

KAKAO

Vo1.05

AI

2017.07

AI CODE

```
func [this *Model] GetEvaluation(sample *Sample) (*Evaluation, float64) {
    score, _ := this._getScore(sample.FeatureKeys)
    loss := this._getLoss(score, sample)
    return &Evaluation{Imps:uint64(sample.Imps), Clicks:uint64(sample.Clicks),
        ScoreSum:float64(sample.Imps) * score, LossSum:loss}, score
}

func [this *Model] GetGradients(sample *Sample) ([]Gradient, float64) {
    score, coefs := this._getScore(sample.FeatureKeys)
    return this._getGradients(sample.Imps, sample.Clicks, sample.FeatureKeys, coefs, score)
}

func [this *Model] UpdateGradient(g *Gradient, q float64, t int64) {
    coef, ok := this.Coefs[g._featureKey]
    if !ok {
        coef = &Coef()
        this.Coefs[g._featureKey] = coef
    }
    coef.Add(g)
    if math.Abs(coef.Z) >= this.Param.L1 {
        coef.W = this._getCoeffW(this.Param, coef)
    } else {
        coef.W = 0.
    }
    if coef.C > math.MaxInt64 {
        coef.C = math.MaxInt64
    }
}
```



KAKAO AI REPORT

Vo1.05

발행일 | 2017년 7월 28일

발행처 | (주)카카오

발행인 | 카카오 정책지원팀

편집 | 김대원, 김명수, 정수현

디자인 | 허진아

메일 | kakaoireport@kakaocorp.com

COVER

카카오 AI 리포트의 표지에선 AI와 관련된 의미 있는 코드들을 매월 소개할 예정입니다.

Vol.05 코드 | 박연진 jaidyn.park@kakaocorp.com

카카오의 서비스에 접속한 사용자의 니즈에 가장 부합하는 광고를 선택, 노출하기 위해서 사용자, 광고 소재 및 매체의 정보를 이용해서 광고 소재를 추천하는 알고리즘입니다.

contents

preface 02

industry

AI와 의료

민현석 | 내가 의료 AI를 선택한 이유 06

안성민 | 데이터 기반 정밀 의료와 AI 14

황상흠 | 딥러닝 기반 의료영상 기술의 진화 18

김남국 | 의료와 AI 신기술의 융합: 과제와 전망 24

정규환 | AI 의료영상 기술 활용 사례 28

양광모 | AI 의료, 이제 윤리를 고민하다 34

review

AI의 진화

정부환 | 더욱 똑똑해진 AI 광고 알고리즘 AI in Kakao 40

이원형 | AI는 인간과 같은 감정을 가지게 될까? 50

information

하반기 주요 AI 컨퍼런스 소개 54

closing

카카오 AI 리포트 5호를 내며

영화 ‘스타워즈 에피소드2: 클론의 습격(Attack of the Clones)(2002)’ 편에서는 의료 로봇이 아마달라 여왕(나탈리 포트만 역)의 쌍둥이 남매 출산을 돕는 장면이 나옵니다. 인공지능 로봇이 새로운 생명의 탄생을 돕고 산모를 돌보는 섬세한 일까지 해내는 모습은 신선한 충격으로 다가왔습니다. AI 기술이 의료 분야에 접목되는 사례가 늘어나면서 이제 ‘AI 의료’는 영화 속 상상이 아닌 현실로 실현되고 있습니다. 실제로 AI 기반 진단 장비를 의료 현장에서 활용하는 병원들이 속속 늘어나면서 과연 의료 분야에서 AI가 어떻게 활용되고 있을까라는 궁금점은 더욱 커져가는 것 같습니다.

카카오 AI 리포트 제5호는 ‘AI와 의료’ 기획 특집을 통해 의료 분야에서 AI 기술을 연구하고 계신 전문가 여섯 분의 글을 실었습니다. AI 기반 영상분석 기술 전문가인 민현석 박사는 본인이 의료 AI 분야를 선택한 이유와 의료 현장에 AI 기술을 접목하며 겪은 생생한 경험담을 들려주십니다. 안성민 가천대 길병원 교수는 ‘데이터 기반 정밀의료와 인공지능’을 주제로 인공지능 시대에 데이터 기반 의학이 왜 더욱 중요해지는가를 설명해주십니다. 여러 의료 분야 중에서도 AI 기반 기술이 상업적으로 가장 활발하게 개발되고, 적용되는 곳은 영상진단 분야입니다. 김남국 서울아산병원 융합의학과 교수, 황상흠 (주)루닛 수석엔지니어, 정규환 (주)뷰노 CTO는 각각 ‘딥러닝 기반의 의료영상 기술’, ‘의료와 신기술 융합을 위한 과제와 전망’, ‘AI 의료영상 기술의 활용 사례’ 등을 주제로 AI 의료영상 기술의 과거와 현재, 미래를 그려주셨습니다. 끝으로, 양광모 삼성서울병원 건강의학센터 교수는 AI 의료 기술이 발전할수록 커지는 의료윤리 딜레마 문제들을 짚어주셨습니다.

AI 기반 추천 기술은 이용자가 원하는 콘텐츠나 서비스를 찾아주는 서비스뿐만 아니라 이용자들에게 가장 도움이 될 광고를 선택해주는 광고랭킹 알고리즘에도 적용됩니다. 카카오 AI 부문의 정부환 님이 AI 기술이 접목되면서 갈수록 더욱 똑똑해지는 AI 광고 알고리즘을 상세하게 설명해주십니다. 카이스트 박사과정 이원형 님은 머잖아 인간의 감정을 이해하는 AI가 등장할 가능성을 기술적, 학문적인 관점에서 진단해주셨습니다.

숨 가쁘게 달려온 2017년도 어느덧 절반이 지나갔습니다. 이번 호에서는 2017년 하반기에 예정된 주요 AI 컨퍼런스를 소개해 드립니다. 다양한 지식 공유의 장과 배움의 기회들을 통해 모든 인류가 함께 AI 시대를 만들어갈 수 있는 ‘모두를 위한 AI’가 되기를 기원합니다.

2017년 7월 28일
카카오 정책지원팀 드림

AI 와 의료



industry	민현석 내가 의료 AI를 선택한 이유	06
	안성민 데이터 기반 정밀 의료와 AI	14
	황상훈 딥러닝 기반 의료영상 기술의 진화	18
	김남국 의료와 AI 신기술의 융합: 과제와 전망	24
	정규환 AI 의료영상 기술 활용 사례	28
	양광모 AI 의료, 이제 윤리를 고민하다	34

AI 기술의 발전이 과연 우리 삶을 변화시켜줄 수 있을까요? 이 질문에 답을 찾기 위한 노력으로 카카오 AI 리포트는 지금까지 다양한 분야, 기술, 활용 가능성 등을 소개해 왔습니다. 이번에는 '의학 분야에 적용되고 있는 AI 기술'을 소개하려고 합니다. 새로운 기술이 개발된다고 해서 바로 돈을 벌 수 있는 산업분야에 응용되기는 쉽지 않습니다. 새로운 기술이 기존에 해결하지 못했던 문제들을 해결할 수 있는 방법을 제시해 준다거나, 기존의 비효율성을 개선한다면 그 기술은 산업 분야의 혁신을 가져올 수 있습니다. 이번 글에서는 의료 AI 스타트업에 종사하시는 AI 전문가들과, 병원에서 환자를 진료하는 의사 선생님 등 현장에서 활동중인 의료 분야 전문가들의 목소리를 통해 AI 기술의 활용 현황 및 한계와 향후 발전 가능성 등에 대해 확인할 수 있습니다

내가 의료 AI를 선택한 이유

의료 AI 바로 알기

필자는 올해 초 3년간의 대기업 연구소 생활을 마치고 AI를 이용해 의료 정보를 분석하는 일을 시작하게 되었다. 얼마 전 마이크로소프트(Microsoft)의 창업자인 빌 게이츠(Bill Gates)는 인터뷰를 통해 자신이 대학생이라면, AI, 에너지, 그리고 생명공학을 공부하겠다고며 유망한 분야 세가지를 꼽았다¹⁾. 세계 최고 부자인 빌 게이츠의 말대로라면 나는 가장 유망한 세가지 중, AI와 생명공학 두가지만 하게 된 것이니 표면적으로는 매우 영리한 선택을 한 것으로 보였다. 대학원 시절 전공은 AI나 생명공학이 아니었고, 회사에서 한 대부분의 일도 AI와는 거리가 있는 일들을 해 왔다. 물론 요즘 같이 AI 인력이 부족한 때에 넓은 시야로 본다면 대부분 AI 관련자로 볼 수 있고, 필자 역시 남들보다 AI 분야를 접하기 용이한 위치에 있었던 것은 사실이다. AI, 그중에서도 의료 AI 분야에 직접 뛰어들면서 느낀 의료 AI의 가능성과 현실적인 어려움, 그리고 앞으로 방향에 대한 생각들을 이 글을 통해 공유하고자 한다.



글 | 민현석 min6284@gmail.com

본업 지우 아빠, 딸바보 공돌이. 부업으로 대학원 시절 전공한 알은 영상처리 지식을 이용하여 잘 나간다는 대기업 AI Lab에서 밥벌이를 하였지만, 부업이 본업을 방해해 의료 AI분야에 뛰어들어 본업도 부업도 충실히 수행하려고 노력하는 중이다. 곧 태어날 둘째 녀석 때문에 기저귀 값이라도 벌기 위해 동분서주하다가 글을 쓰는 지경에 이르렀다. AI 붐을 타고 이 분야에 구경 왔지만, AI 전문가가 아닌 AI를 도구처럼 쓸 수 있는 도메인 전문가가 되길 희망하는 말 많은 딸바보이다. 아.. 그리고 이런 소개글을 빌어, 아내에게 사랑한다는 말을 전하고 싶은 경상도 남자이기도 하다.

AI 영상인식 전문가의 의료 AI 도전기

필자가 대기를 나와 의료AI를 시작한 이유는, 의료 AI는 어느 정도 모험을 해 볼만한 가치가 있는 분야라고 느꼈기 때문이다. 이렇게 생각할 수 있었던 이유는 필자가 기술 기반 스타트업들을 자문할 때 요구했던 3가지 요건들인 ‘문제’, ‘고객’, ‘시장’ 때문이다.

기술로 풀어야 할 많고 중요한 문제

대부분의 기술 기반 서비스라는 것은 기존 기술의 성능이 비즈니스 모델에서 고객이 바라는 성능과의 차이가 있을 때, 간극을 어떻게 매워나가냐에 따라 두 가지로 나눌 수 있다. ‘포켓몬 고’와 같이 기술이 부족할지라도 재미있는 콘텐츠와 UI(user interface)로 그 간극을 메울 수 있는 분야가 있고, 그런 콘텐츠가 아니라 정말 기술의 성능을 향상하여서 간극을 메워야 하는 분야가 있다. 의료 서비스나 자율주행 서비스는 고객이 만족할만한 기술적 성능을 꼭 만족시켜야 시장에 나갈 수 있는 대표적인 서비스이다. 이런 서비스에서의 기술적인 실패는 포켓몬 고와 같은 게임이나 영화 추천과 같은 서비스와 달리 치명적인 피해를 가져올 수 있다는 점이다. 기존 기술에 비해 압도적인 성능을 보였다고 해서, 바로 돈을 벌 수 있는 게 아니며 많은 문제들을 차근차근 해결하면서 발전해야 한다. 내가 의료 AI를 선택한 이유도 바로 기술로 풀어야 할 중요한 문제가 많은 분야이기 때문이다.

현재 AI 기술은 AGI(artificial general intelligence, 인간이 하는 많은 지적 활동들을 동일한 지능으로 할 수 있는 AI를 칭한다)가 아니라 특정 분야 문제를 잘 풀기 위한 좁은 AI(narrow AI)이다. 현재 딥러닝(deep learning) 기술의 발전은 문제를 잘 풀 수 있는 AI가 되었음에는 의심의 여지가 없다. 좋아진 AI는 정말 좋은 도구임에 의심의 여지가 없다. 많은 학생들 및 연구자들이 다양한 온라인 강의들²⁾을 통해 딥러닝을 공부하고 문제를 풀 수 있는 도구를 알게 되었다. ‘이제 어떻게 하지?’ 좋은 도구를 가졌으면, 그 도구를 가지고 해결할 수 있는 좋은 문제를 찾아야 한다. 좋은 문제를 찾는 것은 특정 분야에 대한 많은 경험과 지식을 가지고 있어야 가능한데 의료 분야의 경우 기술로 해결할 수 있는 많은 문제들이 이미 존재하고 있다. AI는 그 해결을 도울 수 있는 좋은 도구이기 때문에 의료 AI는 매력적인 분야로 보였다. 그리고 난 한국의 대표적인 공돌이라 포켓몬고 같은 재밌는 콘텐츠를 만드는 일보다는 확실한 문제를 차근차근 풀어나가는 게 적성에 맞다. 그래 나 재미없다.

어디에도 없는 스승 ‘의사’라는 고객

두 번째 요건은 ‘고객’이다. 스타트업 심사를 가거나 상담을 가면 처음부터 끝까지 ‘고객’에 대해서 질문한다. 기본적으로 모든 서비스의 목적은 고객의 사용자 경험을 극대화시키는 것에 있다고 생각한다. 그러기 위해서 고객이 누구인지 명확히 아는 것이 중요하다. 의료 서비스의 최종 고객이 ‘환자’일수는 있겠지만, 현재 의료 AI의 고객은 ‘의사’라고 생각한다. 앞서 설명한 것처럼 분야의 기술 실수는 실수는 환자의 생명, 혹은 인생을 좌우할 수 있다. 그리고 아직 AI란 도구는 그 선택에 책임을 질 수 있는 단계는 아니다. 그렇기에 의료 AI의 고객은 그 결정에 책임을 질 수 있는 의사여야 한다. 의사가 좀 더 좋은 선택을 할 수 있도록 도와주고, 좀 더 편하게 선택할 수 있도록 도와주고, 지금까지 풀지 못 한 문제를 풀 때 좋은 도구가 되어 주는 것이 현재 의료 AI의 목적이다. 즉, 의사가 의료 AI의 ‘고객’이 되는 것이다

서비스를 만드는 데 있어 고객은 좋은 스승이 될 수 있다. 어떤 서비스건 처음부터 완벽할 수 없기에, 좋은 사용자 피드백은 좋은 서비스를 만들 수 있는 밑거름이 된다. AI를 이용한 서비스 분야도 이와 동일하다. 아니 고객의 피드백이 더 중요할 수 있다. 의사들은 자신들의 분야에서 무엇이 필요하고, 어떤 문제가 있는지 정확히 알고 있는 고객이면서 의료 AI가 갈 방향을 정확히 안내할 수 있는 좋은 안내자이다. AI의 딥러닝 기술은 왜 좋은지 왜 되는지 등이 완벽히 증명되지 않은 블랙박스(black box)와 같은 측면이 많은 기술이지만 예측의 정확도는 기존의 기술을 압도하고 인간 의사 결정과 비등한 결과를 보여주고 있다. 하지만, 그 예측이라는 것이 학습 데이터에 포함된, 이미 의사가 결정낸 정답만을 따라 하는 측면이 있다. 의사가 진찰할 때, 환자의 표정, 이력, 그리고 의사의 다양한 경험을 가지고 환자를 판단하지만, 대부분의 의료 AI는 입력된 영상만을 가지고 판단하고 있기 때문에 환자에 대한 정보가 제한적일 수밖에 없다. 이런 문제를 보완하여 의료 AI 기술을 발전하도록 이끌 수 있는 스승 같은 고객이 바로 의사이다. 이런 스승 같은 고객이 있기에, 난 다른 도메인이 아니라 의료 AI란 도메인을 선택할 수 있었다.

우리는 고객들의 직업을 뺏으러 가는 몽상가가 아니다. 단지 고객들이 가지고 있는 문제를 쉽게 해결할 수 있는 기술을 개발하기를 원하고, 이에 기반해 지금껏 해결하지 못했던 의료의 문제들을 풀기를 원한다. 그렇기에 우리의 스승이자 고객인 의사들이 문제를 해결해 나가는 동업자가 되길 희망한다. 도와주세요.

유행을 타지 않지만 꼭 지켜야 하는 시장

세 번째 요건은 ‘시장’이다. 스타트업들을 상담할 때, 나는 늘 없는 시장을 만들어 내는 것은 있는 시장에 들어가 살아남기보다 10배는 더 힘들다고 말해왔다. ‘의료’라는 시장은 인류 역사상 가장 오래된 시장이고, 가장 큰 시장 중 하나이며, 사라지지 않을 시장이기도 하다. 그렇기 때문에 문제를 잘 해결하는 의료 AI 기술을 만든다면, 시장이 없어서 굶어 죽을 염려는 비교적 하지 않아도 된다. 또한 시장에 대해 말할 때, 스타트업들에 내가 하는 충고 중 하나는 구글, 아마존, 페이스북이 잘 하고 있고, 잘 할 거 같은 시장은 피하란 것이다. 딥러닝 기술의 가장 중요한 점 하나가 데이터 기반(data-driven) 기술이라는 점이다. 각 영역에서 데이터를 가장 많이 확보할 수 있는 플랫폼이 있는 분야는 피하는 것이 좋다. 대학원 시절 마모그램(mammogram) 영상을 연구해 볼 기회가 있었는데, 외국 환자들의 영상자료에 비해 한국 환자들의 영상 자료에 하얀 무늬가 더 많아서 쉽게 구별할 수 있다는 특징이 있었다. 이는 아시아 여성들의 유방에 섬유질이 더 많기 때문에 나타나는 현상이다. 이처럼 단순히 X-ray 영상 하나만으로도 인종별, 대륙별 차이가 나타나고 있다. 의료 AI 기술이 더 발전하면서 영상뿐 아니라 유전자 정보, 생활 습관 등 다양한 정보들이 의료 데이터로 활용될 것이다. 이렇게 되면 같은 병을 다루는 기술일지라도 다른 데이터를 가지고 연구를 해야 한다는 것이다. 100%의 예측 성능을 자랑하는 미국의 기술이라도 한국 의료 시장에서는 100%의 성능을 장담할 수 없다는 것이다. 그렇기 때문에 미국이 의료 AI 기술에서 앞서 나가고 있다고 해서, 혹은 중국의 의료 데이터 양을 따라갈 수 없다고 해서 이 시장을 포기하지 말아야 할 이유가 생긴 것이다.

Big data vs. Good data: 풍요속의 빈곤, 의료 AI가 어려운 이유

지난해 11월 구글은 의학저널인 JAMA(Journal of the American Medical Association)에 딥러닝 기술로 안저 영상(retinal fundus photographs)을 판독해 당뇨병성 망막변증(diabetic retinopathy, DR; 당뇨의 합병증으로 혈관이 좁아지고 막히면서 발생함)을 진단하는 방법에 대한 논문을 게재했다³. 안과 전문의 54명과 함께 약 12만 장의 안저 영상 데이터로 구글의 딥러닝 모델인 인셉션V3(Inception-v3)을 학습시켜 유능한 안과의사에 버금가는 진단 예측 결과를 보인 것이다. 이 논문은 중요한 시사점을 우리에게 던져 주었는데, 질 좋은 데이터가 많은 경우 일부 의료 진단 영역에서는 의료 AI 기술은 의사보다 빠르고 정확한 진단 성능을 보여 준다는 것이다. 최근 구글은 이 기술을 다른 병리영상(imaging and pathology) 영역으로 확대하려고 하고 있다. 이렇게만 보면 조만간 의료 문제 대부분이 해결될 것 같고, 의료 AI 분야는 곧

구글과 같은 대기업에 밀려 없어질 영역이라고 생각될 수 있다. 단지 깃헙(Github)등에 공개된 좋은 딥러닝 기술 코드를 가져와 대용량 데이터에 적용하면 의사에 버금가는 AI 진단 기술을 쉽게 만들 수 있지 않을까라고 생각할 수 있다. 그리고 한국의 경우도 수많은 대형 병원에서 쌓는 많은 데이터들을 이용하면 곧 문제들이 해결될 것이라고 생각할 수 있다. 그러나 현실로 한 걸음 들어가 보면, 풍요 속에 어떤 빈곤이 있는지 확인하게 된다. 한국의 병원에 충분한 데이터가 있을까? 의료 AI 분야를 시작하면서 놀라게 된 것은 대부분의 한국 대형 병원들은 엄청난 데이터를 보유하고 있다는 것이다. 최근 AI가 대세가 되고 나서부터 이들 병원들은 이 빅데이터(big data)에 AI 기술을 적용하기 원하면서 AI 전문가들과 많은 미팅을 갖고 있다. 하지만, AI 전문가들이 하는 말은 ‘활용할 데이터가 없다’라는 것이다.

의사는 학습 데이터를 만드는 사람이 아니다!

이러한 현상이 발생하는 가장 근본적인 이유는 의료 데이터 생성에서부터 출발한다고 본다. 많은 의료 데이터는 의사가 AI 학습용으로 만든 것이 아니라 환자를 진단하고 그 내용을 기록하기 위한 것이다. 의사들은 의료 영상에 정확한 세그멘테이션(segmentation)을 하기 위해 병변(病變)의 위치를 그려놓거나 하지 않는다. 또한 전체 영상에서 병변이 발견된 경우 모든 병변을 꼼꼼히 하나하나 다 찾지 않는다. 아니 찾을 필요도, 찾을 시간도 없다. 영상을 기반으로 진단을 했다면 그 진단 내용과 근거를 기록하고, 전체 영상 내에 존재하는 모든 병변의 위치를 정확히 표시하느라 시간을 낭비하지는 않는다. 또한 모든 의료 영상에 대해 구글의 연구 사례처럼 다수의 의사가 다시 확인하며 살펴볼 이유도 존재하지 않는다. 또 다른 데이터관련 문제 중 하나는, 동일한 병에 관한 동일한 형태의 데이터일지라도 다른 병원, 다른 기계, 다른 시약을 사용하였을 경우, 동일 환자에 대해서도 영상의 특징은 달라질 수 있다는 점이다. 예를 들어 X-ray 기계도 제조사마다 차이가 다르게 나타난다. 그러나 이런 예외상황은 의사가 환자를 진단하는 데에는 큰 문제가 되지 않지만, AI 기술을 적용하는데 문제를 어렵게 만드는 장벽인 것이다. 의사들은 ‘데이터를 만드는 사람’이 아니라 환자를 치료하기 위해 필요한 정보를 기록하는 것이다. 그렇기에 병원들이 차곡차곡 오랜 노력을 들여 쌓은 데이터에 바로 AI 기술을 적용하기에는 많은 어려움이 존재하고, 그 문제를 해결하기 위해서는 많은 노력이 필요한 상황이다.

