이공학개인기초연구지원사업 최종(결과)보고서

											양건	A101
① 부처/	사업명(대)	기초연구사업					보안등급(보안, 일반)				일반	
② 사 역	업 명(중)	이공학개인기초역			구지원사업 공개가능여부			(공개	, 비공개)		공개	
③ 세부/	사업명(소)			기본연구 / 보호연구 / 지역대학우수과학자								
④ 과제/	성격(기초, 응	용, 개발)	기초	기초 ④-1 실용화 대상여부(실용화, 비실용화) 비실용화						용화		
⑤ 과 제 명		국 문			생존치유	모형어	서의	비모수	적추론	<u>-</u>		
		영 문		Nonparametric inference in a survival cure model								
⑥ 주관	연구기관				전	.북대학	학교					
⑦ 협동 ^c	연구기관											
② 즈리	연구책임자	성 명	है	o)	ド성 준		직급	(직위)		조교	수	
○ 十七、	선무색임사	소속부	-서	통:	계학과		전	공		통계	학	
		9	연구개발비	및 참여	여연구원수	(단위	: 천원	, M·Y)				
	정부출연금	기업체부딤		담금		정투 출연	 트의	¹ 외 상대국		합계 경		참여
년 도	(A)	현금 (C)	현물 (D)	1	소계 E=(C+D) 출연 (B				$C = (\Lambda + E)$		+E)	연구원수
1차년도	32,000	(0)	(D)	-	0					32,	000	3
2차년도	,				0					,	0	
3차년도					0						0	
4차년도					0						0	
5차년도					0						0	
합계	32,000		0	0	0		0		0	32,	000	3
10 총연구				20	18. 06. 01				(12개	월)		
	도협약연구기	간	기재하지 않음									
12 당해역	연도연구기간	7) 1+1 ?	2018. 06. 01 ~ 2019. 05. 31 (12개월)					2)			
③ 참여	기업	ਰੇ-	소기업수		대기업수	7		기타			계	0
		상대국연구기관수		<u></u>	상대국연구기		ગો ધો-મો		<u></u>	0 상대국연구책임자수		ŭ
④ 국제-	공동연구	경대국원구기원구		1	6 네 수 단 17		112 11 81		테크 단 [역 표시]			
 관계 규	·정과 모든 ㅈ	기시사항을	을 준수하면서	국가	연구개발사약	업에 U	다라 수	-행 중인] 연-	구개발과저]의	최종보
관계 규정과 모든 지시사항을 준수하면서 국가연구개발사업에 따라 수행 중인 연구개발과제의 최종보 고서를 붙임과 같이 제출 합니다.												
	2019년 6월 20일											
	주관연구책임자 : 양 성 준											
	주관연구기관장 : 김 동 원											

[※] 전자접수이므로 주관연구책임자 및 주관연구기관장 서명(인, 직인)은 생략

【주요항목 작성요령】

- ① 부처사업명(대), ② 사업명(중), 보안등급(일반), 공개가능여부(공개)는 수정하지 않음
- ④ 과제성격 및 ④-1 실용화 대상여부는 수정하지 않음
- ⑤ 과제명은 당초 연구과제명(과제명 변경을 재단에서 승인받은 경우는 승인된 과제명)을 기재함
- ⑥ 주관연구기관은 한국업적통합정보(KRI)에 입력된 기관명과 동일해야 하며 약어를 사용하지 않음 (서울대→서울대학교)
- ⑦ 협동연구기관은 본 사업과 관련이 없으므로 기재하지 않음
- ⑧ 연구책임자의 인적사항 기재
- ⑨ 정부출연금은 전체연구기간동안 기 지원받은 연도별 총 연구비(간접비 포함)를 기재 하며, 참여연구원은 연구책임자를 제외한 참여 연구인력(연구보조원 포함) 인원수를 기재함
- ⑩ 총연구기간은 연구시작일부터 연구종료일까지의 총연구기간 및 개월 수를 기재함
- ※ 최종보고서 작성 대상 기간은 총 연구기간임.(당해연도연구기간이 아님)
- ※ 셀 보호된 내용 수정방법: 해당셀 선택>마우스 우측버튼 클릭>표셀속성 선택>셀 선택>셀보호 해제

