환자의 MLR 설명력은 19%, 비재발 환자의 MLR 설명력은 88%이다.

XLminer 이용	재발 환자	비재발 환자
Total sum of squared errors	18405.122	28653.094
RMSE	20.002784	39.897852
Average Error	2.42511E-14	13.71272
R ²	0.2081538	0.8883325
Adjusted R ²	0.1901573	0.8797427

표 7 재발환자와 비재발 환자의 MLR 에러값

2.2. 유방암 진단 및 예후 관리 사이트

Youtube의 TED영상의 강연자 Brittany Wenger는 2012년에 유방암에 영향을 주는 조직을 정확하게 평가하는 웹 기반의 인공 뇌를 개발하였다. 유방암을 진단하는 글로벌 신경망 클라우드 서비스를 만들고 환자 데이터를 연동해서 의사들이 이용하도록 하였다. 이 서비스는 99.1%의 정확도로 종양을 찾아낸다. Brittany Wegner가 개발한 클라우드 서비스는 데이터를 실시간으로 연동하고 신경망을 계속 발전시킨다. 재작년에 760만 개에 가까운 데이터를 얻어서 신경망을 더 정교하게 만들었다. 신경망 클라우드 서비스는 사람이 직접 덩어리 두께, 세포 크기의 균일성, 세포 모양의 균일성, 응집성, 단일 상피 세포의 크기, 세포질이 없는 세포의 수, 염색질의 거친 정도, 정상적인 핵소체, 유사 분열의 9가지 특징을 직접 1부터 10까지의 척도로 점수를 매기면 결과가 나온다.

	Xcyt	글로벌 신경망 클라우드 서비스
검진 의학기술	미세 침 흡인 생검 (FNA)	
서비스 대상	의사	의사, 환자
평가 방법	컴퓨터가 자동으로 변수 값 측정	사람이 1-10에서 척도를 매김
변수	30개	9개
정확도	97.5%	99.1%
특징	소프트웨어	클라우드 소프트웨어
개발연도	1995	2012
상용화 여부	-	O
장점	진단여부, 재발시기 모두 예측	클라우드-정보 공유 및 신경망 발전

표 8 유방암 진단 소프트웨어 비교 (Xcyt & 글로벌 신경망 클라우드 서비스)

처음에 R 프로그램 코드를 만들면서 Xcyt와 글로벌 신경망 클라우드 서비스를 합쳐서 더 뛰어난 서비스를 만들고 싶었다. 유방암 진단과 재발 시기를 함께 예측하고 병원의 데이터와 연동해서 모델을 계속 발전시키고, 재발 시기에 맞춰서 환자에게 내원 알림 서비스를 해주는 클라우드 서비스를 개발하려 하였다. 하지만, 이러한 소프트웨어를 만드는데 어려움이 있었다.

첫 번째로, 유방암 진단과 재발 시기를 예측하는 방법은 인공지능망이 가장 뛰어났지만, 인공지능망은 일종의 블랙박스로 활성 함수를 알기 힘들어서 변수를 입력받아도 결과값을 알 수 없었다. 두 번째로, 현재 가지고 있는 데이터가 너무 적어서 재발 여부를 잘 예측하는 모델을 찾기 힘들었다. 병원의 데이터도 얻을 수 없었다. 세 번째로, 모델이 정확하지 않고, 데이터 수가 적고, 데이터를 연동할 수 없었기 때문에 환자에게 내원 알림 하는 서비스의 정확도가 떨어진다. 따라서 현재 가지고 있는 데이터로 높은 정확도를 가지는 서비스를 만드는 것보다 간단하고 직관적인 모델과 서비스를 만들기로 하였다.

유방암 진단 코드에 이용한 데이터 마이닝은 분류 트리이다. 93.23%의 정확도로 extreme area>884.5 이면 악성, extreme area<884.5 이면 양성이다. 유방암 예후 코드에 이용한 데이터 마이닝은 선형 회귀 분석이다. 재발 환자의 재발 시기를 예측할 때는 $TTR = 82.04068 + (-3.0808938) * \text{mean radius}$, 비재발 환자의 무질병 기간을 예측할 때는 $DFS = (-118.54294) * \text{extreme concavity} + 1310.2919 * \text{extreme fractal}$ 에 대입하였다.

자바 스크립트를 이용하여 사이트를 만들었고, 5개의 변수를 입력받은 후에 유방암을 진단하고, 재발 시기를 예측한다. 재발과 비재발을 분류하는 데이터마이닝 방법은 얻지 못했기 때문에 이 사이트에서 환자를 재발과 비재발로 분류하지 못 한다. 하지만, 유방암을 진단받은 환자가 재발 환자로 분류될 경우 재발 시기와 비재발 환자로 분류될 경우 무질병 생존 기간을 추정하여 알려준다.