의료 데이터는 바둑 기보와 다르다

알파고는 결국 세계 랭킹 1위인 커제를 눈물짓게 하고 은퇴를 선언했다. 알파고는 많은 기사들의 기보, 많은 기사들과의 대결, 그리고 스스로와의 대결을 통해 성장했다. 한마디로 많은 시뮬레이션 데이터를 통해 학습할 수 있었다는 것이다. 그러나 의료 분야는 시뮬레이션 데이터를 생성하기 아주 어려운 분야이다. 대부분의 의료 데이터는 오랜 관찰에 의해 생성된 실제 데이터이다. 예를 들어, 우리가 10년 후 특정 암의 재발률에 관한 예측 모델을 만든다고 가정해 보자. 이를 위해 암에 걸린 사람들의 10년 후 데이터를 확보해야 한다. 그러나 인간의 삶이란, 바둑과는 비교도 안 되는 변수와 노이즈가 존재한다. 그리고 그 많은 변수 중 10년 후 재발 데이터에 무엇이 관련 되어 있는지 정확히 아는 사람은 없다. 그렇기에 어떤 정보를 관찰해야 하는지도 정확히 알기 어렵다. 유전자에 의해 달라질 수도 있고, 생활환경에 의해 달라질 수도 있다. 또한 10년 후까지 데이터를 생성하지 못할 수도 있다. 인간의 삶이란 정해진 규칙에 의해 움직이고 평가할 수 있는 바둑판 위의 바둑알이 아니다. 그렇기에, 존재하는 데이터를 잘 활용하는 방법뿐 아니라, 향후 데이터를 어떻게 쌓아가야 할지 충분히 토론해야하고, 그 데이터를 잘 확보하기 위한 좋은 플랫폼도 필요한 것이다.

병원은 환자의 데이터가 있는 곳이다

병원은 환자의 데이터가 있는 곳이다. 참 당연한 말이다. 그러나 이 말의 뜻을 잘 들여다보면, 왜 병원 데이터가 어려운지 알 수 있다. 환자라는 단어를 사전에서 찾아보면 ‘병들거나 다쳐서 치료를 받아야 할 사람’이라고 나온다. 어딘가 병들거나 치료를 받아야 하기에 의식이 되는 부분에 대한 의학 데이터가 주로 존재한다. 그렇다는 건 완벽히 정상인 데이터에 비해 어딘가 아픈 사람의 데이터가 많다는 것이다. 또한 병들거나 다쳐서라는 것은 어떤 특정 병이 있을 경우에 병원에 찾게 된다는 것이다. 그러나 모든 병이 똑같은 확률로 똑같은 상황에서 나타나지는 않는다. 그렇기에 어떤 병에 대한 데이터는 많고, 어떤 병에 대한 데이터는 적을 수밖에 없는 표본수의 차이가 발생하게 된다. 건강한 사람의 데이터와 환자 데이터, 특정 병, 혹은 병의 진행 상황에 따른 데이터의 불균형은 데이터 기반(data-driven) 기술인 딥러닝(deep learning)에서는 성능 저하의 한 가지 요인으로 작용하고 있다.

의료 데이터는 일반 데이터와 다르다

마지막으로 말하고 싶은 것은 의료 데이터는 일반 데이터와 그 성질이 다르다는 것이다. 딥러닝이 가장 빠르게 적용되고

발전된 분야 중 하나는 이미지 인식 분야이다. 그래서 의료 영상도 같은 방식으로 쉽게 적용할 수 있을 것이라고 생각할 수 있으나, 의료 영상과 일반 이미지는 그 성격이 다르다. 예를 들면, ImageNet등에서 공개된 이미지 영상 사이즈(pixels)는 224x224 정도이다. ImageNet에서 있는 영상을 그 사이즈로 줄였을 때, 보통의 사람들도 어렵지 않게 이게 고양이인지 개인지 구별할 수 있다. 하지만, X-ray 이미지의 사이즈로 줄이면 의사들도 특수한 경우를 제외하고는 간암인지 위암인지 알기 쉽지 않게 된다. 의료 영상은 그 목적이 다르기 때문에 일반 영상에서 쓰인 기술을 바로 가져다 쓰면 안 되는 경우가 있다. 최근 미국 New York University 조경현 교수팀은 X-ray에서의 유방암 판독을 위한 딥러닝 연구에서 최대한 원 영상 사이즈를 유지할 때 그 성능이 좋다는 결과를 발표했다⁴. 딥러닝은 기존의 다른 기술에 비해 높은 성능을 보여주고 있지만, 의료 영역의 새로운 문제를 해결하기 위해서는 의료 데이터에 대한 이해를 바탕으로 새로운 기술로 개선되어야 할 필요가 있다.

의료 데이터 문제를 해결하기 위한 노력들

최근 AI 분야에서 가장 각광받고 있는 딥러닝은 좋은 데이터가 많으면 많을수록 결과가 좋아지는 기술이다. 최근 MIT 연구 결과는 최신 딥러닝 기술들은 데이터가 충분할 경우 데이터에 오류가 좀 포함되어 있더라도 좋은 성능을 유지할 수 있다고 발표하고 있다⁵. 하지만, 의료 분야 접근방식은 부족한 데이터로 안정적인 성능을 내는 기술을 개발하거나, 인위적인 생성, 변형을 통해 부족한 데이터를 보충하는 방식을 취하고 있다.

AI with Data

데이터의 양이 부족한 경우, 딥러닝에서 일반적으로 사용되는 방법은 이미 학습된 모델(pre-trained model)을 미세조정(fine-tuning)하는 것이다. 최근 발표된 연구 결과에서는 이런 방법이 의료 데이터에서도 잘 적용된다고 하였다⁶. 다른 분야에서 잘 학습된 모델이 있을 경우 부족한 양의 의료 데이터 문제를 조금 완화시킬 수 있다. 학습 데이터 부족 현상을 극복하기 위해 기본적으로 사용하는 기술은 데이터 증강(data augmentation) 기술이다. 이때 주의할 점은 현실과 너무 동떨어지거나 기존 특징을 왜곡할 수 있는 데이터 증강은 오히려 학습만 어렵게 하거나 학습 성능을 낮추기도 한다. 이런 문제를 극복하기 위한 방법 중 하나로 카카오브레인의 김남주 소장팀은 펄린 노이즈(Perlin noise)를 이용한 효과적인 데이터 증강 기법을 이용하였고, 제한적인 의료

데이터 환경에서 의미 있는 결과를 얻은 연구 결과를 발표하였다⁷. 의료 데이터 부족은 학습 데이터가 충분하지 않다는 문제뿐 아니라 적당한 평가 데이터가 없다는 문제도 존재한다. 이런 문제점을 보완하기 위해 최근 역 테스트(reverse testing) 방법을 의료 학습 모델의 평가에 적용하는 연구가 발표되었다⁸. 이 연구는 학습 모델의 평가 데이터가 없는 경우, 역분류정확도(reverse classification accuracy) 기반으로 원래 분류(classfication) 성능을 유추하는 방법에 관한 것이다.

AI for Data

최근 존재하는 않는 데이터를 생성하거나 상황에 맞게 변형된 데이터를 생성하는 방법이 많이 연구되고 있다. 최근 딥러닝 분야에서 가장 이슈가 되고 있는 기술인 GAN(Generative Adversarial Networks)이 바로 그것이다. GAN은 실제 데이터와 구별하기 힘든 가짜 데이터를 생성하는 기술이다. 최근 GAN 기술을 이용하여 의학 데이터를 생성하거나 수정하는 연구가 많이 이루어지고 있다. 노스캐롤라이나 대학(University of North Carolina)은 GAN 기술을 이용하여 MRI(magnetic resonance imaging) 영상을 기반으로 CT(computer tomography) 영상을 생성하는 연구결과를 발표했다⁹. CT가 MRI와는 달리 방사선 피폭량이 많기 때문에 CT를 많이 찍는 건 다량의 방사선 노출의 위험이 있다. 이 기술은 MRI의 영상을 기반으로 CT 영상을 생성하여 사용하는 기술로서 비용이나 안전도 면에서 유리한 기술이다. CT는 방사선 노출에 대한 문제가 있어 저선량(low-dose)으로 촬영하는 경우가 많은데 이럴 경우 기존 방법에 비해 데이터에 잡음이 많아지는 문제가 발생한다. 이때 잡음을 제거한 영상을 생성하기 위해 딥러닝 기술을 활용한 연구가 이루어지고 있다¹⁰. 부족한 의료 데이터 문제를 해결하기 위한 다양한 AI 기술이 연구되고 있다. 이렇게 생성한 데이터는 여러 규제에 자유롭기도 하고, 부족한 의료 데이터를 보완해주기 때문에 좋은 방향으로 각광받고 있다. 하지만, 진짜 의학지식에 기반한 것인지에 대해서는 여러 검증이 필요한 상황이다.

Data 다다익선(多多益善)

앞서 소개한 의료 데이터 부족 문제를 해결하기 위한 연구들도 여전히 생성된 데이터가 현실을 모두 반영할 수 있냐에 대한 질문에 정확한 답을 하지 못하고 있다. 이런 연구가 진행돼야 하는 것은 맞지만, 본질적인 문제 해결을 위해 어떻게 의료 데이터를 ‘잘’ 확보하는지가 핵심 문제이다. 이렇게 확보한

데이터를 연구자들이 연구에 활용하기 위해서는 좋은 규격을 가지고 있어야 하고, 합리적인 규제 완화가 동반되어야 한다. 아주대학교 박래웅 교수님의 연구¹¹에 따르면 2015년 우리나라 3차 의료기관의 포괄적인(comprehensive) 전자의료기록(electronic medical record, EMR) 도입률이 11.6%에 불과하다고 한다. 의료 데이터를 체계적으로 확보하기 위한 관련 법령도 필요하다. 미국은 의료정보보호법(Health Information Portability and Accountability Act, HIPAA)을 제정하여 의료 데이터를 체계적으로 확보할 수 있도록 한 관계 법령을 정비하고 있다. HIPPA에는 최근 이슈가 되는 랜섬웨어(ransomware)에 관련된 사항까지 포함하고 있을 정도로 현실을 잘 반영하고 있다. 이를 참고하여 한국도 의료 데이터를 체계적으로 확보하며 연구할 수 있는 환경을 만들어야 한다. 데이터가 없다면, 아무리 의사와 AI 연구자가 좋은 기술을 만들어도 문제를 해결할 수 없다.

의사 with AI

대학시절 개인과외를 하던 시절을 되돌아보면, 내 수업의 첫 과정은 학생과 앉아서 학생이 무엇을 아는지 확실히 분석하는 것이었다. 학생이 푼 문제를 조목조목 네가지 종류로 분류했다. 1. 알고 맞춘 것, 2. 알고 실수로 틀린 것, 3. 모르고 찍어 맞춘 것, 그리고 4. 몰라서 틀린 것. 해결하기 가장 쉬운 문제는 4번 몰라서 틀린 경우이다. 모르면 알려주면 된다. 그래도 모른다면 또 알려주면 된다. 2번의 ‘알고도 틀린 경우’는 원인이 무엇인지 더 자세히 분석해야 한다. 문제 스타일인지, 자신의 버릇인지에 대해 원인을 더 자세히 살펴봐야 한다. 가장 분석하기 힘든 것은 3번 ‘모르고 찍어 맞춘’ 문제이다. 채점만 해서는 이 문제를 찾을 수 없다. 여러 번 비슷한 문제를 반복적으로 풀면서 그런 실마리를 찾든가, 아니면 정말 학생을 잘 파악해야 한다. 갑자기 학생의 학업 성향을 이야기하게 된 것은 AI 기술의 학습과도 연관이 되어있기 때문이다. 요즘 AI 기술들의 정확도 역시 채점된 점수로서 그 성능을 평가하기 때문이다. AI 기술이 해결한 문제들 중에도 학생들이 푼 것처럼 4가지 유형의 결과가 숨어있을 텐데, 이에 대한 분석이 없는 경우가 많다. 시가 틀렸을 경우, 데이터가 부족해서 틀렸는지, 그 문제에 존재하는 특정 노이즈 때문에 틀렸는지, 아니면 AI 기술이 그 문제와 맞지 않는지에 대한 분석 없이 확립화된 정확도라는 잣대로 평가하고 있다.

허세가 아닌 정확한 정보를

딥러닝의 예측의 정확도는 흔히 %로 표시가 되지만, 확률을 의미하지는 않는다. 그러다 보니 통계적으로 신뢰구간을 결정하기 힘들어 결과에 대해서 얼마나 신뢰를 해야 하는지 알 수 없다. 개와 고양이 사진을 구분하는 딥러닝 모델은 개와 고양이가 아닌 다른 사진을 넣어도 개 혹은 고양이로 결정하게 된다. 딥러닝 모델은 모르는 문제에 대해 학습된 결과만을 제시하고 있다. 즉 모르는 걸 모른다고 하지 않는다. 이런 문제를 개선하기 위하여 불확실한 상황의 AI 추론(AI reasoning under uncertainty)의 연구가 이어져오고 있다¹². 예를 들어 딥러닝 모델이 의료영상을 보고 암이라고 판단한 경우 신뢰할 수 있는 판단 수준을 제공하는 것이다. AI 결정의 신뢰도가 낮은 경우 의사가 추가적인 판단을 하든지, 추가 검사를 할 수 있게 한다¹³. 의료 분야에서 AI 결정은 치명적일 수 있기 때문에, AI 결정은 최대한의 정보를 의사에 전달해서 의사의 추가적인 결정에 도움을 주어야 한다. 신뢰도가 낮은 AI 결정이 데이터의 부족 때문인지, 문제가 특수하기 때문인지에 대한 정보의 전달도 필요하다. AI 전문가들이 99%의 예측 성능을 가진다고 하더라도 쉬운 문제를 학습해 달성한 99%의 성능인지 또 예측하지 못한 1%가 얼마나 치명적인 결과일 수 있는지에 대한 정보를 정확하게 전달할 수 있어야 한다.

최근 MIT Technology Review에 실린 “The Dark Secret at the Heart of AI” 기사¹⁴는 AI 기술의 핵심인 딥러닝의 문제점을 지적하고 있다. 현재 우리는 똑똑한 AI 모델을 만들 수는 있지만 아무도 그것이 어떻게? 왜? 동작하는지 완벽하게 모른다는 점이다. 어떤 근거로 판단을 했는지 설명하지 못하면, 고객 입장에서 아무리 좋은 성능을 보인다고 하더라도 판단을 믿기 쉽지 않다. AI 연구자들이 할 수 있는 방법은 고객이 이해할 수 있는 분야의 정보로 표현하는 것이다. 의료 영상에 대한 판단이 이루어진 경우, 영상의 어떤 부분이 판단에 영향을 주었고, 어떤 부분이 바뀌면 판단이 바뀌는지 등에 대한 정보를 고객에게 전달하는 것이다. 의료 AI의 판단을 의사가 분석할 수 있게 해야 하고, 추가적인 분석을 AI 모델에 반영해 의료 AI 기술을 발전시켜 나가야 한다.

Pictures are not taken in a vacuum

“Pictures are not taken in a vacuum – an overview of exploiting context for semantic scene content understanding” 제목의 논문¹⁵이 있다. 난 늘 이 문장이 좋았다. 논문의 핵심 내용은 사진 바깥에 존재하는 정보에 대한 것이다. 사진이 찍힌 시간, 장소, 사진을 찍을 때 저장되는 메타데이터 그리고 이 사진을 찍기 전후에 찍었던 사진들, 이런 정보가 모두 하나의 사진을 이해하기 위한

상황(context) 정보가 된다는 내용이다. 논문이 발표된 후 시간이 꽤 지나서 논문에 나왔던 기술들은 현재 쓰이지 않는 경우가 많지만, 상황 정보와 사전 지식(prior knowledge)을 사용하여 더 나은 성능을 보일 수 있다는 것은 의료분야에도 시사하는 바가 있다. 의료 데이터는 환자가 살아온 인생, 환경 그리고 유전 정보와 의학적 데이터들을 포함하고 있다. 이 모든 정보들이 분석하고자 하는 의료 데이터를 더 잘 이해하게 해주는 또 다른 정보가 된다. 오랜 기간 동안 의사들이 연구하고 쌓아온 의료 정보는 AI를 학습시킬 수 있는 중요한 교과서인 것이다.

의학의 추가적인 정보를 활용한 논문을 소개하고자 한다.

“Prediction of Kidney Function from Biopsy Images Using Convolutional Neural Networks” 논문에선 콩팥의 병리 영상을 이용하여 콩팥 병을 판단하는 데 사용되는 중요한 수치 중 하나인 eGFR(estimated glomerular filtration rate)의 12개월 후 값을 예측하는 연구를 소개하였다. 병리 영상만 이용해 학습할 때보다, 현재 eGFR 정보를 모델에 추가했을 때, 학습 속도가 2배 빨라지며 예측 오류도 줄어든다는 것이다. 미래 eGFR 값과 현재 eGFR 값은 상호관계가 있기 때문에 학습을 위한 추가 정보를 제공해 준 것이다. 의사들의 의학 지식을 기존 AI 학습의 데이터와 접목한 경우 좋은 성과가 나오고 있는 것이다. AI 전문가가 단순히 데이터를 많이 가지고 있다고 해서 이런 결과가 나오는 것은 아니다. 데이터의 양으로 본다면 중국과 경쟁이 되지 않는다. 최근 중국에서는 정부 주도로 엄청난 양의 의료 데이터를 수많은 인력을 활용하여 확보하고 있다. 이는 훌륭한 학습 데이터로 활용될 것으로 예상된다. 그러나 앞서 말한 바와 같이 의료 데이터는 기보가 아니다. 환자의 상황에 맞게 좋은 처방을 하고, 상태를 잘 살필 수 있어야 좋은 데이터를 만드는 것이다. 이런 측면에서 한국은 의료 기술과 데이터를 확보하기에 좋은 환경을 갖춘 병원들이 존재한다. 의학적 지식과 경험을 가진 의사와 AI 전문가가 함께 한다면 중국과 미국이 주도하고 있는 의료 AI 시장에서 경쟁할 수 있다고 생각한다.

기술은 마케팅이 아니라 고객 질문에 답을 하는 것

네이처 지(Nature)에 실린 “Publish houses of brick, not mansions of straw” 제목의 글에서 논문 리뷰는 논문이 사실일 경우의 임팩트를 보는 것이 아니라, 논문이 사실인지를 판단하는 과정이라는 메시지를 전달하고 있다. AI와 관련된 논문은 서로 임팩트 경쟁을 하듯이 기술을 소개하고 있다. 하지만, 기술을 의료 분야에 적용하기 위해서는 임팩트도 중요하지만 다음 두 가지가 먼저 지켜져야 한다. 첫째, 기술이 탄탄한 사실에 근거해야 한다. 둘째, 테스트 데이터에서만 존재하는 결과가 아니라 의료

데이터에도 동일하게 적용되어야 한다. 두 가지 중 하나라도 만족하지 못하면 그 기술은 의료 분야에 적용하기 힘들다. 일반 AI 연구자와 의료 AI 연구자를 한마디로 구분하자면, AI 연구자는 AI의 문제를 해결하는 연구자이고, 의료 AI 연구자는 AI 기술을 활용해 의료의 문제를 해결하는 것이다. 의료 AI에서는 기술을 의료 AI에 적용하기 위해서 당연히 가져야 할 기본적인 질문에 답하는 것이 중요하다. 예를 들어, 새로운 영상 인식 기술이 나왔다면, 의료 AI에서는 인증 차이에 따른 변인은 없는지, 병원 장비에 따른 차이가 발생하는지, 데이터가 양이 적을 때 잘 적용되는지 등 사소해 보이는 문제를 고민해야 한다. 이를 해결한다고 해도 유명 저널에 논문을 실을 수는 없지만 우리의 고객인 의사들이 가질 수 있는 기본적인 질문에 답을 해나가야 한다. 다시 한번 강조하지만, 서비스의 목표는 고객이고, 우리의 고객은 의사, 그리고 최종적으론 환자이다. 높은 예측 정확도와 임팩트 있는 논문 같은 마케팅에 좋은 답이 아니라, 고객이 원하는 질문에 답을 하기 위해 노력해야 한다. 그래야 우리는 고객과 함께 성장할 수 있다. 그래야 정말 의사들이 해결하고 싶은 의료 문제들에 시가 활용될 수 있다.