가

〈 연구결과 요약문 〉

양식A202

			0 711202				
연구개요	연구에서 주로 나타나는 /에 대한 과대추정 등 부정 러한 모형을 고려함으로써 있는데, 최근까지는 주로 연구는 치유모형에서 비모	생존자료의 특성 중 중도절 확한 결과를 주는 경우가 , 기존 생존모형의 한계를 모수적/준모수적 접근에 초 수적인 접근법을 도입하여	적 접근에 대한 연구이다. 의학 단만을 고려하는 모형은 생존율 있다. 치유개체를 추가적으로 고 극복할 수 있는 것으로 알려져 점이 많이 맞추어져 있었다. 본 모형에 유연성을 더하고자 하는 공변량의 효과를 비모수적으로				
연구 목표대비 연구결과	mixture cure model 하여료 분석을 통해 각 방법를 model 하에서 측정오차가 거한 비모수적 추론 방법을을 보였다. 또한, promot (varying coefficient)함수를 시하였다. 이에 대한 이론다. 당초 계획한 연구가 중과가 도출되지 못한 면이	론들의 특징을 관찰하였다. 존재하는 공변량의 효과에을 제안하고 기존 모수적 특ion time cure model 하여를 통하여 모형화하기 위한적 연구는 아직 진행 중이용료시점까지 완료되지 못하였으나 두 건의 석사학위는	적 방법론들을 소개하고 실증자 그리고, promotion time cure 대한 추론 시, Bootstrap에 근 방법에 비해 우수한 성질을 가짐 에서 공변량의 효과를 가변계수 방법론 제안 및 모의실험을 실 며, 향후 성과가 도출될 예정이 나였고, 논문출판 등의 정량적 성 근문이 작성되었고, 가변계수함수 록이 완료될 것으로 기대한다.				
연구개발결과의 중요성	치유모형은 생존자료의 특성 중 대표적인 중도절단 이외에 치유개체의 존재를 추가적을 고려하는 것으로, 전통적인 생존모형에 비해 더 정확한 분석/다층적인 해석을 가능케 해 준다. 이에, 실제 의학연구 등에 활용될 시 그 유용성이 매우 크다고 판단된다. 본 연구는 그 동안 연구결과가 제한적이었던 치유모형에 대한 비모수적 접근법에 대한 연구로 다른 연구자들의 관심 또한 불러일으킬 것으로 생각되며, 향후 더다양한 생존자료의 특성들을 고려한 진일보한 모형에 대한 연구로 확장될 수 있을 것으로 기대된다.						
중심어	치유모형 가변계수	생존자료	중도절단				

※ 표양식 변경 및 삭제불가능하며 이미지, 수식, 표의 삽입을 금지하고 특수문자 기호는 전각기호만을 이용하여 작성함 ※ 본 요약문은 정보제공용으로 활용되므로 핵심적인 내용을 중심으로 이해하기 쉽도록 기재하고 한 장 이내로 작성함

본문 작성 양식(본문 1~3번 항목까지 총 5페이지 내외로 작성)

가

※ 작성 시 작성요령은 제출 시 삭제하기 바람.

〈 목 차 〉

1.	연구개발과제의 개요
2.	연구수행내용 및 연구결과
	연구개발결과의 중요성 ······
	_, , ,
	참고문헌
5.	연구성과

<별첨> 주관연구기관의 자체평가 의견서

1. 연구개발과제의 개요

(1) 연구의 개요 및 목표

본 연구의 목표는 치유개체(cured subjects)들이 존재하는 생존자료(survival data)에서 공변량이 생존시간(survival time)에 미치는 영향을 규명하기 위한 효과적인 모형을 제안하는 것이다.