그림 7 유방암 진단 화면

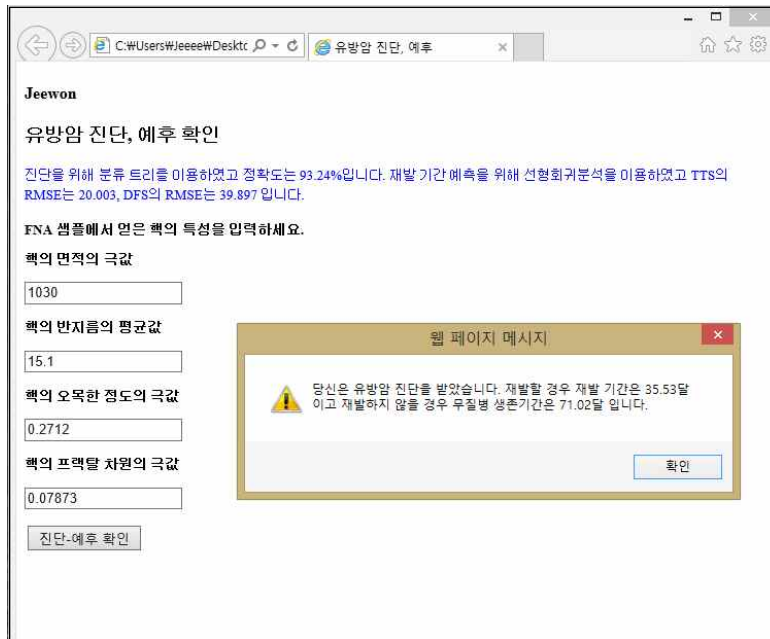


그림 8 유방암 재발/무질병 기간 예측 화면

```

a=82.84;
b=3.88;
c=118.5;
d=1310.3;

TTS=a*b*radius;
TTS=TTS.toFixed(2);
DFS=c*concavity*d*fractal;
DFS=DFS.toFixed(2);

/*종종 확인 */
if(earea < 884.5)
{
    alert ("유방암 진단을 받았습니다.");
}
else
{
    alert ("당신은 유방암 진단을 받았습니다. 재발할 경우 재발 기간은 "+ TTS + "달 이고 재발하지 않을 경우 무질병 생존기간은 "+ DFS + "달 입니다.");
}

```

그림 9 자바 스크립트 코딩(조건 함수)

3. 결론

진단 데이터 569개와 예후 데이터 196개를 다양한 방법의 데이터 마이닝을 이용하여 분석하였다. 유방암을 진단하고 예후를 추정하는 높은 정확도의 신경망을 만들고자 하였으나 데이터 수가 적어서 만들지 못했다. 따라서 대신 정확도가 90% 이상이고 간단한 서비스를 만들어 보았다. 최종 목표는 백만 개에 이르는 데이터를 얻어서 신경망을 더욱 정교하게 만드는 것이다. 신경망을 이용하여 진단, 예후 관리를 모두 하는 똑똑한 소프트웨어를 만든다면 많은 환자께 도움이 될 것으로 생각한다. 그리고 이러한 분석 방법이 다른 암, 질병에도 적용된다면 삶의 질을 높일 수 있을 것이다.

References

-Articles

O. L. Mangasarian, W. N. Street, and W. H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43(4):570-577, July-August 1995.

K. P. Bennett and O. L. Mangasarian. Neural network training via linear programming. In P. M. Pardalos, editor, *Advances in Optimization and Parallel Computing*, pages 56-67, Amsterdam, 1992. North Holland.

O. L. Mangasarian. Multi-surface method of pattern separation. *IEEE Transactions on Information Theory*, IT-14:801-807, 1968.

조경숙 외 4명, 『성인간호학』, 현문사, 2013.

-Sites

지금여기, 「R을 이용한 인공지능망 실습 2」, 다음 블로그 '열려있는 소프트웨어 이야기', 2013.11.14. (http://blog.daum.net/_blog/BlogTypeView.do?blogid=0m6cn&artid=3&categoryid=4®dt=20131114203226)

최윤선, 유방암의 치료 및 병기별 예후 (5년 생존률), 네이버 블로그, 2015.05.18.. (<http://blog.naver.com/younsun3/220363128281>)

<https://cloud4cancer.appspot.com/>

<https://www.youtube.com/watch?v=oJFdalfNy3Y&feature=youtu.be>

<http://eng.ajunews.com/view/20120726000603>

<http://pages.cs.wisc.edu/~olvi/uwmp/cancer.html>

<http://www.broadinstitute.org/mpg/xcyt/>

<http://dollar.biz.uiowa.edu/xcyt/>

<https://news.google.com/newspapers?nid=1368&dat=19940324&id=hpZQAAAIAAJ&sjid=EhMEAAAIAAJ&pg=7028,6499876&hl=ko>

-Data Sources

[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))

[https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Prognostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Prognostic))

-Image Sources

google search