글을 마무리하며

의료 AI 분야만큼은 많은 분들이 함께 해야 많은 문제를 해결할 수 있다고 생각하기에 욕먹을 각오를 하고 글을 적어 봤다. 연구 미팅을 위해 병원 로비에서 약속 시간을 기다리던 중, 어느 할머니 손에 꼭 쥐어진 버스표 한 장을 발견하고 ‘아! 내가 이 분야에 대해 하나도 아는 게 없구나’를 알게 되었다. 진료를 받기 위해 먼길을 와야만 했던 상황과 현실이 과연 데이터에 반영될 수 있을까? 모델 학습을 위해 제공받은 데이터가 정리가 되지 못해서 불평만 하던 자신을 돌아보면서 내가 의료 AI를 이해하고 있는 것인가 다시 한번 생각하게 되었다. 의료 AI는 AI와 다르다. AI 기술에 대한 이해 못지않게 의료 환경을 이해해야 한다. 그런 이해를 위해 조금이나마 도움이 되고자 이 글을 적어 보았다.

***1** 참고 : <http://www.cnn.com/2017/05/15/billionaire-bill-gates-reveals-his-biggest-regrets-and-best-advice.html> ***2** 참고 : AI 유튜브 강의 소개, http://lboon.kakao.com/kakao-it/aireport_03_youtube ***3** 참고 : <http://jamanetwork.com/journals/jama/article-abstract/2588763> ***4** 논문 : Geras, Krzysztof J., et al. "High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks." arXiv preprint arXiv:1703.07047 (2017). ***5** 논문 : Rolnick, David, et al. "Deep Learning is Robust to Massive Label Noise." arXiv preprint arXiv:1705.10694 (2017). ***6** 논문 : Tajbakhsh, Nima, et al. "Convolutional neural networks for medical image analysis: Full training or fine tuning?." IEEE transactions on medical imaging 35.5 (2016): 1299-1312. ***7** 참고 : https://www.slideshare.net/ssuser77ee21/a-pixel-topixel-segmentation-method-of-dild-without-masks-using-cnn-and-perlin-noise?qid=75235b71-85d9-43b8-a6bf-ffcefeb26953&v&from_search=1 ***8** 논문 : Valindria, Vanya V., et al. "Reverse Classification Accuracy: Predicting Segmentation Performance in the Absence of Ground Truth." IEEE Transactions on Medical Imaging (2017). ***9** 논문 : Nie, Dong, et al. "Medical Image Synthesis with Context-Aware Generative Adversarial Networks." arXiv preprint arXiv:1612.05362 (2016). ***10** 논문 : Wolterink, Jelmer M., et al. "Generative Adversarial Networks for Noise Reduction in Low-Dose CT." IEEE Transactions on Medical Imaging (2017). ***11** 논문 : Kim, Young-Gun, et al. "Rate of electronic health record adoption in South Korea: A nation-wide survey." International Journal of Medical Informatics 101 (2017): 100-107. ***12** 논문 : Gal, Yarin, and Zoubin Ghahramani. "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning." international conference on machine learning. 2016. ***13** 논문 : Kendall, Alex, and Yarin Gal. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?." arXiv preprint arXiv:1703.04977 (2017). ***14** 참고 : <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> ***15** 논문 : Luo, Jiebo, Matthew Boutell, and Christopher Brown. "Pictures are not taken in a vacuum—an overview of exploiting context for semantic scene content understanding." IEEE Signal Processing Magazine 23.2 (2006): 101-114.

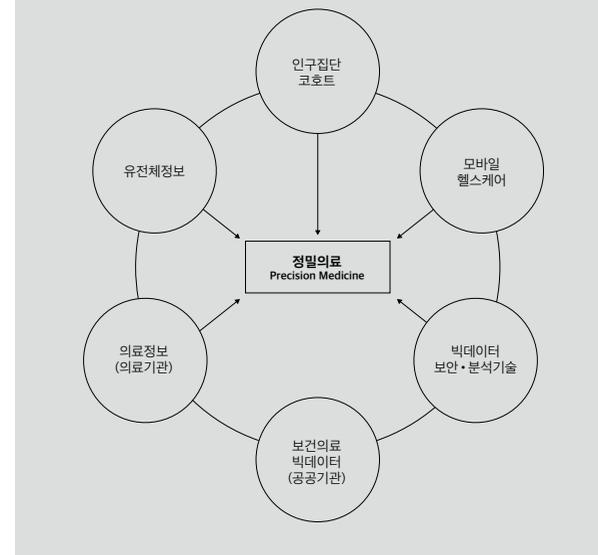
데이터 기반 정밀 의료와 AI

썬 마이크로시스템즈는 “네트워크가 컴퓨터”라는 슬로건과 인터넷의 기반언어라 할 수 있는 자바로 유명하다. 2009년 오라클에 합병되기 전까지 30년 가까이 IT 업계의 강자로 군림한 이 회사는 4명의 창업자에 의해 1982년 설립되었다. 그 중 한 명이자 썬 마이크로시스템즈의 초대 CEO, 지금도 실리콘 벨리의 IT 구루 중 한 명으로 꼽히는 비노드 코슬라는 2013년 다음과 같은 말을 남겼다. “앞으로 10년 간은 의학에서 데이터 과학과 소프트웨어의 기여가 생물학 분야 전체의 기여보다 더 클 것이다(In the next 10 years, data science and software will do more for medicine than all of the biological sciences together)¹⁾.” 이 말을 이해하려면 정밀의료(precision medicine)로 대표되는 의학의 큰 변화 및 이와 함께 등장하는 인공지능과의 관계에 대해 파악해볼 필요가 있다.

오바마 전 대통령은 2015년 2월 정밀의료 추진계획을 전격적으로 발표했고, 미국 국립보건원(NIH)은 이에 맞추어 정밀의료의 개념을 다음과 같이 정의했다. “정밀의료란 유전자, 환경, 생활습관 등의 개인적 차이를 고려하여 질병을 예방하고 치료 기술을 개발하기 위한 새로운 의학적 접근방법을 말한다”²⁾.

위의 정의는 포괄적이고 모호하다. 정밀医료를 이해하기 위해서는 NIH가 내세운 시기적 타당성에 초점을 맞출 필요가 있다. NIH는 정밀의료의 시기적 타당성으로 유전체 해독 기술의 발전, 빅데이터 사용 기술의 발전, 의생명 분석 기술의 발전을 들었다. 진단 및 영상검사와 같은 의생명 분석 기술의 발전은 지난 수십 년간 지속적으로 이루어져왔기 때문에 사실상 정밀医료를 견인하는 두 개의 수레바퀴는 유전체 기술과 빅데이터 기술이다.

[그림 1] 정밀의료의 구성 요소들



이런 관점에서 볼 때 정밀의료란 데이터 기반 의료이다. 데이터 분석은 히포크라테스 이후 모든 의사에게 필수적인 요소이다. 의사는 환자가 진료실에 들어오는 순간부터 데이터를 수집하고 분석한 뒤 치료를 위한 의사결정을 내린다. 지난 1세기 동안 질병에 관한 치료법이 눈부시게 발전했지만 보다 더 비약적으로 발전한 것은 환자에 관한 데이터를 수집하는 방법과 데이터의 양이다. 데이터 수집 관점에서 볼 때 의학은 지난 1세기 동안 청진기의 시대에서 진단 및 영상의학의 시대를 거쳐 유전체와 빅데이터의 시대에 진입했다. IBM의 분석에 의하면 한 사람은 평생동안 0.4테라바이트(TB)의 임상데이터(병원에서 생산된)와 6TB의 유전체데이터, 1100TB의 라이프스타일 데이터를 생산한다³⁾. 이제 의사들은 병원 안에서 생산되는 데이터뿐만 아니라 모바일 기기 등에서 생산되는 다양한 라이프스타일 데이터를 환자 치료에 활용해야 하는 새로운 도전에 직면하고 있다. 요컨대 의학은 데이터

과학이고, 정밀의료는 데이터 기반 의료이다.

의학이 데이터 과학이라면, 데이터 과학의 도구인 인공지능과 의학의 결합은 필연적이다. 몇 가지 예를 들어보자.

1. 인공지능 내시경 마취 솔루션

우리나라에서도 위내시경이나 대장내시경검사를 할 때 소위 수면내시경이 선호된다. 세계 최대의 바이오-제약기업인 존슨앤드존슨은 2009년 수면유도 마취로봇인 세더시스(Sedasys)를 개발했고 2013년 미국 식품의약국(FDA)의 허가를 받아 미국, 캐나다, 호주 병원에서 성공적으로 사용한 바 있다. 세더시스는 내장된 인공지능을 이용해 환자의 혈중 산소 함량, 심장박동 수 등의 신체 징후를 모니터링하면서 투약량을 조절한다. 세더시스를 이용하면 마취 비용을 1/10로 대폭 줄일 수 있다. 불행히도 존슨앤드존슨은 마취전문협회 등의 반발 때문에 2016년 3월 세더시스의 생산 및 판매를 중단했다⁴⁾.

2. 인공지능 영상판독 지원 솔루션

영상의학 발전의 역사적 배경을 고려하지 않고 현대 IT 기술 측면으로만 볼 때 가장 납득하기 어려운 분야가 바로 영상판독이다. 현대 병원에서 다루어지는 영상이미지는 모두 디지털 이미지도다. 그리고, 영상학과 전문의는 디지털 이미지를 눈으로 들여다보면서 아날로그적으로 분석한다. 물론, 여기에는 의사의 지식과 경험, 통찰력이 종합적으로 가미된다. 인공지능이 본격적으로 도입되면서 영상판독에는 완전히 새로운 장이 열리고 있다. 유방 엑스선 검사는 유방암 조기 진단을 위해 널리 사용되고 있다. 미국에서는 두 명의 영상의학전문가가 유방엑스선 사진을 검토하도록 정해져 있지만, 이미 오래전부터 인공지능 솔루션이 영상의학 전문의 한 명을 대체할 수 있다. 실리콘 벨리의 투자전문가인 앤디 케슬러가 10년 전에 출판한 “의사가 사라진다(The End of Medicine)”라는 책에 이와 관련한 내용이 생생하게 소개되어 있다⁵⁾. 우리나라에서도 다양한 노력들이 있다. 서울아산병원 영상학과 의료진은 뷰노코리아와 함께 골 연령 판독 인공지능 소프트웨어를 만들었다. 뼈의 엑스선 영상을 가지고 소아과 환자의 뼈 나이를 판독하는 것이다. 평균 5분 걸리는 판독 시간을 20초로 줄일 수 있다⁶⁾.

3. 인공지능 중환자실 솔루션

중환자실은 병원에서 가장 데이터가 많이 생산되는 곳이다. 중환자실에서는 환자에게 다양한 센서를 부착한 뒤 각종 징후를 지속적으로 모니터링한다. 지금까지는 이러한 데이터를 효율적으로 사용하지 못했고 따라서 어떤 사건이 발생할 경우 의료진이

글 | 안성민 smahn@gachon.ac.kr

가천대 길병원 중앙내과, 의과대학 유전체외과학교실에서 일하고 있습니다. 의과대학을 졸업한 뒤 웰번 대학에서 유전단백체로 박사학위를 받았고, 데이터 기반 연구인 유전체 연구와 이를 임상 분야에 증개하는 역할에 힘써왔습니다. 데이터를 다루고 거기에서 의미를 찾는 데이터 과학자로서 발전해 왔으며 데이터 과학의 도구로서 인공지능을 활용하는 일에 관심이 있습니다.

신속하게 대처하는 일에 초점이 맞추어졌다. 인공지능의 도입이 이러한 풍경을 완전히 바꾸어 놓고 있다.

아주대 의대와 아주대병원은 응급, 중환자 생체 정보를 통합 저장 및 분석하는 인프라 구축을 2016년 11월 완료했고, 인공지능 기반의 통합 모니터링 시스템을 구축할 계획이다. 이렇게 되면 환자의 응급 상황을 최대 3시간 전에 예측하고 선제적으로 대응할 수 있는 새로운 개념의 환자 관리가 가능하다. 또한 하나의 인공지능 기반 모니터링 센터가 다수의 중환자실을 효율적으로 관리하는 것이 가능하다⁷⁾.

4. 인공지능 암진료 지원 솔루션(IBM Watson for Oncology, WfO)

의료계에서는 2016년 가천대 길병원에 처음 도입된 IBM 왓슨포온콜로지라는 인공지능 솔루션이 화제를 모으고 있다. IBM 왓슨은 IBM이 개발한 인지컴퓨팅(cognitive computing) 플랫폼을 말한다. 왓슨은 IBM 창업자의 이름이다. WfO를 개발하기 위해 IBM은 인지컴퓨팅 플랫폼에 암치료와 관련한 방대한 자료를 학습시켰고, 최종적으로는 메모리얼 슬로언 케터링 암센터(MSKCC)의 암치료 데이터에 최적화시켰다. 요컨대, WfO는 MSKCC에서 내리는 치료의사결정과 동일한 결정을 내리도록 훈련된 인공지능 솔루션이며, 새로운 연구결과와 약물, 치료방침 도입에 따라 지속적으로 업그레이드되고 있다. 물론, 여기에는 MSKCC의 암치료가 전세계에서 가장 모범적인 수준이라는 전제가 깔려있다.

최근 국내 여러 신문에 WfO에 관해 “인공지능과 의사”간의 차이를 부각한 글들이 실린 바 있다. 이는 근본적으로 이 WfO가 작동하는 방식과 암치료에 대한 잘못된 이해에서 기인한 것이다. WfO는 특정 암환자가 MSKCC에 갔을 때 그곳에서 받을 수 있는 치료 옵션을 추천한다. IBM에 따르면 WfO의 추천 옵션과 MSKCC에서 이루어지고 있는 치료는 99.9% 일치한다. 2017년 4월 6일 중앙일보에 실린 것처럼 인도 마니팔 병원과 WfO의 일치율 비교 데이터를 “실력”의 간접적 평가로 접근하는 것은 큰 오해의 소지가 있다⁸⁾. 인도 마니팔 병원에서 이루어지고 있는 암치료 옵션과 WfO의 추천을 비교하는 것은 “의사 대 인공지능”의 비교가 아니라 마니팔 병원의 특이 치료 패턴과 MSKCC의 치료 패턴을 비교하는 것이다. 예를 들어, 폐암의 경우 미국은 물론 우리나라도 고가의 표적치료제가 다수 사용되고 있지만, 이를 경제적으로 감당하기 어려운 나라에서는 효과는 떨어지지만 기존의 세포독성항암제를 사용하고 있다. 이런 경우 당연히 WfO의 추천 옵션과 해당 병원 암치료 옵션과의 일치율은 떨어질 수 밖에 없다(마니팔 병원에서는 폐암의 경우 WfO와 일치율 17.8%). 가천대 길병원도 다양한 암종에서 이러한 일치율 분석을 진행하고 있으며

이 결과를 바탕으로 WfO의 길병원내 사용을 최적화하고 있다.

일치율은 WfO의 실력의 기준이 될 수 없을뿐더러 WfO의 사용에 장애가 되지 않는다..

인공지능의 본격적인 도입은 의학에 어떤 영향을 미칠 것인가? 병원에도 소위 4차 산업혁명이 일어날 수 있을까? 이 질문에 답하기 위해서는 현대의학과 전문성, 인공지능의 관계를 파악해야 한다. 현대의학과 병원은 지난 한 세기 동안 모더니즘적인 발전을 통해 현재의 체제로 분화해왔다. 예를 들어, 외과와 내과로, 다시 내과가 신장내과, 류마티스내과, 중양내과, 심장내과, 소화기내과, 호흡기내과 등으로. 이러한 모더니즘적 분화의 중심에는 “전문성”이 자리잡고 있다. 즉, 중양내과와 심장내과 사이에는 차별화된 전문성이 존재하며, 전문성을 확보하기 위해서는 일정 기간의 수련이 필요하다는 개념이다. 현대병원은 차별화된 전문성을 바탕으로 각 분과가 개별 환자를 보면서 동시에 협업하는 모더니즘적 건축물이다.

인공지능은 전문성을 제공한다. 전문성을 설명하는 모델만 7가지 이상이 되고, 전문성을 구성하는 여러 요소가 있지만 가장 중요한 것은 결국 전문지식이다. 전문지식은 이론적 지식뿐만 아니라 적용 가능한 실용적 지식과 능력을 포함한다⁹⁾. 예를 들어, 외과의사는 특정 수술에 관한 이론적 지식을 완벽하게 갖추어야 하는 것은 물론이지만 이를 구현할 수 있는 실용적 지식과 능력이 없다면 전문성을 인정받을 수 없다. 따라서, 의학에서는 인공지능이 인간의 전문성을 완벽하게 대체하는 것은 어려우며 인공지능과 인간 과의 적극적인 협업모델이 주류를 이루게 될 것이다.

아직도 많은 사람이 인공지능이 나온가 인간이 나온가를 질문하고 있다. 분야에 따라 진행 속도는 다르겠지만 이미 이런 질문 자체에 큰 가치는 없으며, 알파고의 바둑 은퇴가 이를 상징적으로 보여준다. 적어도 의학에서 우리가 해야하는 질문은 다음과 같다. “인공지능이 의사의 전문성을 적어도 부분적으로 제공할 수 있다면, 반대로, 의사가 인공지능을 통해 추가적인 전문성을 확보할 수 있다면 전문성에 기초한 현대의학의 구조는 어떻게 바뀔 것인가?”

지금까지 설명한 내용을 바탕으로 이 질문에 답을 시도해보면 다음과 같다. 모더니즘적으로 분화해온 의학 분과의 전문성을 인공지능을 통해 적어도 부분적으로 확보할 수 있다면(분과마다 정도의 차이가 있지만), 분과 역할의 재조정이 이루어질 것이며, 개별 의사의 역량은 확대될 것이다(추가적 전문성 확보를 통해). 그리고, 이를 바탕으로 새로운 구조의 병원 및 서비스가 출현하면서 현대의학의 포스트모더니즘 시대가 열리고 이는 헬스케어의 4차 산업혁명과 연결될 것이다.

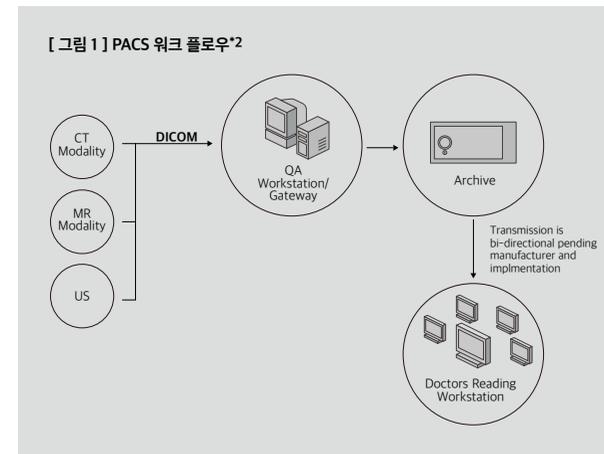
¹⁾ 참고 : TechCrunch Disrupt SF 2013. <https://techcrunch.com/2013/09/11/vinod-khosla-in-the-next-10-years-data-science-will-do-more-for-medicine-than-all-biological-sciences-combined/> ²⁾ 참고 : Precision Medicine Initiative. <https://obamawhitehouse.archives.gov/blog/2015/01/30/precision-medicine-initiative-data-driven-treatments-unique-your-own-body> ³⁾ 참고 : 2014 IBM Health and Social Programs Summit ⁴⁾ 참고 : 매일경제, “질대# 일자리는 료밖에 뵈기지 않는다, 2016.3.30. <http://news.mk.co.kr/newsRead.php?year=2016&no=235945>. ⁵⁾ 참고 : 엔디 케슬러 저/김승욱 역, “의사가 사라진다”. 프로네시스, 2006. ⁶⁾ 참고 : 조선비즈, [4차혁명 생생현장]① “5분 걸리던 영상판독 20초로 줄었죠”...딥러닝 기반 X-레이 판독 공동연구 서울아산병원을 기다. 2017.1.1. http://biz.chosun.com/site/data/html_dir/2016/12/30/2016123000684.html ⁷⁾ 참고 : Science Times 2017.7.14. <http://www.sciencetimes.co.kr/?news=%EC%A4%91%ED%99%98%EC%9E%90-%EC%83%81%ED%83%9C-%EC%98%88%EC%B8%A1%EB%8F%84-ai%EB%A1%9C> ⁸⁾ 참고 : 중앙일보, 중환자 상태 예측도 AI로, 2017.4.6. <http://news.joins.com/article/21445412> ⁹⁾ 참고 : 리처드 서스킨드 및 대니얼 서스킨드 저/위대선 역, “4차 산업혁명 시대, 전문직의 미래”. 와이즈베리, 2016.

딥러닝 기반 의료영상 기술의 진화

컴퓨터를 이용하여 의료영상을 분석하고 진단하고자 하는 시도는 꽤 긴 역사를 가지고 있습니다. 컴퓨터 보조 진단(computer-aided diagnosis, CAD) 개념은 지금으로부터 약 50년 전 미국의 Gwilym S. Lodwick이라는 의사가 처음 제안했습니다¹. 이 연구에서 그는 흉부 X선 촬영 영상을 기반으로 어떤 폐암 환자의 일 년 후 생존 여부를 예측하는 시스템을 개발했습니다. 하지만 당시에는 영상을 스캔하여 디지털화하는 기술이 없었고 자연스럽게 이러한 영상을 처리할 수 있는 컴퓨팅 기술도 없었기 때문에 영상으로부터 중요하다고 판단되는 예측 변수들을 손수 추출했습니다.

실제로 의료영상을 스캔하고 이렇게 디지털화된 영상을 컴퓨터를 이용하여 처리하기 시작한 연구는 1970년대에 등장합니다. 이때부터 여러 영상처리 기법들을 이용하여 추출한 객체의 가장자리, 선분 등의 영상 특징들을 활용하기 시작합니다. 이러한 특징들에 기반한 수학적 모델링을 통해 규칙기반(rule-based) 시스템이 만들어지는데 이는 비슷한 시기에 인공지능 분야에서 유행했던 전문가 시스템(expert system)과 유사합니다.