생존자료라 함은 어떤 사건이 일어나기까지의 시간(생존시간)을 기록한 자료로서 time-to-event data로 표현하기도 한다. 예컨대, 특정 질병에 걸린 후 환자들의 생존기간이나 완치 판정 후 재발까지의기간, 또는 실직 이후 새로운 직업을 얻을 때까지의기간 등을 기록한다면 이는 모두 생존자료의 범주에 포함된다고 할 수 있다. 전통적인 생존분석 방법론에서는 충분히 오랜 시간 관찰한다면 모든 개체가언젠가는 관심대상이 되는 사건을 경험한다고 가정하는 경우가 많았으나 이러한 가정이 실제 상황을 잘반영하지 못할 수 있음이 여러 연구자들에 의해 인지되었다. 예를 들어, 완치된 환자들 중 일부는 다시는 재발을 경험하지 않을 수 있으며, 실업자들 중 일부는 재취업을 포기할 수도 있다. 치유개체란 이렇듯 관심대상이 되는 사건을 경험하지 않게 되는 일부 개체들을 지칭하며 이를 함께 고려하여 생존자료를 분석하는 모형을 치유모형(cure model)이라 한다.

한편, 치유개체의 존재 하에서, 환자의 성별, 나이, 특정검사수치 등과 같이 개체의 특성을 나타내는 공변량들이 생존시간에 어떤 영향을 주는지를 규명하고자 하는 연구들이 있어 왔는데 mixture cure model(MCM) 과 promotion time cure model(PTCM)이 대표적인 두 가지 접근법이었다.

이 중 PTCM은 생존분석에서 매우 유용한 Cox의 비례위험모형(proportional hazard model)을 치유개체가 존재하는 경우로 확장한 것으로 볼 수 있으며 조건부생존함수를 다음과 같이 가정하는 것이다.

$$S(t|x) = \exp(-\theta(x)F(t))$$

여기서 $F(\cdot)$ 는 적절한 baseline 분포함수이며 $\theta(x)$ 가 공변량이 조건부생존함수에 주는 영향을 반영하

가

게 되는 함수이다. $\theta(x)$ 를 어떻게 선택하느냐에 따라 다양한 추론이 가능해지게 되는데 간단하게는 $\theta(x) = \exp(x^\top \beta)$ 와 같이 둘 수 있으며 이는 Cox의 비례위험모형과 같은 꼴을 가지게 된다.

본 과제는 치유모형에서의 비모수적 접근법에 대한 연구를 목표로 하였으며 특히, PTCM에 기반을 둔 모형에서 $\theta(x)$ 의 선택을 모형에 유연성을 더하고 자료에 존재할 수 있는 복잡한 특성을 더 잘 설명해 낼 수 있는 방식으로 하고자 하였다. 또 다른 목표는 본 연구가 생존자료에 존재할 수 있는 다양한 특성들을 반영하는 진일보한 생존모형 연구를 위한 발판이 되도록 하는 것이었다.

PTCM은 1990년대 중반 이후 (Yakovlev & Tsodikov (1996)) 활발하게 연구되기 시작하였으며 Cox의 비례위험모형과의 연관성으로 인해 많은 관심을 끌고 있다. 그러나 최근까지도 공변량의 효과를 모수적으로 모형화하는 연구가 대부분이었다. 전술한 것처럼

$$S(t|x) = \exp(-\theta(x)F(t))$$
, $\theta(x) = \exp(x^{\top}\beta)$

와 같은 모형을 고려하는 것이 대표적인 예다. 최근 미국의 한 연구진에서 spline방식을 이용하여 공변 량의 효과를 비모수적으로 추정하는 방법이 제안된 것이 현재까지 알려져 있는 유일한 비모수적 접근법 이다 (Chen & Du (2018)). 본 연구에서는 공변량의 효과 $\theta(x)$ 를 가변계수항(varying coefficient term)을 이용하여 다음과 같이 모형화하고자 하였다.

$$S(t|x) = \exp(-\theta(x)F(t))$$
, $\theta(x) = \exp(x^{\top}\beta(u))$.

여기서 $\beta(\cdot)$ 는 공변량 x에 대한 계수함수이며, 이것이 또 다른 변수 u에 대한 함수로 표현된다고 가 정하여 모형에 유연성을 더하고자 하는 시도이다.