1980년대에 들어서 CAD 시스템의 발전을 가속화시키는 여러 요인들이 등장합니다. 그중 가장 중요한 요인은 바로 의료영상 저장 및 전송시스템 (picture archiving and communication system, PACS)의 도입입니다 ([그림1] 참고). 디지털화된 영상이 의사들의 판독 능력에 미치는 영향이 검증된 이후 이 PACS는 가장 효율적이고 경제적으로 의료영상을 저장하고 전송할 수 있는 시스템으로 자리 잡습니다. 다른 한 가지는 CAD를 바라보는 패러다임의 변화입니다. 이전에는 CAD의 개념이 모호하여 주로 컴퓨터를 이용한 진단 자동화에 초점이 맞춰져 있었다면 이 시기부터 CAD의 개념이 보조 진단으로 확실하게 자리 잡게 됩니다. 즉, CAD 시스템을 의사의 판독 이후 보조 기구로 활용했을 때 원래의 판독 능력보다 나아지지만 하면 충분히 가치가 있다는 것입니다. 이는 CAD 시스템의 판독 능력이 전문가의 그것과 비슷하거나 상회하지 않아도 상호보완적인 역할을 할 수만 있다면 활용 가치가 있다는 것인데 그렇다고 해서 이러한 개념이 CAD 시스템 개발 과정에 직접 반영되지는 않았습니다. 단지 가치를 평가하는 방식과 기준만 변화한 것이지요.



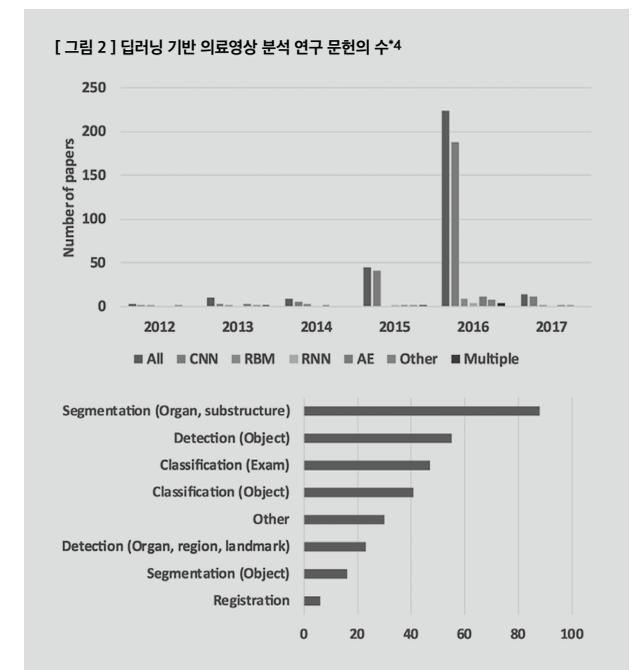
CAD라는 개념의 대중화에 가장 큰 역할을 한 곳은 미국 시카고 대학의 Kurt Rossmann Laboratories for Radiologic Image Research 그룹입니다³. 이곳에서는 의료 현장에서의 효과가 가장 클 것으로 판단된 혈관영상 (vascular imaging), 흉부 X선 촬영 영상, 유방촬영영상 분석을 주요 연구과제로 선정하여 선도적인 연구를 진행했습니다. 이 연구 그룹은 1993년 북미영상의학회(Radiological Society of North America, RSNA)에서 자신들이 개발한 CAD 시스템을 시연하기 위해 직접 118명의 영상의학전문의를 초청하여 검증하기도 했습니다.

이 시기에 활용된 영상 분석 기법은 기술적인 관점에서 패턴 인식 혹은 기계 학습으로 분류할 수 있습니다. 영상으로부터 주요 특징들을 추출하여 이 추출된 특징들로 영상을 벡터화한 후 다양한 기계학습 분류기법들을 활용합니다. 이런 방식의 기술이

주류를 이뤄 개발되어 오다가 최근 들어 딥러닝을 기반으로 하는 인공지능 기술의 혁신적인 발전으로 접근방법이 급격히 바뀌게 됩니다. 다루는 문제에 따라 중요한 특징들을 직접 디자인하고 추출하던 이전의 방식들이 데이터로부터 문제의 해결에 최적화된 특징들을 학습하는 방식으로 변화한 것입니다. 이러한 주류 접근 방식의 변화는 일반적인 컴퓨터 비전 연구에서의 변화와 그 맥락을 같이 합니다. 규칙 기반의 전문가 시스템에서 시작하여 추출된 특징 벡터를 기반으로 분류기를 학습하는 방식이 주류를 이루어 오다가 최근 들어 딥러닝으로 수렴하는 추세입니다.

딥러닝 기반의 의료영상 분석

의료영상 분석의 세부적인 주요 과제들은 일반 영상의 경우와 상당히 유사합니다. 영상을 분류(classification) 하는 것을 시작으로 객체의 검출(detection), 객체 경계의 추출(segmentation), 서로 다른 영상의 정합(registration) 등이 의료영상 분석에서 중요한 과제들이라고 볼 수 있습니다. 기본적으로 영상을 입력으로 하기 때문에 영상에서 특징을 추출하는데 특화된 컨볼루션 신경망(convolutional neural networks, CNN)이 가장 많이 활용됩니다.



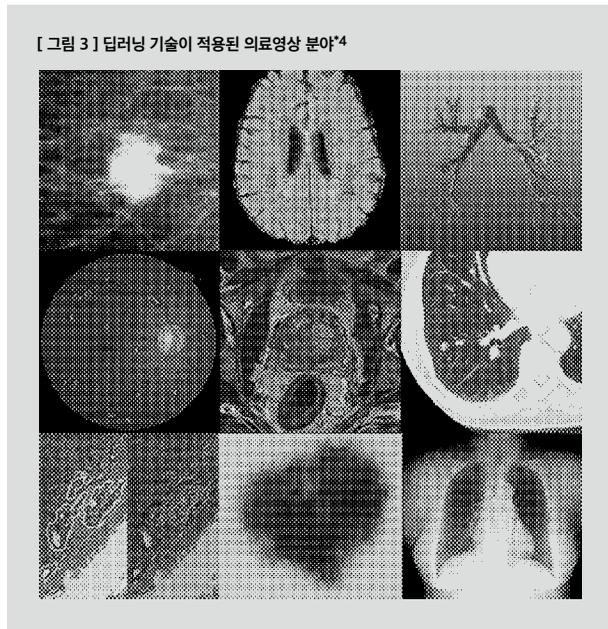
[그림2]의 위 도표는 연도별로 딥러닝 관련 기술을 활용한 의료영상 분석 연구들의 수를 나타냅니다. 2015년 이후로 딥러닝, 특히 CNN을 이용한 연구 논문들의 수가 가파르게 증가하는 것을 확인할 수 있습니다. 아래 도표는 딥러닝을 활용한 의료영상 분석 연구들에서 다루고 있는 과제의 빈도수를 나타내고 있습니다.

글 | 황상홍: shwang@lunit.io

인공지능 기술이 인류의 번영에 도움이 될거라 믿고 있는 엔지니어입니다. 오랜 시간에 걸쳐 축적된 지식의 혜택을 전세계에서 누구나 누릴 수 있게 된다면 얼마나 멋진가요. 그런 세상을 꿈꾸며 의료 분야의 인공지능을 연구하고 있습니다.

가장 많은 연구결과가 발표된 과제는 장기 혹은 특정 구조의 경계 검출이었고, 그 뒤를 이어 병변의 검출과 검사 단위의 분류 연구가 활발했습니다. 상대적으로 영상의 정합 연구는 그 수가 적었는데 의료영상 분석에서는 시차를 두고 촬영된 영상 사이의 변화가 진단에 있어 중요한 정보이기 때문에 앞으로 많은 연구 결과들이 나올 걸로 생각됩니다.

CNN을 활용한 의료영상 분석 연구들 중 가장 화제가 되었던 연구는 구글에서 2016년 발표한 당뇨병성 망막증의 진단⁵ 과 올해 스탠퍼드 대학에서 발표한 피부암 진단입니다⁶. 두 연구 모두 모델을 개발하는데 활용한 학습 데이터의 수와 학습된 모델의 성능으로 주목을 받았습니다. 공통적으로 약 10만 건 이상의 영상을 학습에 활용했는데 이는 기존의 딥러닝을 활용한 의료영상 분석 연구에 비해 훨씬 큰 규모의 학습 데이터입니다. 그리고 모두 숙련된 전문의 수준에 뒤지지 않는 분류 성능을 보였습니다. 특히 스탠퍼드 대학의 연구는 모바일 기기에 쉽게 탑재되어 활용될 수 있기 때문에 그 활용가치는 굉장히 크다고 평가받습니다. 구글의 연구에서 주목할만한 점은 약 6만 장의 학습 데이터 이상에서는 성능 향상이 없었다는 것입니다. 일정 수준 이상의 예측 능력을 확보하기 위해 필요한 학습 데이터의 수는 일반화하여 결론 내릴 수 없는 어려운 문제입니다. 하지만 의료영상은 대부분 통제된 상황에서 획득되는 만큼 일반적인 영상에 비해 데이터 간의 산포가 예측 가능하고, 그래서 이미지 넷에 비해 상대적으로 적은 수의 학습 데이터로도 좋은 일반화 성능을 보일 수 있습니다.



[그림3]에 보이는 다양한 의료영상 분야에서 현재 딥러닝 기반 모델링 방식이 가장 좋은 성능을 내는 것으로 알려져 있습니다⁴. X선 영상에서 CT, MRI, 나아가 병리 조직 영상까지 거의 대부분의

영역에서 딥러닝 기술의 도입이 아주 빠르게 진행되고 있고 좋은 예측 성능들이 보고되고 있습니다. 이렇게 빠르게 확산될 수 있는 이유는 앞에서 언급한 바와 같이 딥러닝 방식이 데이터로부터 주요 특징들을 스스로 학습하기 때문에 주어진 과제와 영상에 최적화된 특징을 직접 디자인할 필요가 없기 때문입니다. 즉, 충분한 양의 학습 데이터만 확보되면 바로 학습을 시작할 수 있고 꽤 높은 확률로 좋은 성능을 얻을 수 있습니다.

지금까지의 의료영상 분석 관련 연구들 대부분은 기술적인 관점에서 봤을 때 기본적인 지도학습의 범주에 속합니다. 다시 말해서 학습 데이터는 입력과 정답, 이렇게 쌍으로 주어지고 입력과 정답 간의 함수관계를 CNN이 학습하는 방식입니다. 앞서 살펴본 바와 같이 의료영상 분석 분야에서 다루고 있는 과제들이 일반적인 영상의 분석 과제들과 공통되기 때문에 전반적인 기술 발전의 흐름이 컴퓨터 비전 분야의 흐름과 매우 유사합니다. 그렇지만 연도별 연구의 수에서 알 수 있듯이 컴퓨터 비전 분야의 연구와 약간의 시간차를 두고 기술 개발이 이뤄지고 있는데, 이는 딥러닝 기반의 기술을 연구할만한 공개된 대규모 데이터를 의료영상 분야에서는 찾기 어렵기 때문입니다.

딥러닝을 위한 의료영상 데이터

딥러닝 기반의 인공지능 기술은 많은 수의 학습 데이터를 필요로 합니다. 그리고 기술 개발의 가속화를 위해서는 이러한 데이터가 연구자들에게 공개되어 쉽게 접근 가능해야 합니다. 현재 하루가 멀다 하고 새로운 시도와 기술이 쏟아지고 있는 컴퓨터 비전 영역의 연구결과들은 이미지 넷 데이터가 없었다면 그 속도가 매우 더뎠을 것입니다. 하지만 지금까지 의료영상 분석의 연구 대부분은 많아야 수천 개의 데이터를 활용했음에도 불구하고 상대적으로 좋은 성능을 보일 수 있었는데 이는 전이 학습(transfer learning)이라는 방식의 효과 덕분이었습니다⁷. 여기서의 전이 학습의 의미는 이미 이미지 넷 데이터 등으로 학습되어 있는 모델을 시작점으로 의료영상의 학습을 시작하는 것을 말합니다. 전이 학습이 효과적인 이유는 이미지 넷 데이터와 같은 일반적인 영상과 의료영상이 어느 정도 공통된 특징을 가지고 있기 때문입니다. 특히 모서리, 선분 등과 같은 낮은 차원의 특징들을 공유하기 때문에 이미 학습된 모델 전체를 시작점으로 삼지 않고 하위 레이어들만 가져와도 큰 효과를 얻을 수 있습니다⁸. 그렇다고 하더라도 역시 학습 데이터는 많으면 많을수록 좋습니다.

의료영상은 그 특성상 대규모의 학습 데이터를 개인 혹은 하나의 기관에서 확보하는 것이 일반적인 영상에 비해 더 어렵습니다. 이런 이유 아래 인공지능 기반 의료영상 분석 기술의

개발에 필수적인 대규모의 의료영상 데이터를 공개하는 시도들이 의료영상 분석 대회라는 형태로 이루어지고 있습니다. 이렇게 의료영상 데이터가 연구를 위해 공개되는 흐름에는 역시 영상의 분석에 있어 딥러닝이 탁월한 효과를 보이고 있기 때문입니다.

이 흐름의 대표적인 예가 작년에 시작된 The Digital Mammography DREAM Challenge⁹와 올해 열렸던 Data Science Bowl 2017¹⁰입니다. DREAM Challenge는 총 8만 4,000명 이상의 수검자들로부터 모은 64만 개 이상의 유방촬영영상을 학습하여 특정 유방촬영영상에 유방암으로 의심되는 조직이 존재할 가능성을 예측하는 대회입니다. 이전까지의 의료영상 관련 대회에서는 찾아보기 힘들었던 대규모의 데이터를 제공했고 이에 많은 연구자들의 관심을 끌었지만, 막상 뚜껑을 열어보니 실제 악성 병변을 가지고 있는 검사의 수는 수백 건에 불과했고 제공된 데이터를 직접 다운로드하여 사용하는 방식이 아니라 아쉬움이 남습니다. 이 대회는 1차 경쟁 단계(competitive phase) 및 2차 협업 단계(collaborative phase)로 나뉘어 있는데 현재 루닛은 1차 단계 결과 상위 8개 팀만을 초청하여 진행되는 2차 단계에 참여하여 과제를 수행하고 있습니다. Data Science Bowl 2017은 약 1,400건의 흉부 CT 촬영 영상을 이용하여 그 영상에 암조직으로 의심되는 병변이 존재하는지 여부를 예측하는 대회입니다. 제공된 데이터 중 실제 악성 종양이 포함된 케이스는 약 360건 정도로 앞서 유방촬영영상 데이터와 같이 정상 데이터에 비해 그 수가 적습니다.

공개데이터 관점에서 올해 가장 흥미로운 소식은 NVIDIA의 GPU 기술 학회(GPU technology conference, GTC)에서 스탠퍼드 대학이 발표한 Medical ImageNet 프로젝트입니다¹¹. 프로젝트의 이름에서 알 수 있듯이 의료영상으로 이미지넷과 같은 대규모 공개 데이터셋을 만들겠다는 내용입니다. 이 데이터셋은 의료영상뿐 아니라 연관된 유전체 데이터, 환자의 전자의무기록 등 상당히 넓은 범위를 모두 포함합니다. 데이터셋의 크기는 대략 0.5 페타바이트, 총영상의 수는 약 10억 건 정도로 예상되고 있습니다. 아직 구체적인 공개시기는 알려진 바 없지만 공개가 되면 관련 연구의 발전에 상당한 기여를 할 수 있을 걸로 기대됩니다.

위와 같이 우리에게 어느 정도 익숙한 방사선영상 이외의 또 다른 형태의 의료영상으로는 병리 조직 슬라이드를 스캔한 병리 영상이 있습니다. 병리 영상의 판독은 조직의 악성 유무와 전이 여부를 판단하고 이러한 진단 결과가 향후 환자의 치료 계획에 영향을 미치는 아주 중요한 행위입니다. CAMELYON 대회¹²는 이 병리 영상을 이용하여 유방암 전이를 검출하는 것을 목표로 작년부터 개최되어 오고 있습니다. 올해 대회에서는 1,000장의 슬라이드 영상이 참가자들에게 제공되었습니다. 2016년에 열린 유방암의 진행 정도를 예측하는 Tumor Proliferation Assessment

Challenge (TUPAC)¹³에서는 약 800장의 슬라이드 영상을 제공했고 이 대회의 총 세 가지 세부과제에서 루닛은 모두 1위를 기록하기도 했습니다.

위에 언급된 대회들에서 상위 성적을 얻은 참가팀들은 모두 딥러닝, 특히 CNN 기반의 모델을 이용합니다. 일반적인 영상 분석과 마찬가지로 의료영상에서도 대부분의 영역에서 현재 가장 좋은 성능을 보이는 기술은 딥러닝임을 다시 한번 확인할 수 있습니다.

의료영상 데이터의 특징과 향후 과제

의료영상은 일반적인 영상과 데이터 측면에서 확연히 다른 몇 가지 특징들을 가지고 있는데 이러한 특징들을 모델링 과정에 반영한 기술 개발이 앞으로 활발해질 것으로 예상됩니다. 이러한 기술은 새로운 알고리즘의 개발, 최적화된 네트워크 모형의 수립 등을 필요로 할 수 있습니다. 예를 들어, 2015년에 소개된 U-Net¹⁴ 은 적은 수의 영상을 이용하여 세포 경계를 검출하는데 효과적인 새로운 네트워크 모형이라고 알려져 있고, 이 연구에서 현미경 영상의 특징을 고려한 탄성 변형(elastic deformation) 방식으로 데이터를 생성하여 좋은 검출 성능을 얻었습니다.

의료영상에서 공통적으로 찾아볼 수 있는 몇 가지 특징들은 아래와 같습니다.

먼저, 대량의 데이터를 얻기 힘들고 게다가 지도학습에 필요한 레이블 정보를 얻는 건 더욱 어렵습니다. PACS가 도입된 이래 병원 내에서 촬영되는 의료영상들은 모두 저장되어 왔기 때문에 데이터 수는 굉장히 많습니다. 하지만 여러 제도적, 사회적 이슈들로 인해 이러한 데이터를 손쉽게 외부에서 접근하기 어렵고 실제 활용하기 위해서는 축적된 데이터를 학습에 활용할 수 있도록 정제하는 작업이 선행되어야 합니다. 이 부분이 많은 시간과 노력을 요합니다. 또한 숙련된 전문가만이 영상을 판독할 수 있기 때문에 지도학습에 필요한 레이블 정보를 얻는 것 또한 많은 시간과 노력이 필요합니다. 예를 들어, 객체 검출 방식을 통해 의료영상에서의 병변의 위치를 알아내고자 한다면 기본적으로 학습 데이터에 병변의 위치가 모두 표기되어 있어야 하는데 PACS에 있는 그대로의 의료영상들은 이러한 정보를 담고 있는 경우가 거의 없기 때문에 새롭게 병변의 위치를 표시해야 합니다. 데이터를 마련하는 데 있어 필요한 자원을 최소화하면서도 좋은 성능을 기대할 수 있는 방법론의 개발이 중요합니다.

다른 특징은 영상의 크기입니다. 2015년 이미지넷 대회의 모든 과제에서 압도적인 성능으로 1위를 차지한 Residual Network¹⁵는 영상을 분류할 때 짧은 변 기준으로 최대 640 픽셀

크기의 영상을 입력으로 받습니다. 반면 흉부 X선 영상은 한번의 크기가 2000픽셀 이상이고 유방촬영영상의 경우 4000픽셀이 넘습니다. 또한 병리 조직 세포를 스캔한 병리 영상의 경우는 한번의 크기가 10만 픽셀보다 큰 경우가 대부분입니다. 만약 영상단위로 분류를 하고자 한다면 상당히 많은 계산 자원을 필요하게 됩니다. 정보의 손실 없이 주어진 데이터를 활용하기 위하여 이러한 엄청난 크기의 입력 영상을 효율적으로 처리할 수 있는 알고리즘의 개발이 필요합니다.

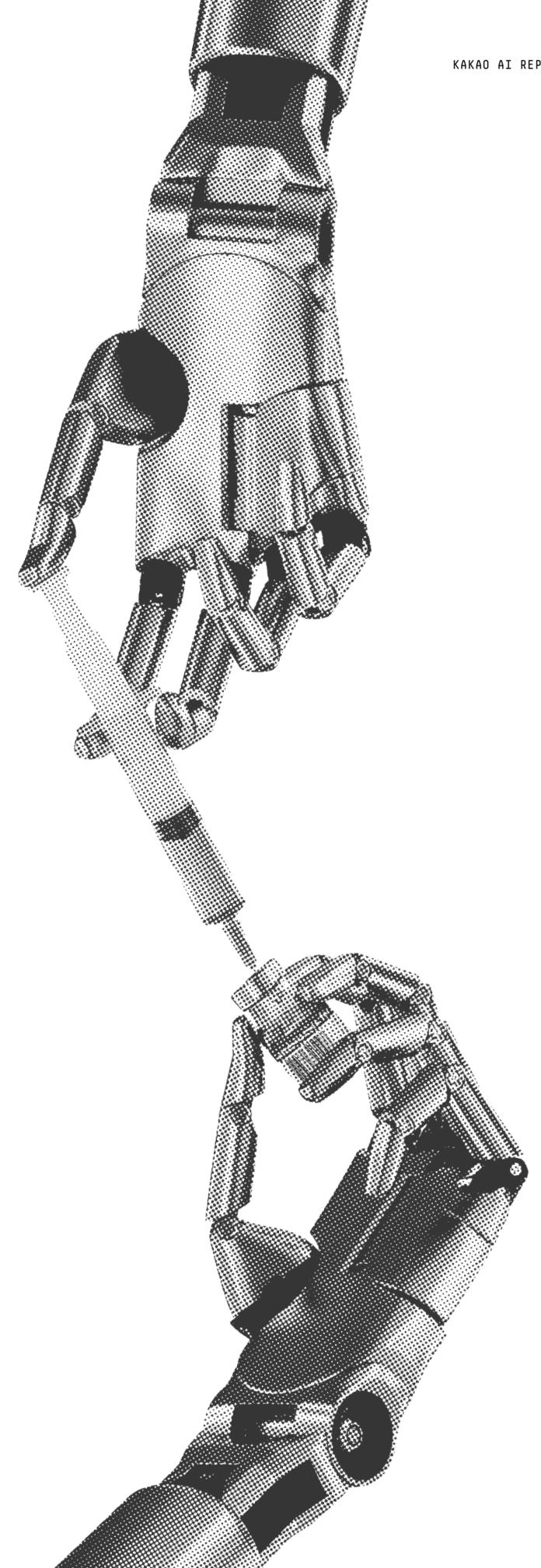
또 다른 특징은 영상에 존재하는 객체의 크기입니다. 일반적인 영상에서 그 영상의 클래스는 특정 객체의 유무로 판단하게 됩니다. 예를 들어, 고양이클래스에 속한 영상은 고양이라는 객체를 가지고 있습니다. 의료영상에서는 보통 정상과 비정상 영상을 구분하는 것이 목적이기 때문에 비정상 병변이 영상의 클래스를 결정하는 객체라고 볼 수 있습니다. 의료영상은 이 객체의 크기가 일반적인 영상에 비해 상대적으로 굉장히 작은 경우가 많습니다. 물론 주어진 데이터안에 존재하는 모든 객체들의 위치를 알고 있다면 문제없겠지만 실제로 그런 데이터를 수집하는 것은 현실적으로 많은 자원을 요구하기 때문에 한정된 정보를 활용하면서도 이렇게 작은 객체를 잘 검출할 수 있는 기술을 필요로 합니다.

언급된 것들 이외에 주어진 의료영상의 고유한 특징을 학습 과정에 반영시키는 방향으로 기술 개발이 이루어진다면 데이터 관점에서 경제적이면서 좋은 성능을 보이는 모델을 얻을 수 있을 것입니다.

마치며

현재까지 보고되고 있는 인공지능 기술을 활용한 의료영상 분석에 관한 연구들을 보면 그 결과가 놀랍습니다. 소개드린 바와 같이 수년간 수련한 전문의 수준의 진단 성능을 데이터로부터 학습된 모델이 보여주고 있으니깐요. 하지만 실제로 널리 활용되기까지는 더 광범위하고 다양한 검증을 거쳐야 합니다. 의료라는 산업의 특성상 실제로 활용되기 위해서는 인공지능 시스템이 좀 더 예측 가능해야 하고, 다양한 상황하에서도 일관된 성능을 보여야 하며, 예측한 결과에 대한 최소한의 해석이 가능해야 합니다. 앞으로 이런 관점에서의 기술 개발과 함께, 개발된 시스템에 대한 많은 임상 연구들이 발표되기를 기대합니다.

***1** 논문 : Lodwick, G. S. (1966), "Computer-aided diagnosis in radiology: A research plan", Investigative Radiology, 1, 72-80. ***2** 참고 : https://en.wikipedia.org/wiki/Picture_archiving_and_communication_system ***3** 논문 : Doi, K. (2007), "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential", Computerized Medical Imaging and Graphics, 31, 198-211. ***4** 논문 : Litjens, G. et al. (2017), "A survey on deep learning in medical image analysis", arXiv:1702.05747v2. ***5** 논문 : Gulshan, V. et al. (2016), "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs", The Journal of the American Medical Association (JAMA), 316, 2402-2410. ***6** 논문 : Esteva, A. et al. (2017), "Dermatologist-level classification of skin cancer with deep neural networks", Nature, 542, 115-118. ***7** 논문 : Shin, H.-C. et al. (2016), "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning", IEEE Transactions on Medical Imaging, 35(5), 1285-1298. ***8** 논문 : Hwang, S. and Kim, H.-E. (2016), "A novel approach for tuberculosis screening based on deep convolutional neural networks", In Proceedings of SPIE Medical Imaging, 9785, 97852W-1. ***9** 참고 : https://www.synapse.org/Digital_Mammography_DREAM_Challenge ***10** 참고 : <https://www.kaggle.com/c/data-science-bowl-2017> ***11** 참고 : https://gputechconf2017.smarteventcloud.com/connect/sessionDetail.ww?SESSION_ID=110157 ***12** 참고 : <https://camelyon17.grand-challenge.org/> ***13** 참고 : <http://tupac.tue-image.nl/> ***14** 논문 : Ronneberger, O., Fischer, P. and Brox, T. (2015), "U-Net: convolutional networks for biomedical image segmentation", Medical Image Computing and Computer-Assisted Intervention (MICCAI), 9351, 234-241. ***15** 논문 : He, K. et al. (2015), "Deep residual learning for image recognition", arXiv:1512.03385.



의료와 AI 신기술의 융합 : 과제와 전망

의학은 본질적으로 윤리적이며 보수적이다. 질병에 대한 불안정한 이해와 제한된 치료법으로 인간의 생명을 다루기 때문이다. 이러한 한계로 인해 의료 산업은 규제 산업의 성격을 가진다. 이는 신기술이 의료에 적용되는 진입 장벽으로 작용했다. 윤리적 문제와 부작용을 무시한 채 선부르게 신기술을 적용하는 행위는 의료의 영역에서는 납득되기 어렵다. 줄기세포가 대표적인 사례다.

줄기세포 외에도 기존의 미충족수요(unmet needs)를 해결하는 방법으로 적극적으로 도입이 고려됐던 신기술들이 알려지지 않은 부작용이나 불안전성(instability) 등으로 인해 의료 산업에서 퇴출된 예는 셀 수 없이 많다. 일례로 설측 교정을 들 수 있다. 설측 교정은 치아의 안쪽 면에 교정 장치를 붙여서 교정 장치가 보이지 않게 교정하는 방법이다. 웃을 때 교정 장치가 보이는 것에 대해 치아 교정에 거부감을 갖고 있던 환자들(주로 10대)을 겨냥해 1980년대 고안됐다. 개발 당시에는 기대치에 못미치는 효과때문에 설측 교정은 의료 현장에서 배제되었다. 1990년대에 초탄성을 가진 니틴올(nitinol)의 도입되기 전까지는 의료 산업에 적용되지 못하다가, 일반 교정의 효과와 유사해지면서 설측 교정은 가격이 비싼 단점을 지녔음에도 임상 현장에서 적극적으로 채택되고 있다.

이처럼 의료와 신기술의 융합은 많은 난제를 갖고 있지만, 수많은 임상 현장의 미충족수요를 해결하기 위해 더욱 적극적으로 활용되어야 한다

글 | 김남국 namkugkim@gmail.com

공대를 나와서 의료영상매개 연구에 뛰어들지 어언 20년. 팔자에 없는 의대교수를 하고 있다. 의료 분야에 더 똑똑한 연구자들이 더 많이 들어올 수 있도록, 가고 역할을 하고 싶다. 언젠가 내가 활활 떠날 수 있기 위해서!

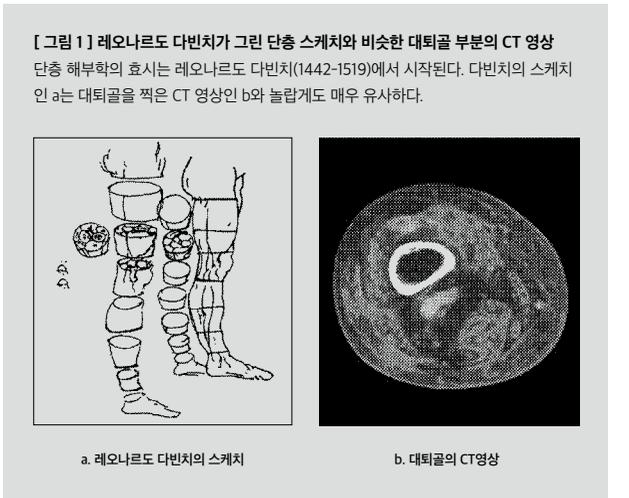
의료영상의 역사

의료영상은 빌헬름 린트겐(Wilhelm Conrad Rontgen)이 1895년 엑스레이(X-ray)를 발견할 때부터 시작되었다고 할 수 있다. 린트겐은 엑스레이를 발견한 공로로 1901년(노벨상 원년)에 노벨물리학상을 수상했다. 처음에는 기술의 미비로 인하여 병원에서 X선관(X-ray tube)과 필름 등을 의료 현장에 적용하기 어려웠다. 하지만, 린트겐이 엑스레이를 발견한 지 6개월도 채 안 되는 시점에, 한 정형외과 의사가 뼈가 부러진 환자를 엑스레이로 촬영했다¹. 엑스레이의 ‘보급’은 다른 문제였다. 의료 현장에서 엑스레이를 활용하는 데 필요한 X선관과 필름 등을 구입하는 것이 녹록찮았기 때문이다. 이런 어려움은 한 현장의 의사가 자신의 친구인 지멘스(Siemens)에게 전하면서 상황은 바뀐다. 지멘스는 세계에서 최초로 의료용 X선관을 상업화했다. 이는 엑스레이가 의료현장에 널리 퍼지는 계기가 되는 동시에, 지멘스가 의료영상 산업 1위, 의료기기 전체에서 2위로 부상하는 시발점이 된다².

엑스레이 외 현재 대표적 의료영상 기술은 물리학적 이론에 기반을 두고 있다. 초음파 영상장비는 2차 세계대전 때 잠수함을 찾기 위한 목적에서 개발됐던 초음파를 사람 몸에 적용한 기술이며, CT로 불리는 전산단층촬영 장비(computed tomography)는 엑스레이의 조직 흡수율 차이를 활용한 것이다. 자기공명영상(magnetic resonance imaging, MRI)은 몸 안의 수소 원자핵을 공명화시켜 그 원자핵 스핀의 고유한 특성을 컴퓨터를 활용하여 영상화시키는 것이다. 앞선 일련의 의료영상 장비들은 1970년대에 개발됐다.

CT와 MRI는 10여개의 노벨 물리학상, 화학상을 받은 원천 기술을 기반으로 하고 있다. 이들 의료기기를 개발한 연구자들도 노벨 생리의학상을 받았다. CT는 1971년에 개발됐는데, 개발된 지 8년 후인 1979년 노벨상 수상의 근거가 될 정도로 의료현장에서는 혁명적인 발명으로 인식됐다. 이러한 높은 평가는 CT가 의학에서 공간의 문제를 해결한 결과로 풀이된다.

CT와 MRI와 같은 단층촬영 영상장비는 고유한 대조(contrast) 매커니즘에 따라 영상을 생성한다. 최근에는 환자의 해부학적인 구조 뿐만아니라, 혈류의 흐름을 볼 수 있는 관류강조영상(perfusion weighted image), 조직의 미세구조를 볼 수 있는 확산강조영상(diffusion weighted image), 폐의 공기의 흐름을 볼 수 있는 대류영상(ventilation image), 4D 유동강조 MR영상(4D flow MRI) 등과 같이 다양한 기능적 의료영상 기법이 의료 현장에 적용되고 있다. 이러한 장비를 통해 확보된 영상의 정확한 분석을 위해서는 물리학과 의학적 관점이 수반된다. 단층촬영 영상장비의 보급은 단층 해부학의 발생으로 연결된다.



의료영상 분석기술의 역사

초기에 필름을 사용하던 엑스레이 촬영 장치를 제외하고, 컴퓨터 방사선(computed radiography, CR), 디지털 방사선 촬영술(digital radiography), CT, MRI, 양전자단층촬영(proton emission tomography, PET) 등 대부분의 의료기기는 디지털로 영상을 처리하고 있다.

하지만, 이전에는 병원 내에 영상들을 효율적으로 유통할 수 있는 인프라가 없어서, 필름을 이용하여 병원 내에서 사용하였다. 1990년대 들어오면서 상황은 바뀌었다³. 컴퓨터 통신을 이용하여 의료영상을 병원내에서 공유할 수 있게 하는 영상저장통신장치(picture archive and communication system, PACS)가 개발되면서, 병원 내에서 누구나 디지털 의료영상을 사용할 수 있게 되었다. 이 때부터 본격적인 의료영상 분석기술이 개발되고 적용되기 시작했다. 1990년대 들어와서 시카고 대학 등이 유방 엑스레이 영상(mammography) 분야에서 컴퓨터보조진단(computer aided diagnosis, CAD)기술을 개발하였다. 이를 기술이전하여 상업화한 R2라는 프로그램⁴이 미국 식약처(food and drug administration, FDA)에서 유방 엑스레이 영상 판독시에 두 번째 판독 의사(second opinion reader)로 사용될 수 있다는 인정을 받게된다. 이는 세계 최초로 컴퓨터프로그램이 영상의학과 의사 대신 판독을 하고, 그 대가로 보험수가를 받을 수 있게 된 사례이다.

의료 영상에는 해부학적 다양성 뿐만 아니라, 질환의 다양성까지 포함하고 있다. 이런 자연적인 다양성을 극복할 수 있는 결정론적(deterministic)인 알고리즘을 개발하는 것은 매우 많은 가정, 뛰어난 개발자, 그리고 기술을 이해하는 의료진과의 협업 등이 필요하다. 추가적으로 생각해 볼 수 있는 방법은 통계학적인 방법이다. 통계적 방법은 (해부학적 및 질환의) 다양성 등을 통계적으로는 정확하게 모델링하기가 어렵고,

모델의 변이(standard deviation)가 커서 개별 환자의 임상진료에 쓰이기에는 부적합하다.

또, 다른 해결 방법은 의사가 해부학을 기반으로 다양한 변이를 이해하듯이, 여러 변이를 가지고 있는 장기를 미리 분할해 놓고, 새로운 데이터가 오면 이를 정합(registration)해서 가장 잘 정합되는 것을 찾아서 분할하는 기법(multi-atlas registration)이다. 이는 다양한 변이를 반영할 수 있다는 장점이 있지만, 계산해야 할 양이 많이 계산에 시간이 매우 오래 걸린다. 결과적으로 정합 알고리즘의 정확도에 기반하고 있어서 일반적으로 적용되기 어렵다.

최근 딥러닝(deep learning)이 보여주는 성능과 속도는 실제 의료에 쓰일만큼 강인하고, 정확하다고 사료된다. 영상의 분류뿐 아니라, 검출(detection), 분할(segmentation), 정규화(normalization), 내용기반질의(content based image retrieval) 등 다양한 분야에 좋은 성과를 내고 있다. 하지만, 지금도 의사들이 처음 보는 희귀하고 다양한 질환이 의학저널에서 사례발표(case report) 형식으로 출간되고 있다.

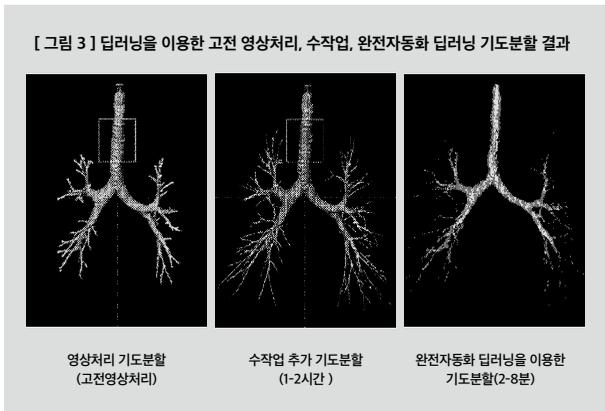
이렇듯이 의료영상에 있는 질환의 다양성과 희소성을 극복해야 한다. 또한, 영상의학 의사가 임상에서 하는 작업을 분류해보면, 약 5만가지 정도로 구분된다고 한다. 데이터 자체가 희소하거나, 구하기 힘든 점을 차치하고라도, 지도학습 방법으로 하나하나 잘 만드는 것도 쉽지 않을 것이다. 오히려, 의학의 근본 원리인 해부학, 생리학, 병인학(Etiology), 병리생리학(Pathophysiology) 등을 기반으로 뼈대를 세우고, 이런 의료영상의 다양한 현상을 보지 않으면, 다양한 질환의 변이에 잘 대응하기 어려울 것이다. 최근 딥러닝에 다양한 기존 방법이나 복잡한 네트워크가 적용되는 것이나 원샷 학습(one-shot learning) 등이 개발되는 것은 이런 특성을 반영하는 것이라 할 수 있다.



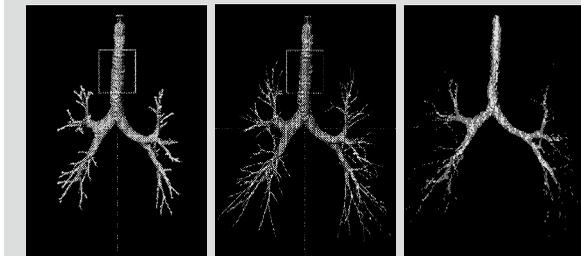
[그림 2] 디지털 의료영상(CT)

플랫폼과 인공지능

의료 인공지능 과정에서 95%는 고전 영상처리기법 및 다양한 수작업을 통해서 데이터 클리닝(cleaning), 5%는 인공지능이 담당한다. 이중 95%를 차지하는 의료영상 전·후 처리의 질을 높이고, 효율적으로 만드는 데 매우 큰 영향을 준다. 그런데, 인공지능이 학습할 의료 데이터를 전처리하거나, 정답을 만드는 것은 전문적 영역의 일이다. 따라서, 이 과정에 의학 지식을 가지고 있는 의료진의 협력은 필수적이다.



[그림 3] 딥러닝을 이용한 고전 영상처리, 수작업, 완전자동화 딥러닝 기도분할 결과



영상처리 기도분할 (고전영상처리)

수작업 추가 기도분할 (1-2시간)

완전자동화 딥러닝을 이용한 기도분할(2-8분)

어떤 문제는 정답을 인간이 수작업으로 만드는 것이 현실적으로 불가능하기도 하다. 대표적인 것이, CT영상에서 기도(airway) 분할이다. 기도 벽의 두께는 폐의 염증 반응을 알 수 있는 대표적인 표지자(biomarker)이기 때문에, 기도 벽의 두께를 정량화하는 것이 의료에서는 매우 중요하다. 하지만, 기도는 폐안에 프랙탈 형태로 분화되어서 실제 CT 영상에 무수히 많은 폐기도가 보이지만, 계속 가늘어져서(tapering) CT영상에서 부분용적효과(partial volume effect)를 만들고, 호흡이나 심장의 움직임 때문에 영상에서 연속성이 유지되지 않는 문제(motion artifact)가 발생한다.

이 문제를 해결하는 것은 고전적인 영상처리 기법으로는 가장 어려운 난제중에 하나였다. 환자 1명의 CT에서 기도의 통로(lumen)를 전문가가 수작업으로 그리면 대략 1주일정도 걸린다. 수작업으로 진행되기 때문에, 완벽하게 통로만 그리지 못할 뿐더러, 기도벽을 조금씩 침범할수 밖에 없고, 이를 완벽하게 고치기도 매우 어렵다. 고전영상처리기법으로 모든 기도의 통로를 분할하고, 손으로 점을 찍으면 끊어진 기도들을 연결하는 프로그램이 수년에 걸쳐 개발됐다. 이제는 어느 정도 시간(1~2시간)만 들이면, 반자동으로 거의 모든 기도 통로를 일관적으로 분할할 수 있게 되었다. 높은 수준의 분할결과(mask)를 정답으로 학습하여, 전자동으로 기도를 2~8분 정도의 짧은 시간안에 분할할 수 있게 된 것이다. [그림 4]에서 볼수 있듯이, 학습이 잘되면, 인공지능이 자동으로 만든 기도 분할 결과가 사람이 만든 정답보다 더 좋은 것을 알 수 있다. 이런 성과는 여러가지

요소가 결합되어서 가능한 것이다. 그러나 기도 분할 프로그램을 개발해서, 이전에 비해 질적으로 다른 수준의 정답을 만들었다는 것이 가장 중요하다.

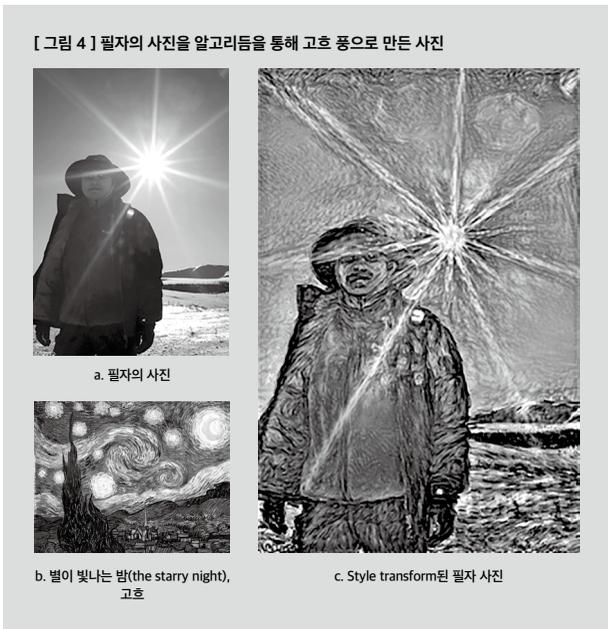
이런 좋은 정답을 만들수 있는 플랫폼을 장악한 회사가 중요하기 때문에 다빈치와 같은 수술로봇을 인공지능으로 자동화하고 싶다고 하면, 인공지능 회사가 로봇을 만드는 것 보다는, 의료로봇 회사가 인공지능을 배우는 것이 더 빠르고 쉬울 것이다. 인공지능 기술을 의료에 적용하기 위해서는 이미 수준 높은 제품을 가지고 있는 의료 업체와 협력하여, 그 제품을 지능화 또는 강화(augmentation)하는 것이 보다 실효적일 것이다. 동일한 맥락에서, 인공지능 회사와 기존의 플랫폼을 가지고 있는 회사들간의 협력이나 인수합병(M&A)이 향후에 보다 활발하게 일어나야 할 것이다.