공변량의 계수를 상수함수가 아니라 다른 변수의 함수로 두는 것은 가변계수모형(varying coefficient model)이라 하며 회귀모형(regression model)하에서 최근 20여 년간 매우 활발하게 연구되어 왔다. 가변계수모형은 잘 알려진 선형모형과 그 기본적인 형태는 유사하면서도 비모수함수를 포함함으로써 모형의 유연성을 크게 증대시켜, 데이터의 동적특성(dynamic feature)을 검출해 내는 데에 매우 효과적인 것으로 알려져 있다. 본 연구는 가변계수모형의 아이디어를 치유모형에 적용한 것으로 볼 수 있으며, 회귀모형에서와 비슷한 효과를 기대할 수 있을 것으로 판단된다.

(2) 연구의 필요성

생존자료라 함은 어떤 사건이 일어나기까지의 시간(생존시간)을 기록한 자료로서 주로 의학연구 분야에서 환자의 수명이나 질병의 재발기간 등을 나타내는 경우가 많지만, 구직자의 실업기간이나 제품의수명과 같이 경제학, 공학 등 다른 분야에서도 빈번하게 나타나는 형태의 자료이다. 이에 대해 생존시간의 분포를 추정하거나 환자의 나이나 혈압 등의 공변량이 생존시간에 어떻게 영향을 주는지를 규명하는 것은 생존분석 분야의 매우 중요한 이슈 중 하나이다.

(가) 보다 정확한 생존율 추정

환자들의 생존시간을 측정하는 생존자료의 연구시, 짧은 연구기간 혹은 발달된 치료법 등으로 인하여 많은 환자들이 사망하지 않은 채로 연구의 종료를 맞게 될 가능성이 있다. 이 경우, 과다한 중도절단비율 등의 문제로 인해 (조건부)생존함수 등에 대한 정확한 추정이 어려운 경우가 많다. 이 때, 치유된 개

가

체가 일부 존재한다고 가정하는 치유모형을 고려한다면, 더 정확한 추론 및 그에 따른 합리적인 결론도출이 가능하게 된다. 대표적으로 치유모형은 환자의 생존율에 대한 과대추정(overestimation)을 피할 수 있게 해 주는 것으로 알려져 있다. 이렇듯 치유개체를 고려하지 않은 고전적인 방법론으로는 부정확하게 (조건부)생존율을 추정하게 되며 이는 전체 중 치유된 개체의 비율이 크면 클수록 더 심해질 것이다. 따라서 치유모형은 고전적인 생존모형의 한계를 극복할 수 있는 훌륭한 대안 중 하나로 볼 수 있어 연구의 필요성이 크다 하겠다.

(나) 모형의 유연성

본 연구에서 제안하는 방법은 공변량 효과에 비모수함수(nonparametric function)를 포함하고 있다. 이는 모형을 훨씬 유연하게 만들어 주어 설명력을 증대시킬 것이며, 기존 모수적 방법론으로는 검출해 낼 수 없는 데이터의 고유한 특성을 찾아낼 수 있다는 장점이 있다. 또한, 비모수적 접근법은 데이터 탐색단계에서도 유용한 것으로 알려져 있다.

(다) 향후 연구의 확장성

전술하였듯이 중도절단만을 고려하는 생존모형으로는 정확한 분석 및 추론이 어려울 수 있고, 각종 치료법의 고도화 등으로 생존시간이 길어지는 상황 하에서는 이런 현상이 더 심화된다고 할 수 있다. 이에 따라, 중도절단 뿐 아니라 다른 다양한 생존자료의 고유한 특성들을 모형 구축과 추론시에 고려하는 방법론들의 개발 및 연구에 많은 관심이 집중되고 있다. 본 연구는 그 중 치유개체의 존재성을 고려하는 모형에 대한 연구이며, 이러한 치유모형에 대한 연구는 그 유용성에 비해 관련연구가 양적으로 충분하지 않아 연구 가능성이 많이 열려 있는 상황이며 특히 비모수적 접근법은 더욱 그러하다. 국내에도 아직 활발한 연구가 이루어지지는 않고 있는 것으로 파악하였다. (Kim (2014) & Choi et. al. (2018)). 본 과제를 바탕으로 하여 국내에 치유모형을 널리 소개하고, 뿐만 아니라 향후 더 다양하고 복잡한 특성을 가지는 (혹은 고려한) 생존모형의 연구로 나아가고자 한다.

2. 연구수행내용 및 연구결과

전술하였듯이 치유모형에는 MCM과 PTCM 두 접근법이 크게 존재한다. 본 과제에서는 두 모형에서의 모수적인 접근과 비모수적인 접근법을 비교하여 보았다.