결어

인공지능의 의료영상 적용이 혁명적이라는 것에는 이견이 없다. 하지만, 영상(image)이라는 말이 환상(imagery)이라는 말과 어원이 같다는 점을 명심해야 한다. 영상장비의 기계적 오류 외 인공지능 자체의 메커니즘적 한계도 상존하는 것이 현실이다. 인공지능을 임상에 적용해서 가치 있는 지능형 의료장비를 개발하기 위해서는, 인공지능 기술 이외에도 다양한 기술을 융합해야 한다. 지능형 의료영상처리가 본격적으로 도입되기 시작하면, 의료의 많은 것이 바뀔것이다. 내가 어렸을때는 많은 사람들이 계산능력과 수학을 동일시해서, 계산능력을 키울 요량으로 주산이나 부기학원을 다녔다. 하지만, 엑셀(excel)이나 계산기가 나온후에는 누구도 엑셀보다 계산을 잘할 수 없고, 지금은 누구도 계산능력과 수학을 동일시하지 않는다. 즉, 수학은 계산능력과는 다른 어떤 것이 되었다. 마찬가지로, 의료에서도 단순하고 귀찮지만 의사들만이 할 수 있는 수많은 잡일이 있다. 인공지능이 이런 귀찮은 단순한 일을 대체할 수 있다면(의사와 동일한 수준에서), 의사들은 단순 노동에서 해방되고, 보다 수준 높은 의료에 집중할 수 있게 될 것이다. 결과적으로, 의학은 지금과는 다른 어떤 것이 될것이다. 인공지능 혁명은 단지 의사의 대치가 아니라, 새로운 의료에 대한 필요성과 이를 이룰 수 있는 도구를 줄 것이다.

내가 좋아하는 화가 반 고흐(Vincent Willem van Gogh)는 여러가지 질병을 앓았지만, 특히 측두엽 이상으로 인한 하이퍼그래피아(hypergraphia)라는 글쓰기 증동병에 시달렸다. 그는 자살하기 전 10여년(1880-1890)동안, 2,000점 이상의 그림과 스케치와, 동생 테오에게 보낸 1,700페이지의 서신을 남겼다. 그가 10년 동안 추구한 것은 자기만의 스타일이었고,

누구나 고흐 그림을 구분할 수 있을 정도의 독자적인 화풍을 만들었다. 하지만, 구글의 ‘딥드림(deep dream)’ 프로젝트는 고흐의 ‘별이 빛나는 밤(the starry night)’ 작품의 스타일 모사(style transform)를 통해, 어떤 사진이라도 ‘고흐 풍’으로 바꿀 수 있다.



[그림 4] 필자의 사진을 알고리즘을 통해 고흐 풍으로 만든 사진

a. 필자의 사진

b. 별이 빛나는 밤(the starry night), 고흐

c. Style transform된 필자 사진

예술의 본질은 사람에게 감동을 주는 것이다. 나는 이런 것이 새로운 예술이 아닐까 한다. 지금은 대부분의 사람들이 예술에 소외되어 있다. 앞으로 인공지능 기술에 힘입어 더 많은 사람들에게 맞춤형 감동을 줄 수 있을 것이다. 역시, 고흐 그림에는 태양이 있어야 한다!

*1 설명 : 기술 도입의 측면에서 의료는 보수적인 영역으로 간주하지만, 사람의 생명을 살리기 위한 목적에서 충분한 검증을 거치지 않은 새로운 기술이라도 가치가 있을 것이라 생각되면, 매우 빨리 적용되기도 한다. 생명을 살리는 데 유용하다고 판단되는 신기술을 도입시키는 것 역시 비윤리적이라는 평가를 받을 수 있다. *2 설명 : 의료 기기는 대개 최초로 제품을 만드는 회사가 해당 분야의 시장의 대부분을 차지한다. *3 설명 : 이보다 훨씬 전인 컴퓨터와 의료영상 도입의 초창기인 70년대에, 시카고 대학에서 X-ray 필름을 스캔해서 폐결절을 찾는 프로그램을 개발한 것 등의 시도가 있었다. *4 설명 : 이 프로그램은 위양성(false positive rate, FPR)이 많아서, 실제 판독 시간을 길게 할 뿐더러, 재검사율(recall rate)을 증가시키고, 의료 비용을 증가시키는 등, 의료의 질적 향상이 없다는 학술적 평가를 받기도 한다. 이 제품은 보험수가 책정 측면에서도 성공했다고 보긴 어렵다. 이미 보험수가 책정되어 있는데도, 미국내에서 임상적으로 쓰는 병원이 2%, 가끔 쓰는 병원이 50%, 한번도 안써 본 병원이 48%가 나오기도 했다. 여기서 알 수 있는 것은 CAD와 같이 의사의 판독을 대체하는 경우는 의학적 정확도와 강건함(robustness)이 매우 중요하고, 동시에 실제 임상현장에서 생산성과 의료의 질적 향상이 이루어져야 한다는 사실이다.

AI 의료영상 기술 활용 사례

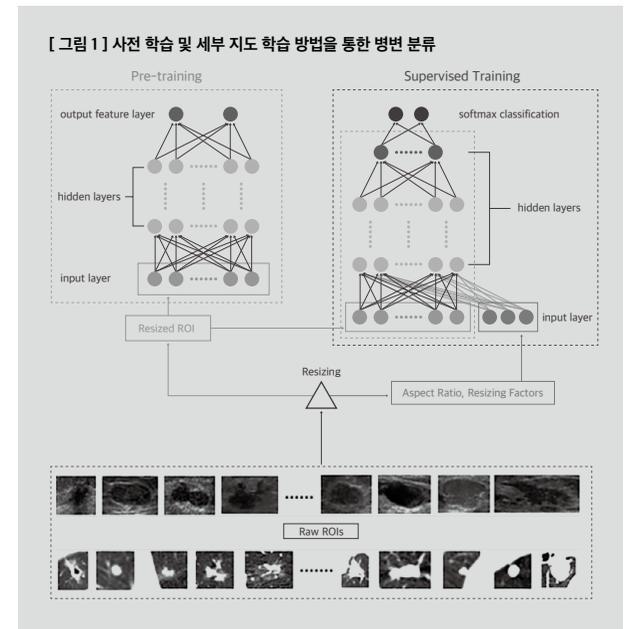
최근 딥러닝을 중심으로 기계학습 기술이 인공지능의 주요 방법론으로 자리 잡으면서 의료 영상분석에서도 기존의 임상적, 경험적 지식이나 규칙에 근거한 특징 추출 및 분석 방법에서 데이터 기반(data-driven)의 일관적이고 객관적인 특징 학습 및 분석 방법론으로 패러다임 이동이 가속화되고 있다. 특히, 일반 영상(natural image)에서 연구되는 다양한 기법들이 의료영상 분석에 적용되는 간격이 점차 짧아지면서 단순히 자연 영상에서 개발된 모델을 의료 영상에 그대로 적용하는 수준이 아닌, 의료 영상의 특성을 고려한 새로운 방법론들이 속속 발표되고 있다. 더 나아가 기존의 임상적 지식을 통합하거나 의료에서 반드시 필요한 해석가능성(interpretability)에 관한 연구들이 진행됨에 따라 실제 임상에서 활용 가능성이 높은 결과들이 등장하기 시작하였다. 본 글에서는 최근에 개발되고 있는 다양한 인공지능 기술들이나 연구 주제들이 의료영상 분석에 실제로 어떻게 적용되고 있는지 다양한 사례를 통해서 살펴보고 앞으로의 방향성과 남은 과제에 대해 논의해보고자 한다.

인공지능 - 의료영상 분석의 주류로

의료영상 분석에서 기계학습 방법은 특정 병변의 검출 및 분류, 인체 기관의 세부 구조 분할, 영상 간의 정합, 유사 영상 검색 등 다양한 영역에서 활용되어 왔다¹. 특히 딥러닝 모델 중 영상인식에 주로 사용되어온 컨볼루션 신경망(convolutional neural network, CNN)은 그 구조가 제안된 지 얼마 지나지 않은 1993년 무렵부터 폐결절 검출이나², 유방조직 미세석회화³ 검출 등 의료영상 분석에 적용된 바 있으나 데이터의 규모나 모델의 크기 및 학습 방법, 그리고 연산 자원의 한계로 실험적인 수준에 머물렀다. 이러한 가운데, 딥러닝이 영상인식에서 새로운 성능적 돌파구를 마련하면서 의료영상에서도 다시 딥러닝이 적극적으로 도입되기 시작하였다. 특히 ImageNet 대회에서 2012년 CNN을 사용한 팀이 큰 격차로 우승⁴한 이후 2015년에 이르러 간접적 비교에서 인간의 영상 인식 수준을 넘어서는 결과를 보인 것과 같이 의료영상에서도 2012년 ICPR(international conference on pattern recognition)의 유방 병리 영상 내 유사분열세포 검출에서 CNN기반의 모델이 우수한 이후⁵, 최근 구글이 JAMA(the journal of the American medical association)에 발표한 안저영상 기반 당뇨병망막병증(diabetic retinopathy, DR) 검출⁶, 스탠퍼드 대학에서 Nature에 발표한 피부암 분류⁷ 등 전문의의 수준에 준하거나 이를 넘어서는 결과들이 등장하면서 큰 반향을 일으킨 바가 있다.

의료영상의 데이터적 특성과 해결방안

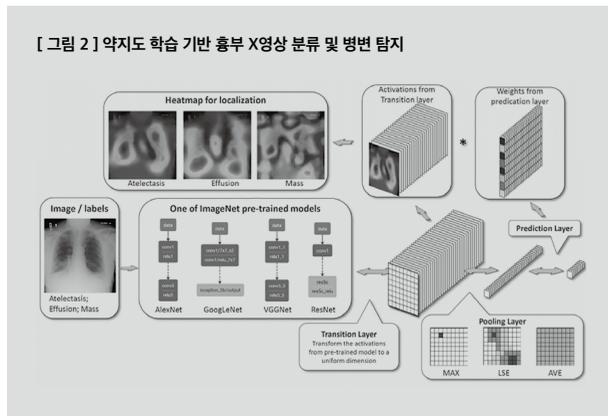
이렇게 딥러닝이 의료영상 분석에 본격적으로 적용되어 오면서 일반 영상에 비해 의료영상이 가지는 다양한 특성들을 고려할 필요성이 있었는데, 이 중 가장 중요한 점은 바로 의료영상의 데이터 측면의 특성이다. 특히 의료영상은 입력 영상의 수는 충분하지만 전문가의 판독이나 표식(label)을 획득하는데 많은 비용과 시간이 드는 경우가 빈번하여 이를 해결하기 위한 다양한 방법들이 제안되어 왔다. 가장 기본적으로, 비지도 학습(unsupervised learning)을 통해 병변의 중요한 영상적 특징을 추출하도록 학습하고, 이를 소수의 전문가 판독 결과를 바탕으로 지도 미세조정(supervised fine-tuning)하는 방법이 있는데, [그림 1]과 같이 유방 초음파와 흉부 CT에서 병변에 대해 양성, 악성에 대한 판단을 위해 SDAE(stacked denoising auto encoder)로 사전학습을 수행하고 수백 개 수준의 병변에 대한 판독 결과와 병변 영역의 크기 및 비율에 대한 정보를 바탕으로 추가 학습을 한 결과 기존의 컴퓨터 보조 진단(computer-aided diagnosis, CADx) 시스템 대비 높은 성능을 보임을 발표한 사례⁸가 있다.



딥러닝의 장점 중 하나는 이미 학습된 모델을 재사용하여 다른 영역의 데이터에 대해 추가로 학습하는 전이학습(transfer learning)이 가능하다는 점인데, 이는 처음부터 새롭게 학습하는 경우에 비해 학습의 속도가 빨라지거나 최종 모델의 성능이 높아지는 효과를 가져온다고 알려져 있다. 특히 이러한 방법은 새로운 영역의 데이터가 부족할 때 더욱 유용하며, 의료영상에서도 일반영상에 대해 이미 학습된 모델을 가져와 의료영상으로 추가 학습시키는 방법으로 모델의 성능을 개선하는 연구들이 수행되어 왔다. 미국 국립보건원(national institute of health, NIH)에는 흉복부 림프절(thoracoabdominal lymph node)를 검출하기 위한 컴퓨터 보조 검출(computer-aided detection, CADe) 시스템을 개발하기 위해 이미지넷에서 이미 학습된 잘 알려진 네트워크를 가져와 추가 학습함으로써 기존 성능을 넘어서는 결과를 얻었으며⁹, 애리조나 주립대와 메이요 클리닉(Mayo clinic)의 연구팀은 대장내시경 영상에서 용종을 검출하는 CADe 시스템 개발을 위해 다양한 조건에서 전이 학습이 최종 성능을 높이는데 도움이 된다는 점을 확인¹⁰하였다. 앞서 언급한 구글의 당뇨병망막병증 판단 모델이나 스탠퍼드 대학의 피부암 분류 모델도 일반영상에서 이미 학습이 된 구글의 Inception v3 모델¹¹을 바탕으로 의료 영상에 추가 학습하는 방법을 택하였다는 공통점이 있다.

글 | 정규환 khwan.jung@vuno.co

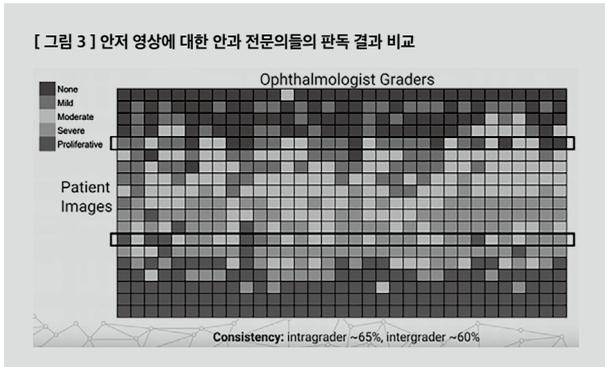
포스텍 산업경영공학과에서 최적화와 기계학습으로 박사학위를 받고 SK텔레콤, SK플래닛, 삼성전자종합기술원을 거쳐 현재는 뷰노의 공동창업자이자 CTO로 일하고 있다. Shallow 아키텍처로 학위를 받았지만 지금은 그걸 까야 먹고사는 형편이 되었고, 가장 무서워하는 곳이 병원이었지만 정신을 차려보니 어느덧 가장 자주 가는 곳이 되어 있었다. 배워서 남주자는 마음으로 살고 있기 때문에 몸은 지쳐도 배우는 것이 많은 지금을 감사하게 생각하고 있으며 우리나라 의료 발전에 작은 기여라도 할 수 있기를 소망하고 있다.



한편, 의료 영상 판독 시스템 개발을 위해 병변의 위치에 대한 표시 정보가 없는 상황에서 영상의 전체에 대한 소견만을 바탕으로 병변의 위치를 검출해내야 하는 약지도 학습(weakly-supervised learning)도 주요 문제 중 하나이다. 대표적인 방법들은 분류 모델이 예측 값을 내는데 가장 큰 영향을 미치는 입력 영상 내 영역을 역추적하는 방법으로서, [그림 2]에 도식화된 NIH의 최근 연구¹²가 이에 속한다. 해당 논문에서는 10만 장 가량의 흉부 X-선 영상의 판독문에서 8개 흉부 질환에 대한 분류 결과를 추출하고 해당 영상을 기반으로 이를 예측하도록 학습시키는데, 기존 모델의 활성화(activation)값이 공간적인 정보를 가지도록 하는 전이층(transition layer)으로 변환하고, 전이층의 여러 특징 지도들을 통합(pooling)한 이후 가중 합(weighted sum)하여 최종 예측값을 산출하도록 변경하였다. 이렇게 되면, 최종 예측값 산출에 대한 각 공간적 특징 지도 별 기여도를 알 수 있고, 해당 특징 지도의 활성화 값을 통해 입력 X-선 영상의 중요 영역을 추적하여 영상에 대한 분류뿐 아니라 병변의 위치도 검출할 수 있다.

위에서 전문가의 판독이나 표시가 부족한 경우에 이를 해결하기 위한 방안들을 소개하였지만, 궁극적으로 딥러닝 모델은 데이터 기반의 접근 방법으로서, 대량의 학습 데이터가 주어진 경우 그 잠재력이 가장 잘 활용될 수 있다. 하지만 데이터 기반 방법론의 특성상 데이터에 품질에 따라 그 성능이 크게 좌우될 수 있다는 점에서 주의를 요하게 되는데, 특히 의료에서 판독의 일관성이나 일치도가 낮은 질환의 경우 이를 고려한 데이터 수집 방법의 수립이 반드시 필요하다. 아래 [그림 3]은 안저 영상에 대해서 미국 안과 전문의들의 판독 결과를 비교한 것인데, 특정 환자에 대해서는 의사마다 5가지 중증도 모두에 걸쳐 판독 결과가 존재할 정도로 일치도가 낮다. 따라서 앞서 설명한 구글은 논문⁶에서는 하나의 영상에 대해 다수의 의사가 판독하도록 하고, 이들의 다수결을 정답으로 학습하도록 하여 일관성을 높이는 노력을 하였다. 스탠퍼드 대학의 피부암 분류 논문⁷에서는 피부질환의 분류체계(taxonomy)를 정의하여 의사간의 세부적인 질환 분류의

불일치가 최종적인 대분류 결과에 영향이 덜 미치도록 하는 방법을 선택하는 방법을 취하기도 하였다.

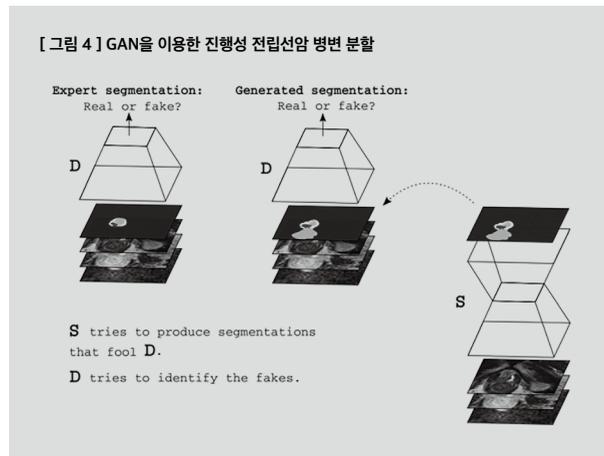


끝으로 의료영상의 데이터 측면에서 또 하나 고려할 점은, 의료영상에 대한 전문가의 판독이나 병변 표시를 위한 편리한 도구의 유무가 데이터 품질 및 규모에 매우 큰 영향을 미친다는 점이다. 구글이 당뇨성망막병증¹³ 연구를 위해 13만 장 가량의 안저 영상을 54명의 전문의에게 판독을 의뢰할 수 있었던 것은, 편리한 인터페이스로 판독에 집중할 수 있는 도구를 제공한 덕분임에 주목해야 한다.

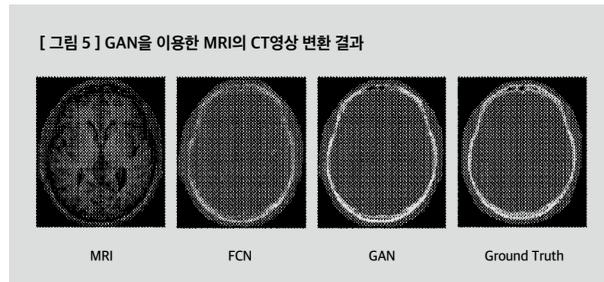
의료영상에서의 GAN

앞서 설명한 바와 같이 의료영상에서 딥러닝을 적용하려는 초기의 시도는 의료영상의 데이터적 특성과 한계를 극복하려는 노력에 집중되었다. 하지만 최근 들어 딥러닝 기반 영상 분석의 적용 범위가 확대되면서, 의료영상에서도 최근 딥러닝 기술들을 적극적으로 도입하거나 혹은 의료분야의 특성을 고려한 다양한 연구들이 이루어지고 있다.

최근에 가장 주목받는 딥러닝 방법 중 하나인 생성적 적대 신경망(generative adversarial network, GAN)도 활발하게 의료에 적용되고 있는데, 인공 의료영상의 생성의 측면에서 바라보기보다는 기존 의료영상에서 연구되던 병변 분할이나 영상 변환과 같은 곳에 적용되어 좋은 결과를 보여주고 있다. 먼저, 독일 암센터 연구팀은 자기공명영상(magnetic resonance imaging, MRI)에서 진행성 전립선암(aggressive prostate cancer) 병변 검출을 위한 GAN기반 영상 분할 방법을 제안하였는데¹⁴, [그림 4]와 같이 전문가의 병변 표시와 분할 모델이 생성한 병변 표시를 구분하는 모델을 학습하고 그 결과를 다시 분할 모델 학습에 반영하는 것을 반복하도록 하였다. 이러한 과정을 통해 분할 모델의 병변 표시 결과가 전문가의 병변 표시 수준에 가까워지도록 학습되고 결과적으로 기존의 영상 분할 방법에 비해 유의한 성능의 개선을 얻게 되었다.



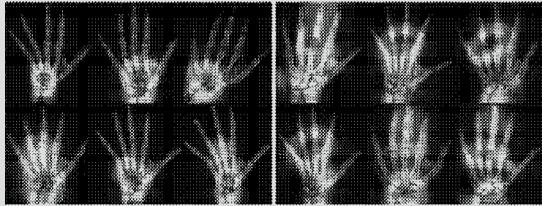
또한 GAN은 서로 다른 영역의 영상 간의 번역(translation) 혹은 변환에 활용될 수 있는데, 의료영상에는 촬영 장치나 목적에 따라 다양한 모달리티(modality)의 영상이 존재하므로 이들 간의 변환은 시간과 비용을 단축하거나 판독의 정확도를 향상시키는데 활용이 가능하다. 미국 노스캐롤리나 연구팀은 [그림 5]와 같이 GAN을 이용하여 MRI를 컴퓨터 단층촬영(computerized tomography, CT) 영상으로 합성하는 연구 결과를 발표하였는데¹⁵, 이때 단순한 GAN 구조로 합성하도록 하면 일반적인 CT 영상과 다르게 흐릿한(blurry) 영상이 생성되므로 실제 CT 영상의 특정 화소와 주변 화소 간의 기울기(gradient)를 모델이 합성한 영상에서도 유지하도록 하는 손실 함수를 추가하여 학습하였다. 그 결과 [그림 5]와 같이 기존의 방법에 비해 보다 선명하고 실제 CT영상과 유사한 합성 영상을 얻을 수 있었는데, 아직까지는 실험적인 수준이지만 이러한 모달리티(modality)간 영상 변환이나 저선량 CT를 표준선량 CT로 변환하는 연구 등이 계속되면 방사선 영상 촬영으로 인한 피폭을 줄이면서도 정교한 방사선치료 계획을 수립하거나 정밀한 판독을 위한 필요한 정보를 얻는데 활용할 수 있다. 이러한 연구들이 지니는 한계는 입력과 목표 출력 영상이 쌍(pair)으로 주어져야 한다는 점으로서 의료에서는 특히 이러한 쌍으로 주어진 영상데이터 확보가 쉽지 않은데, 최근 들어 쌍으로 주어지지 않은 두 영역의 영상 간 변환에 대한 연구들이 등장하면서¹⁶ 이러한 제약도 해소되고 있으므로 앞으로 관련된 다양한 연구 결과들이 발표될 것으로 기대된다.



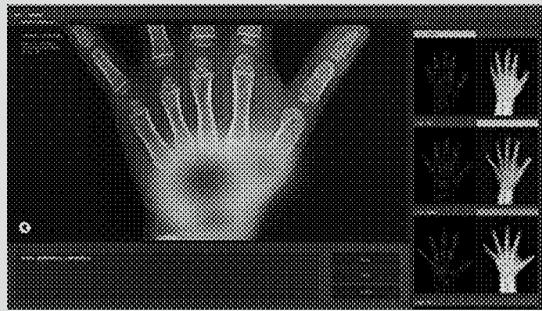
의료영상 인공지능 모델의 해석가능성 및 설명력

의료영상 분석에 딥러닝을 적용할 때 또 하나 고려해야 할 점은 바로 모델의 판단에 대한 설명이 가능한지 여부, 즉 해석 가능성이다. 데이터 기반의 접근법인 딥러닝 모델의 특성상 예측 정확도가 높다고 하더라도 질환의 종류나 중증도에 따라 데이터 수집 경로가 달라 발생하는 데이터 편중(bias)이나 질환과 관련 없는 노이즈를 학습하여 판단을 내릴 가능성이 있다. 따라서 판단의 근거를 시각화하여 모델이 유의미한 특징을 학습하는지 혹은 새로운 영상적 특징을 발견하였는지를 파악해 보는 것은 학습 결과에 대한 신뢰도를 높이고 새로운 영상적 표시자(biomarker)를 발견하는데 중요한 역할을 한다. 이러한 딥러닝 모델의 판단 근거의 시각화 방법도 꾸준히 발전해 왔는데, 가장 단순한 방법은 입력 영상의 관심 영역을 폐색(occlusion)하여 예측 결과의 변화 정도를 보는 방법으로서 해당 영역의 중요도가 클수록 출력 값의 변화가 크다는 점을 가정으로 한다. 이 방법은 허버드 협력병원인 메사추세츠 종합병원(Massachusetts general hospital, MGH)이 개발한 수부 골연령(hand bone age) 판독 시스템¹⁷에 적용되었는데, 아래 [그림 6-a]와 같이 영상이 집중하는 곳일수록 붉은 영역으로 표시되며 실제 연령대별 주요 뼈 구조의 변화를 반영하는 것을 확인할 수 있다. 하지만 폐색 기반의 시각화는 전체 입력 영상 영역을 슬라이딩 윈도우(sliding window)방식으로 가려가며 반복적으로 모델의 출력 값을 확인하여야 하므로 계산 비용이 크고, 폐색 영역의 크기와 모델의 구조에 따라 시각화된 주요 영역이 세밀하지 못한 단점이 있다. 또 다른 시각화 방법으로는 앞서 NIH의 흉부 X-선에 대한 약지도 학습과 같이 공간적 정보를 가지는 특징 지도의 가중합 하는 방법으로 [그림 6-b]와 같이 뷰노의 골연령 판독 시스템¹⁸이 이러한 방법을 적용하고 있다. 이 방법의 장점은 전방향 계산(forward pass)한번으로 시각화에 필요한 정보를 모두 얻을 수 있어 가장 빠르다는 점이며, 단점은 역시 사용되는 특징 지도의 크기에 따라 시각화의 세밀도가 낮을 수 있다는 점이다.

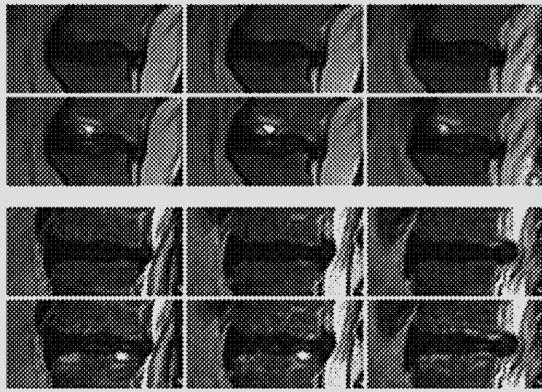
[그림 6] CT영상 및 판독 시스템 시각화



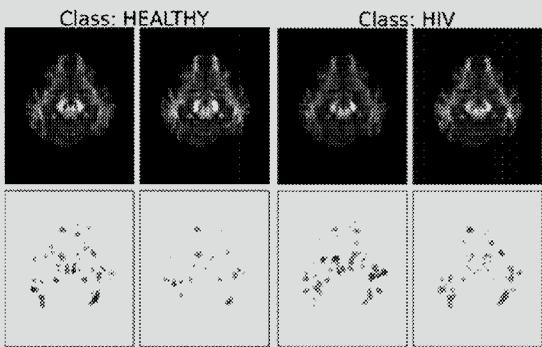
a. MGH의 골연령 판독 시스템의 시각화



b. 뷰노의 골연령 판독 시스템의 시각화



c. 척추 MRI병변 탐지 시각화



d. HIV 환자 뇌 MRI 병변 시각화

통해 입력 영상의 화소별 편미분(partial derivative) 값을 시각화하는 방법이다. 옥스퍼드 대학에서는 [그림 6-c]와 같이 척추 MRI에서 여러 질환을 동시에 찾아내거나 점수화하는 다중 작업 모델¹⁹⁾의 판단 근거 시각화에 이 방법을 적용하였다. 이 방법은 화소 단위의 매우 세부적인 시각화가 가능하고 역전파를 한번 계산하여 시각화에 필요한 정보를 얻을 수 있어 비교적 빠르지만, 시각화의 결과가 입력영상의 넓은 영역에 분산되는 경향이 있어 직관성이 떨어지고 해석이 용이하지 않은 단점이 있다. 최근 들어서는 예측 차 분석(prediction difference analysis)을 통한 시각화 방법²⁰⁾이 제안되었는데, 이는 관심 영역의 모델 출력 값에 대한 중요도를 평가하기 위해서 원본 영상에 대한 출력 값과 관심 영역을 주변화(marginalize)한 영상의 출력 값의 차이를 관련 점수(relevance score)로 하여 시각화하는 방법이다. 기본적으로 관심 영역을 대치하면서 출력의 변화를 보는 페색 방법과 유사하지만, 해당 영역을 모두 특정 값으로 대치해버리는 것이 아니라 관심 영역의 주변부로부터 샘플링하여 대치한다는 차이점이 있다.

결과적으로 [그림 6-d]와 같이 뇌 MRI로부터 HIV 환자와 정상환자를 분류하는 모델의 판단 근거를 시각화한 결과, 직관적이고 해석이 용이한 결과를 얻을 수 있다.

의료영상 인공지능의 남은 주제와 미래

이상으로 인공지능 기반의 의료영상 분석의 사례들과 고려해야 하는 주제들에 대해서 살펴보았다. 이러한 주제들 외에도, 인공지능 모델의 판단 결과에 대한 불확실성을 추정함으로써 판단의 불확실성이 높은 경우 의사를 개입시키는 방법도 제시되고 있으며²¹⁾, 임상적 혹은 의학적 지식을 함께 반영한 모델 학습 방법, 질환적으로 유사한 영상을 검색하는 모델 개발, 조영 증강(contrast enhanced) 영상이나 추적 관찰(follow-up)영상과 같은 3차원 이상의 고차원 영상 데이터에 대한 분석, 마지막으로 인공지능 모델 학습 및 활용 시의 의료영상의 보안과 관련된 연구²²⁾²³⁾ 등 이제 연구가 시작되었거나 본격적인 연구가 필요한 주제들이 많이 남아 있다.

하지만 세계최초로 미국 식약청(food and drug administration, FDA) 승인을 받은 딥러닝 기반의 심장 MRI 정량화 소프트웨어 Cardio DL²⁴⁾의 사례처럼 현재까지 개발되어 널리 쓰이는 기술도 잘 정의된 문제와 고품질의 데이터와 결합된다면 인공지능 기반 의료영상 분석 기술이 실제 임상에 폭넓게 활용되는 시기의 도래가 가속화 될 수 있다. 국내에서도 식품의약품안전처가 빅데이터 및 인공지능 기술이 적용된 의료기기에 대한 허가심사 가이드라인을 마련하고 여러 인공지능 기업들이 인허가 절차에 돌입한 상황으로 멀지 않은 시점에 국내에서도 인공지능 기반의

의료 영상 분석 소프트웨어들이 활용되기 시작할 것이다. 인공지능 기술은 향후 영상 정보뿐 아니라 다양한 환자 건강 정보나 유전체 정보를 통합하여 분석함으로써 질병의 조기 발견, 예후 및 생존 예측 등을 통해 의료 지출을 감소시키고 환자들의 삶의 질을 향상할 수 있는 잠재력을 가지고 있다. 따라서 병원과 기업, 그리고 임상 연구자와 인공지능 연구자들의 보다 적극적인 만남과 협업을 통해 데이터 기반의 정밀 의료의 구현과 확산의 가능성이 보다 높아질 것으로 기대된다.

*1 논문 : Litjens, G. et. al, "A Survey of Deep Learning in Medical Image Analysis", arXiv, 2017. *2 논문 : Lo, S-C. et. al, "Computer-assisted diagnosis of lung nodule detection using artificial convolution neural network", SPIE Medical Imaging, 1993. *3 논문 : Chan, H.P. et. al, "Computer-aided detection of mammographic microcalcifications : pattern recognition with an artificial neural network", Medical Physics, 1995. *4 자료 : <http://image-net.org/challenges/LSVRC/2012/results.html> *5 논문 : Ciresan, D.C., et. al, "Mitosis detection in breast cancer histology images using deep neural networks", MICCAI, 2013. *6 논문 : Gulshan, V. et. al, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs", JAMA, 2016. *7 참고 : Esteva, A. et. al, "Dermatologist-level classification of skin cancer with deep neural networks", Nature, 2017. *8 논문 : Cheng, J-Z., et. al, "Computer-aided diagnosis with deep learning architecture: application to breast lesions in US images and pulmonary nodules in CT scans", Scientific Reports, 2016. *9 논문 : Shin, H-C., et. al, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning", IEEE trans. on. Medical Imaging, 2016. *10 논문 : Tajbakhsh, N., et. al, "Convolutional neural networks for medical image analysis: fine tuning or full training?", IEEE trans. on. Medical Imaging, 2016. *11 자료 : <https://research.googleblog.com/2016/03/train-your-own-image-classifier-with.html> *12 논문 : Wang, X., et. al, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases", CVPR, 2017. *13 자료 : <https://www.youtube.com/watch?v=oOeZ7lgEN4o> *14 논문 : Kohl, S., et. al, "Adversarial networks for the detection of aggressive prostate cancer", arXiv, 2017. *15 논문 : Nie, D., et. al, "Medical image synthesis with context-aware generative adversarial networks", arXiv, 2016. *16 논문 : Zhu, J-Y., et. al, "Unpaired image-to-image translation using cycle-consistent adversarial networks", arXiv, 2017. *17 논문 : Lee, H., et. al, "Fully automated deep learning system for bone age assessment", Journal of Digital Imaging, 2017. *18 논문 : Kim, J. R., et. al, "Computer-assisted program using deep learning technique in determination of bone age: evaluation of the accuracy and efficiency", American Journal of Roentgenology, 2017. *19 논문 : amaludin, A., et. al, "SpineNet: automatically pinpointing classification evidence in spinal MRIs", MICCAI, 2016. *20 논문 : Zintgraf, L. M., et. al, "Visualizing deep neural network decisions: prediction difference analysis", ICLR, 2017. *21 논문 : Leibig, C., et. al, "Leveraging uncertainty information from deep neural networks for disease detection", bioRxiv, 2016. *22 논문 : Shokri, R., et. al, "Privacy-preserving deep learning", ACM SIGSAC CCS, 2015. *23 논문 : Papernot, N., et. al, "Semi-supervised knowledge transfer for deep learning from private training data", ICLR, 2017. *24 자료 : <https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#5520cc62161c>

또 다른 시각화 방법으로는 입력 영상의 각화소의 변화에 따른 출력 값의 민감도를 측정하는 방법으로, 역전파(back propagation)를

AI 의료, 이제 윤리를 고민하다

유발 하라리 작가는 최신작 <호모 데우스>를 통해 인공지능의 등장으로 인류가 멸망할 것으로 예언을 했다. 하라리는 역사학자로서 <사피엔스>란 책을 통해 세계적인 유명세를 얻은 작가이다. 그는 인간의 역사를 인지혁명, 농업혁명, 과학혁명이라는 틀로 분석한다.

역사적으로 우리의 선조격인 호모 사피엔스는 약 20만 년 전에 등장했다. 그러다가 7만 년 전쯤 도구를 사용하기 시작했다. 이 시기를 인지혁명으로 규정했다. 인류가 똑똑해지기 시작했다는 의미다. 이후 1만 2,000년 전에 수렵생활을 중단하고 농경지에 정착해 생활하기 시작했다. 이를 농업혁명이라고 불렀다. 개인으로 보면 수렵생활보다 편하지 않지도 않고 감염병에도 취약했지만 인류 전체로 봤을 때는 매우 효율적인 시스템으로 해석했다. 과학혁명은 약 5백 년 전에 일어난 자본주의와 에너지 생산과 소비의 확대 등 산업혁명을 전반적으로 일컫는 말로 정의했다.

유발 하라리는 신작 <호모 데우스>에서 '생명공학에 따르면 인간의 행동은 인간의 자유의지가 아닌 일종의 호르몬의 영향에 따라 움직인다'며 인간행동 자체를 일종의 알고리즘으로 봤다. 그렇기에 인간이 특정한 선택을 하기 전에 알고리즘으로 만들어진 인공지능이 그 선택을 대신 하도록 할 수 있고, 이런 기술적 발달은 대규모 실직 사태와 최종적으로 인류의 종말로 인도할 가능성도 있다는 것이다.

인공지능에 도덕과 윤리가 필요한가

체스에 이어 인공지능이 바둑에 있어서도 인류를 앞섰다. 구글의 알파고가 이세돌 9단을 꺾으면서 인공지능에 대한 관심이 하늘을 찌르고 있다. 정부도 나서서 인공지능 등을 활용한 4차 산업을 육성하겠다고 말하고 있다. 그러나 이런 관심의 이면에는 불안감도 도사리고 있다. 가장 큰 것은 인공지능이 우리의 일자리를 대체할 가능성이다. 안타깝게도 이 부분은 현실이 될 가능성이 많다. 수많은 연구기관에서 현재의 일자리 중 단순한 업무부터 대체 될 것이라고 예측하고 있으니 말이다. 그 다음의 불안감은 영화 터미네이터나 매트릭스에 나오는 인공지능처럼 인간을 괴롭히는 나쁜 역할을 하지는 않을까 하는 걱정이다. 이런 사태를 막기 위해서라도 인공지능의 윤리가 필요하다.

올해 열린 대한의료정보학회 춘계학술대회는 인공지능의 윤리에 대한 심도 깊은 토의가 열렸다. 토의에 참여한 한국정보화진흥원의 이영주 수석은 여섯 가지 이슈를 대비해야 한다고 했다. 첫째, 과연 안전하게 작동할 것인가, 둘째, 프라이버시 문제, 셋째, 사용자의 오남용 문제, 넷째, 누가 어떻게 책임질 것인가, 다섯째, 인간의 고유성은 어디까지인가, 여섯째, 인공지능에 대한 막연한 공포 등이었다. 그 중에서 가장 크고 당면한 문제는 안전문제다. 이미 상식이 됐듯이 자율주행 자동차는 기술적으로 만들어졌고, 법과 제도만 보완되면 언제든 실제 도로에서 볼 수 있다. 그러나 자율주행 자동차에도 큰 결함이 있는데 바로 트롤리 딜레마(trolley problem)를 해결 할 수 없다는 것이다.

윤리와 철학을 연구하는 사람에게는 익숙한 주제일 것이다. 트롤리 딜레마란 윤리학의 사고 실험으로 기차가 두 선로에 있는 사람을 칠 수밖에 없다면 많은 사람을 칠 것인가, 소수의 사람을 칠 것인가 하는 철학적 문제를 다룬다. 기차 대신에 자율주행 자동차로 치환해 둘 경우 인공지능은 어떤 판단을 할 것인가. 좀 더 심한 가정도 가능하다. 앞에 갑자기 나타난 유치원생 무리를 칠 것인가, 아니면 탑승자만 다치도록 외벽에 부딪히는 것을 택할 것인가. 혹은 이런 윤리적 이슈가 '철학자들만 관심 있는 사항'으로 치부하기도 한다. 그러나 가능성이 적을 뿐 언제든지 있을 수 있는 일이다. 2015년 기준으로 우리나라에서 발생한 교통사고는 23만 건이 넘는다. 이 중에 보행자 사고는 40%에 가깝다.

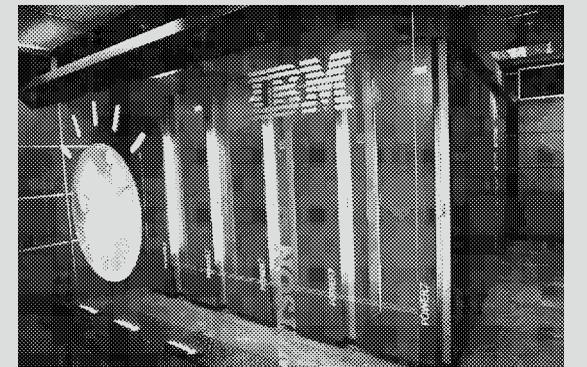
IBM 왓슨은 과연 윤리적일까?

작년 12월 우리나라 최초로 인공지능 왓슨이 길병원에 도입됐다. 언론 보도에 따르면 진료를 본 대다수의 환자가 의료진의 처방과 왓슨의 처방이 다를 때 왓슨을 따르겠다고 밝히고 있다. 이렇게 왓슨을 절대적으로 신뢰해도 될까?

토론에 참석했던 서울대병원 혈액종양내과 김범석 교수는 국내에 들어온 IBM 왓슨도 윤리적인 측면에 문제가 있다고 봤다. 환자의 사회경제적인 요건을 고려하지 않고 교과서적인 '최상'의 치료를 권하기 때문이다. 게다가 이런 왓슨이 최상의 처방은 현재 우리나라 의료시스템에서는 임의비급여 항목으로 '불법'에 해당되는 경우도 심심치 않게 있다고 한다.

알다시피 우리나라 의료시스템은 사회보험의 성격이 짙은 건강보험공단이란 단일보험 시스템이다. 최단 기간에 전국민 의료보험을 시행한 만큼 허점도 많다. 대표적으로 보험료가 싸고, 보험의 보장성도 낮고, 병원에 돌려주는 수가도 낮다. 그러다 보니 의학적으로 효과가 있다는 결론이 나온다고 바로 보험 적용을 해주지는 않는다. 게다가 이를 보험적용이 되지 않는 '비급여'로 처리하는 것도 애초에 불가능하다. 적발되면 '임의비급여'로 병원이 받은 돈 이상을 뱉어내야 하기 때문이다. 이런 속사정을 알 리가 없는 왓슨은 최선의 논문을 검색해 결과를 알려주고 있다.

[그림 1] IBM의 슈퍼컴퓨터 '왓슨'



왓슨을 이용했을 때 책임의 한계도 여전히 모호하다. 만약 왓슨이 잘못 진단했을 경우 누가 책임질 것인가. 도입한 병원의 책임인가, 아니면 이용하겠다고 결심한 개인의 책임인가, 그것도 아니면 제조사인 IBM의 책임인가. 법인처럼 사람이 아니더라도 책임을 물을 수 있도록 하는 방안도 검토되고는 있다지만 아직 실체화되지는 않고 있다.