(1) MCM에서 모수적/준모수적/비모수적 방법론 비교연구

MCM은 조건부 생존함수를

$$S(t|x) = p(x) + [1 - p(x)]S_0(t|x)$$

과 같이 모형화하는 것이다. 여기서 p(x)는 치유율을 나타내며 보통 로지스틱 모형 등을 이용하여 추정하는 것이 일반적이다. 한편, $S_0(t|x)$ 는 치유되지 않은 개체들의 조건부 생존함수로, 이를 어떻게 모형 화할지에 대한 연구가 활발히 이루어지고 있다. 본 연구에서는 특정분포족을 가정하는 모수적인 방법, Cox모형에 의해서 추정하는 준모수적인 방법(Peng (2003)), Beran 추정량(López-Cheda et. al. (2017))에 의해서 추정하는 비모수적인 방법 및 각 방법론을 적용할 수 있는 $R(\xi)$ 계 소프트웨어) package들을 소개하고, 세 접근법을 실제 자료분석에 적용하여 그 특성 및 차이를 관측하였다. 본 연구는 과제에 소속된 연구원의 석사학위논문으로 제출되었다.

(2) 측정오차가 있는 PTCM에서의 비모수적 추론

종종 공변량이 측정오차와 함께 관측되는 경우가 있다. 이 경우, 이를 무시하면 추정효율이 저하됨이 잘 알려져 있다. 본 연구에서는 측정오차가 존재할 때 공변량의 계수(β)에 대한 신뢰구간을 비모수적으로 구성하는 방법에 대해 고려하였다. 공변량이 X이고 실제로는 측정오차 U와 함께 W=X+U의 형태로 관측된다고 가정하자. (Ma and Yin (2008)) 에서는 U의 분포가 정규분포라고 가정하고 이에 근거하여 β 의 추정량의 점근적 분포를 유도후 이를 이용하여 신뢰구간을 제시하였다. 만약, 측정오차에 대한 정규성 가정이 어긋난다면 이는 부정확한 추론이 될 가능성이 있다. 본 연구에서는 최근 많이 활용되고 있는 Bootstrap 방법을 이용하여 β 에 대한 신뢰구간을 구축할 것을 제안하였고, 다양한 측정오차의 분포 하에서 모의실험을 실시하여 정규성을 가정하는 것보다 안정적인 추론이 가능해짐을 확인하였다. 본 연구 또한 과제 소속 연구원의 석사학위논문으로 제출되었다.

(3) PTCM에서 공변량의 계수를 가변계수함수로 모형화하는 연구

추정방법론의 제안과 이론적 연구/모의실험을 포함하는 연구를 계획하였으나, 현재 추정방법론 및 모의실험 연구가 진행 중이다. 가변계수함수에 국소선형모형화를 실시하고 kernel 방법에 의한 추정을 시도하였다. 간단한 모의실험 결과 모수적인 방법론에 비해 유연한 결과를 준다는 사실을 일부 확인하였고, kernel 방법 적용을 위한 평활모수의 선택 방법론에 초점을 맞추어 연구를 진행중이다. 이론적인 부분에 대한 연구는 향후 보완 예정이다.

본 연구자는 생존모형에 대한 연구 및 연구과제 수행 경험자이다. 하지만, 치유모형에 대한 연구는 처음 시도했던 것으로 다소 도전적인 선택이었으며, 본 주제의 연구를 위해 전술한 (1),(2)등의 연구 등을 통하여 치유모형 방법론 및 비모수적 접근법 등에 대해 탐구하고, 모의실험 등을 통해 치유모형에서 비모수적 접근법들이 어떠한 특성을 가지는지 파악하고자 하였다. 또한, 연구의 방향 설정 및 이론적 접근 등을 위해 생존모형 관련 권위자인 해외 전문가를 방문하여 공동연구를 진행하기도 하였다 (2019.04.14.~2019.04.19. 벨기에 카톨릭루뱅대학 방문).

현재 가시적인 성과가 미흡한 면이 있으나 (1),(2)등의 과정을 거치며 두 건의 석사학위논문이 작성되었고 현재 (3)의 연구가 계속 진행 중이며 본 과제의 연구기간이 1년으로 단년 과제였음을 고려할때, 추가적인 성과가 조만간 도출될 것으로 기대한다

3. 연구개발결과의 중요성

(1) 연구결과 자체의 의미

치유모형은 최근 활발하게 연구되고는 있지만 실제 현장에서 사용되는 빈도는 상대적으로 낮으며 많이 알려져 있지 않다. 본 연구과제에서 제안하는 모형은 그 유연성이 매우 커 복잡한 자료의 특성 추출에 유용하며, PTCM은 Cox 모형과 직접적으로 연관되어 있기 때문에 실제 현장의 분석가들에게도 크게 낯설지 않아 그 적용에 큰 장벽이 없으리라고 판단되며 매력적인 대안 중 하나가 될 수 있을 것으로 생각된다. 본 연구과제 종료 이후 본 연구자와 협업하고 있는 의학연구단 등을 중심으로 치유모형의 적용가능성을 적극적으로 타진해 볼 계획이다.

가 ・

가

PTCM에서 비모수적 접근법은 거의 연구된 바가 없어 본 연구는 국제적으로도 경쟁력이 있으며 본 연구를 통해 국내 치유모형의 연구 또한 더욱 활성화되기를 기대한다.

(2) 후속연구로의 연결

처음 본 연구과제 제안 시, 본 연구를 통하여 치유모형에 대한 경험 및 노하우를 축적하여 이후 후속 연구로써 다년과제의 제안을 염두에 두었고, 과제제안서에 명시한 바 있다. 본 연구자는 진일보한 생존 모형에 대한 다양한 추론을 목표로 하는 연구과제를 2018년 하반기에 제안하였고, 최종 선정되어 현재 연구를 진행 중이다.

4. 참고문헌

- [1] Chen, T., & Du, P. (2018). Promotion time cure rate model with nonparametric form of covariate effects. Statistics in medicine, 37, 1625-1635.
- [2] Choi, S., Zhu, L., & Huang, X. (2018). Semiparametric accelerated failure time cure rate mixture models with competing risks. Statistics in medicine, 37(1), 48-59.
- [3] Kim, Y. J. (2014). Cure Rate Model with Clustered Interval Censored Data. Korean Journal of Applied Statistics, 27(1), 21-30.
- [4] López-Cheda, A., Cao, R., Jácome, M. A. and Van Keilegom, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. Computational Statistics and Data Analysis. 105, 144-165.
- [5] Ma, Y. and Yin. G. (2008). Cure Rate Model with Mismeasured Covariates under Transformation. Journal of the American Statistical Association. 103, 743-756.
- [6] Peng, Y. (2003), Fitting semiparametric cure models. Computational Statistics and Data Analysis. 41, 481-490.
- [7] Yakovlev, A. Y., & Tsodikov, A. D. (1996). Stochastic models of tumor latency and their biostatistical applications. World Scientific.

5. 연구성과

학위배출인력 성과정보										
과제번호	학위취득연월	학위구분		지근						
			성명	성별	대학	학과	- 진로			
2018R1D1A 1B07051215	201902	석사	이종성	남성	한국외국어대학교	통계학과	취업(박사 후연구원 포함)			

가 가 .

국내외 과학자교류 성과정보											
과제번호	파견유치	인력교 류구 분		파	파련국가	파견유치					
파제인호	기간		성명	소속기관	직위	최종학위	전공	유치자국적	목적		
2018R1D1A 1B07051215	20190414 ~2019042 0	해외파견	양성준	전북대학교	조교수	박사	통계학	벨기에	공동연구		

후속연구추진 성과정보											
과제번호	연구기간	시업명	연구과제명	선행과제와의 관	과제성격	수행구분	과제책임자				
기재근포				계	7/1107		성명	소속기관			
2018R1D1A 1B0705121 5	201903~ 202402	신진연구자지원 사업	복잡한 생존 자료의 비모수 적/준모수적 모형화	치유모형 연구를 비탕으로 더 다양 하고 복잡한 특성 을 고려하는 생존 모형에 대한 확 장 연구 수행	기초	과제책임자 수행	앙성준	전북대학 교			