더 큰 문제는 이런 인공지능 활용이 보편화 됐을 때 생기는 의료진의 '의존증'이다. 지금 추세로 발전해 나간다면 머잖아 모든 병원에서 왓슨과 같은 인공지능을 병원 전산처방시스템과 함께 사용하게 될 것이다. 그렇게 된다면 인공지능이 권하는 처방순위에 따라 많은 의료진들이 의식하지 않은 채 처방을 내리게 될 것이다. 이미 이와 비슷한 경험을 전산처방시스템을 도입하면서 경험한바 있다. 인공지능이 특정 치료나 치료제를 우선적으로 추천한다면 관련된 회사의 주식은 폭등하게 될 가능성도 배제할 수 없다.

한편 아직은 인공지능의 윤리를 논하기에는 이르다는 주장도

글 | 양광모 kmhi.yang@samsung.com

비뇨기과 전문의로 공중보건의 생활을 하다가 블로그에 매료되어 기자가 되기로 결심했다. 블로그 미디어인 코리아헬스로그 편집장과 종이 매체인 청년의사신문의 편집국장을 거쳤다. 그러다 언론과 병원이 바라보는 의료시스템에 대한 생각의 간극이 크다고 느껴 병원으로 돌아갔다. 현재는 삼성서울병원 건강의학센터 교수로 있으며 같은 기관 커뮤니케이션실 차장과 대한의료정보학회 홍보이사를 겸하고 있다.

있다. 예방의학 전문의이자 변호사인 단국대학교 박형욱 교수는 ‘책임 문제는 재물손괴죄란 형법에 따라 소유주가 책임을 지도록 하면 된다’면서 ‘아직 (인공지능의 책임을 묻는다는 것은) 요원한 일이다. 이용하는 의사와 소속 병원의 책임을 묻도록 하면 된다’고 주장했다. 그는 만약 왓슨이 더 발전하게 되어 거의 필수적인 도구로 인정받게 될 경우에는 이를 활용하는 방법에 대해 설명하도록 의무화 하는 것도 방법이라고 했다. 지금도 작은 뇌종양일 경우 감마나이프로 수술하거나 전통적인 개두술로 수술하는 방법 모두 유효한데, 이런 경우 환자에게 설명하지 않으면 의사가 법적으로 책임을 지게 되어 있다.

하지만 여전히 더 복잡해질 알고리즘에 의해 추천되는 결과물을 검증하는 문제는 여전히 요원했다. 서울아산병원 빅데이터센터의 감혜진 교수는 ‘인공지능은 매우 복잡한 함수와 같다’며 ‘겉에서 보는 것과 달리 뒤편에서는 복잡한 알고리즘과 새로 추가된 무수한 정보들이 있기 때문이다.’라고 했다. 미래에는 의사들의 새로운 직업으로 ‘알고리즘 감청사’가 생길 것이란 전망도 가능해 보인다.

인공지능에 윤리를 요구해선 안 된다

인공지능에게 윤리를 학습하도록 해서는 안 된다는 주장도 있다. 한국정보화진흥원의 이은주 수석은 ‘인간에게도 윤리적 딜레마는 풀기 어려운 문제’라며 ‘인간이 풀지 못하는 딜레마를 인공지능에게 해결해달라고 해선 안 된다’고 했다. 인간이 윤리적으로 완벽하지 않은데 어떻게 인공지능에게 윤리를 요구할 수 있겠느냐는 뜻이다. 앞서 예를 들었던 단일보험의 문제는 건강보험심사평가원의 기준을 인공지능이 학습하지 못했기 때문에 생긴 일이기 때문에, 소위 ‘심평의학(심사 기준을 뜻함)’에 대해 학습시키면 해결될 수 있다는 것이다. 하지만 정말 답을 내리기 어려운, 의학적, 윤리적 판단 상황에 있어서는 인공지능의 결론을 기대할 수 없다고 했다. 예를 들어 태아를 살릴 것인가 산모를 살릴 것인가, 부족한 인공호흡기를 50세 환자와 60세 환자 중에 누구에게 적용시킬 것인가, 에이즈 환자를 수술해야 하는데 방호복이 없을 때에는 어떻게 할 것인가 등이다.

인간에 있어서는 개인 의사의 가치판단에 의해 어떻게든 결론이 난다. 중증의 50세 남성보다 경증의 60세 남성이 회복 가능성이 더 높다고 판단해 60세 남성에게 인공호흡기를 적용시키고, 50세 남성은 의료진이 엠부백(ambu bag)을 시행하도록 할 수도 있다. 반대로 50세 남성이 중요한 사람이라서 인공호흡기를 적용시키고 60세 남성을 수동으로 산소 공급할 수도 있다. 태아와 산모 사이의 딜레마와 에이즈 환자 수술의

문제도 마찬가지다. 개인이 받은 교육과 가치체계에 따라 판단이 내려지겠지만 어떤 판단을 하더라도 나름의 이유가 있기 마련이다. 그러나 이런 판단을 기계에게 맡길 수는 없다는 주장이다.

애초부터 인공지능 왓슨이 병원에서 이용되는 것 자체가 윤리적인지 확인을 해야 한다는 주장도 있었다. 단국대학교 박형욱 교수는 ‘의료적으로 활용하려면 식약처에서 유효성과 안정성을 인정해야 한다. 그러나 지금 도입된 왓슨은 유효성과 안정성에 대해 공식적으로 인정받지는 않은 것으로 알고 있다’고 했다. 실제로 왓슨을 도입한 길병원 등에서는 환자가 원할 경우 자문을 받을 수 있도록 하지만, 비용은 청구하지 않고 있다. 아니 못하고 있다고 표현해야 더 정확하다.

인간이 먼저 자신을 제대로 아는 것이 선행돼야

감혜진 교수는 트롤리의 딜레마를 예를 들며 인간의 윤리를 언급했다. 그는 ‘대다수의 사람들은 최대다수의 최대행복이란 공리주의에 따르는 원칙을 좋아한다. 트롤리의 딜레마와 같은 상황에서 많은 사람을 살릴 수 있도록 설정하는 것을 공개적으로 지지한다. 자율주행 자동차도 공리주의적 관점에서 다수를 살리는 선택을 하도록 하라는 것이다. 이 원칙에 따르면 다수 사람들이 갑자기 도로에 튀어 나왔을 때 자율주행 자동차가 자기 스스로 벽에 부딪히는 선택을 하도록 해야 한다. 그러나 조사를 해보니 이렇듯 공리주의 관점을 지지했던 사람들의 대다수가, 실제 그렇게 세팅한 자율주행 자동차를 구매하겠느냐고 물으니 구매하지 않겠다고 했다. 인간의 모순이 단적으로 드러난 것이다’고 했다. 인간이 스스로 윤리적이지만 않은데 어떻게 인공지능의 윤리를 논할 수 있느냐는 지적이다.

박형욱 교수는 ‘의료에 있어서 윤리적 측면이 모호한 부분이 있다. 장기이식을 하려면 질병관리본부 장기이식관리센터에 등록을 해야 하는데, 나는 젊고 건강한 이식 대기자인데 내 앞에는 노인 이식 대기자가 있다. 노인인 경우 이식한 장기가 제대로 생착되지 않아 결과적으로 장기를 낭비하는 경우도 생길 수 있다. 하지만 현재는 먼저 신청한 사람부터 제공하고 있다’며 철학적 관점을 법률로 강제할 수도 있음을 언급했다.

어떻게 보면 인공지능이 보편화되지 않아 사회적 합의점이 도출되지 않은 상태이기 때문에 벌어지는 논란일 수도 있다. 언론 보도에서나 접한 자율주행 자동차나, 특정 병원에 있는 의료용 인공지능을 생각해보면 사실 그렇기도 하다. 모든 자동차가 자율주행으로 움직이고, 모든 병원에서 인공지능을 활용한다면 논의가 조금 더 단순해 질 가능성도 있다. 유발 하라리는 <호모 데우스>를 통해 인간이 스스로 판단하는 것을 포기하는 시점에

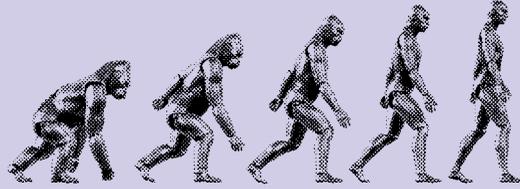
대해 이야기하고 있다. 나를 나보다도 더 잘 아는 인공지능이, 나를 대신해 대선 후보자를 선택해주고, 나를 대신해 배우자를 선택하는 일도 가능하다는 것이다. 이미 일부 논문에서는 페이스북의 ‘좋아요’를 분석해 사람의 성향을 상당히 높은 확률로 예측했다고 보고하고 있다.

만약 인공지능에 의존해 주요한 판단을 내리게 된다면 어떻게 될까. 분명 지금과는 다를 것이다. 유발 하라리는 지금의 자유주의 시스템은 붕괴할 것이라고 비판적으로 예측했다. 대중의 시대는 끝나고 일부 엘리트들이 나머지를 지배하는 구조로 갈 것으로 전망했다. 그렇다고 ‘유발 하라리’의 말을 너무 마음에 담아둘 필요는 없다. 꽤 신뢰할 만한 역사학자긴 하지만 미래를 예측하는 전문가라고 보기는 부족한 부분도 있기 때문이다.

하지만 인공지능 시대를 맞이해 시사점은 분명히 있다.

인간이 적절하게 개입하지 않으면 좋지 않은 결과로 치달을 수 있다. 이를 사전에 막기 위해서는 인공지능의 윤리에 대해 더 깊은 고민이 필요하다.

AI의 진화



review

정부환 | 더욱 똑똑해진 AI 광고 알고리즘

40

이원형 | AI는 인간과 같은 감정을 가지게 될까?

50

전 세계에서 시장 가치(market capitalization)가 가장 높은 회사 중 하나인 알파벳(Google의 모회사)은 매출 90%를 광고로부터 벌어들이고 있습니다. 광고는 인터넷 서비스에서는 없어서는 안 될 존재입니다. 서비스를 이용자들의 경험을 방해하지 않고 어떻게 하면 효과적인 광고를 할 수 있을까? AI에서 그 해답을 찾을 수 있습니다. 첫 번째 글은 AI 기술이 어떻게 카카오 광고 알고리즘에 활용되고 있는지 설명하고 있습니다. 상상 속의 로봇은 영화 속에 등장하는 '아이로봇'처럼 인간을 이해하고, 인간과 교감할 수 있기를 기대하고 있습니다. 현재 AI 기술로 과연 인간의 감정을 이해할 수 있을까요? 두 번째 글을 통해 확인할 수 있습니다.

더욱 똑똑해진 AI 광고 알고리즘

카카오의 광고랭킹 알고리즘

‘알파벳(Alphabet Inc., Google의 모기업)’은 어떤 회사인가?라는 질문에 많은 사람들은 검색 서비스 회사나 안드로이드 OS를 만드는 회사 정도로 답한다. 이 글을 읽는 분들이라면 텐서플로우(TensorFlow)를 만든 회사 또는 알파고(AlphaGo)를 만든 딥마인드(DeepMind)의 모회사 정도로 답할지도 모른다. 같은 질문을 페이스북(Facebook Inc.)에 대해 한다면, 소셜미디어(SNS), 인스타그램(Instagram) 혹은 왓츠앱(WhatsApp) 서비스를 제공하는 회사라는 답변이 가장 많을 것이다. 한국을 대표하는 기업들 중 하나인 삼성전자는 스마트폰을 팔아서 돈을 벌고, 현대자동차는 자동차를 팔아서 돈을 버는 반면, 구글과 페이스북은 제품이나 서비스를 파는 것이 아닌 광고를 통해 대부분의 매출을 올리고 있다. 현대자동차가 자동차 회사라면, 구글과 페이스북은 광고 회사인 셈이다. 구글은 광고를 위해 검색 서비스와 안드로이드 OS를 만들고, 페이스북도 광고를 위해 타임라인과 인스타그램 등의 서비스를 제공한다. 광고의 정의와 범위에 따라 달라지겠지만 우리가 알고 있는 대부분의 인터넷 기업들을 광고 회사로 봐야 한다. 광고 비즈니스를 알면 인터넷 기업의 진면모를 제대로 볼 수 있지만, 아쉽게도 많은 이들이 - 심지어 IT 회사의 종사자마저도 - 인터넷 기업들이 어떻게 돈을 벌고 있는지 자세히 알지 못한다.

이 글은 여러분이 기대하는 AI, 특히 딥러닝(deep learning)에 관해서는 다루지 않는다. 이게 무슨 반트렌드적인 발상인가라고 의문을 표할 수도 있겠으나, 회사의 근간을 이루는 기술일수록 더 보수적으로 적용되기 마련이다. 여전히 은행의 레거시 시스템(legacy system)은 코볼(COBOL)로 짜여졌다는 설이 있듯이, 기업의 중요한 자산일수록 더 검증된 기술을 사용하는 경향이 있다. 광고 시스템에 사용되는 알고리즘은 정확도도 중요하지만 짧은 시간 안에 안정적으로 서비스하는 것이 더 중요하다. 그렇기에 부분 기술로는 딥러닝을 적용한 케이스나 논문들이 최근 소개되고 있지만, 광고 시스템 전체를 딥러닝으로 대체하는 데는 여전히 시간이 필요하다. 2006년 10월에 넷플릭스(Netflix)는 추천 정확도(root mean square error, RMSE)를 10% 높이는 알고리즘을 개발한 팀에게 100만 달러를 상금으로 주겠다는 넷플릭스 프라이즈(Netflix Prize)를 진행했고, 실제 목표를 달성한 알고리즘이 2009년에 개발됐다. 하지만, 그 알고리즘을 바로 제품 단계(production level)에서 넷플릭스 추천 엔진에 탑재하지 못했다². 이렇듯 딥러닝 기술도 광고 시스템에 적용하기 위해서는 확장(scale up) 과정이 필요하다.

광고를 단순히 기업의 매출 수단으로만 생각한다면 광고의 본질을 볼 수 없다. 흔히 서비스와 광고를 배타적인 관계로 보는 경향이 있다. 서비스에 광고가 들어오는 순간 서비스의 정체성을 상실한다고 생각한다. 그렇기에 광고가 서비스의 사용자 경험을 해치지 않게 자연스럽게 녹아들고 때론 더 유용한 정보를 제공하는 매개가 되도록 노력한다. 재미있고 창의적인 광고 소재를 만드는 것에 더해서, 그 광고가 노출될 서비스 또는 매체의 본질을 파악하고 그걸 소비하는 사용자들을 제대로 이해하는 것이 필요하다. 이 글에서 설명할 광고랭킹은 사용자, 매체/서비스, 광고 (소재와 광고주), 그리고 맥락을 연결하는 것이다.

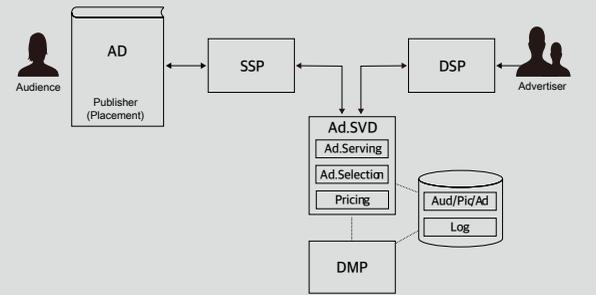
광고 시스템 개요

카카오의 광고 플랫폼에서 사용하는 기계학습 알고리즘을 이해하기 위해서 먼저 광고 시스템이 어떻게 구성되고 어떤 순서로 광고가 집행되는지를 간략히 알아보도록 하자.

【 광고 집행 흐름 】

1. 광고주는 DSP(demand side platform)를 통해 광고 소재 창작, 타게팅 조건, 입찰 방법 등을 등록한다.
2. 오디언스(audience)³가 웹 페이지에 접속하면 광고 매체를 관리하는 SSP(supply side platform)를 통해서 광고를 요청한다.
3. 광고 요청을 받은 광고 서버는 지면 정보를 받아와서 카카오 내부에서 분석, 저장한 오디언스/지면/광고 정보를 기반으로 현 상황에 가장 적합한 광고를 선택해서 SSP에 전달한다. 추가로 프리퀀시 캐핑(frequency capping), 어뷰징 체크(fraud detection), 예산 관리 등을 통해서 광고 노출을 제어하고, 약속된 액션(노출, 클릭, 전환 등)에 맞게 과금한다.
4. 이상의 모든 과정은 DMP(data management platform)를 통해서 기록(log)으로 남긴다.

【 그림 1 】 광고 집행 플로우와 시스템 구성



광고 시스템은 일반 서비스에 활용되는 CMS(content management system)와 크게 다르지 않으나, 역할에 따라서 DSP, SSP, 애드서빙, DMP 등으로 구분한다.

【 광고 시스템 】

1. DSP(demand side platform)

DSP는 광고주와의 접점이다. 즉, 카카오가 관리하는 모든/특정 지면에 광고를 노출하려는 광고주가 접속해서 1) 광고의 목적을 설정하고 (캠페인 등록), 2) 메시지가 전달될 타겟 오디언스, 노출될 지면, 집행 기간, 예산 및 입찰 조건을 설정하고 (광고 그룹 등록), 3) 오디언스에게 전하고 싶은 메시지와 형식을 설정하고 (광고 소재 등록), 그리고 4) 최종적으로 집행된 광고의 효과를 확인하는 보고서 기능을 제공한다. 특히, 니즈가 명확한 검색광고가 아닌 일반 배너광고(display ad)에서 오디언스 타게팅 기능이 중요해지고 있다. 오디언스 타게팅은 오디언스의 기본 데모그래픽 (성별, 나이대, 거주지 등)⁴, 카카오 내 활동을 분석해서 추정된 관심사 등으로 타겟 그룹을 분화한다. 현재 카카오는 마이더스(Midas)와 클릭스(Clix)라는 DSP를 제공하고 있으며, 현재 모멘트(Moment)라는 새로운 통합 DSP도 CBT(closed beta test) 중이다.

2. SSP(supply side platform)

DSP가 광고주와의 접점이라면, SSP는 카카오 광고를 노출하는 매체 또는 발행자(publisher)와의 접점이다. 카카오 광고는 카카오 내부 지면 (카카오톡, 카카오 스토리, 다음앱, 다음PC포털 등)과 외부 네트워크 지면 (카카오와 다음 서비스 이외의 외부 언론사 홈페이지, 독립 개발자의 앱 등)에 노출된다. 이런 모든 지면을 등록해서 광고를 제공할지 여부를 판단, 관리하는 것이 SSP의 역할이다. 그리고 최근에는 프로그래매틱 바이잉(programmatic buying) 또는 실시간 비딩(real-time bidding, RTB) 등이 애드테크(AdTech)의 트렌드인데, 이런 AdX(AdExchange)의 광고 중계(mediation)도 SSP의 역할이라 할 수 있다. 이미 카카오는 크리테오, GDN(Google Display Network) 등의 외부 DSP가 제공하는 광고도 애드핏(AdFit, 카카오 SSP)를 통해서 내부 지면에 노출하고 있다.

글 | 정부환 ben.hur@kakaocorp.com

그냥 논문 읽고 연구하고 논문 쓰기를 좋아하던 학생이 생계를 위해서 검색 데이터를 모델링하고, 쇼핑과 뉴스 등의 콘텐츠를 추천하는 알고리즘을 개발하고, 광고 플랫폼에서 정밀한 타게팅을 고민하다가 지금은 광고랭킹을 개선하면서 10년의 세월을 보냈습니다. 그러는 사이에 빅데이터의 흐름을 경험하고 최근에는 딥러닝을 필두로 한 AI 기술의 폭발에 기술적 고민보다는 철학적 고민에 더 깊이 빠집니다.

3. AdServer

애드서버 또는 애드서빙(AdServnig) 시스템은 DSP와 SSP에 별도로 존재해서 광고 요청에 가장 적합한 광고를 선별하는 기능과 어떤 매체에 광고를 제공할지를 조정하는 기능을 한다. 편의상 하나의 가상 애드서버가 있는 것으로 도식화했다. 크게 1) 광고를 매체(또는 SSP)에 제공하는 애드서빙(AdServing), 2) 현재 맥락에서 접속한 오디언스 및 매체에 가장 적합한 광고를 선택하는 애드 셀렉션(Ad Selection), 그리고 3) 오디언스가 약속한 액션을 취했을 때 얼마의 금액을 광고주에게 요청할 것인가를 정하는 과금액 산정(pricing) 기능을 담당한다.
그외에 광고주의 일예산 관리 (예산 분배 및 광고 중지), 오디언스의 광고 피로도를 고려한 프리퀀시 캐핑, 광고 어뷰징 제어, 과금(billing), 그리고 실시간 광고중계 등의 기능도 담당한다.
이번 글은 애드 셀렉션(Ad Selection)을 중심으로 더 자세히 설명할 예정이다.

4. DMP(data management platform)

DMP는 광고 집행에 필요한 모든 데이터와 집행 중에 발생하는 모든 이벤트 로그를 저장/관리하는 시스템이다. DSP, SSP, 애드서버 등에 DMP 기능이 분산돼있으나 편의상 하나의 가상 시스템으로 그렸다. DSP를 통한 광고주 및 광고 데이터와 SSP를 통한 매체 데이터 뿐만 아니라, 카카오가 적법하게 수집한 데이터로 추정/분석한 오디언스 정보도 관리한다. 그리고 광고를 집행 하면서 발생하는 광고 요청, 노출, 클릭, 전환 및 빌링 로그 등을 저장한다. 특히 최근에는 광고에서 전환(conversion)이 중요해져서 CTS(conversion tracking system) 및 MAT(mobile app tracking) 등 외부 데이터를 관리하는 역할도 한다.

5. DR(data repository)

데이터 저장소는 DSP와 SSP에 등록된 모든 데이터와 DMP를 통해서 수집한 모든 정보와 로그를 저장해서 요청시에 빠르게 제공해주는 곳이다. 특히 애드 셀렉션에서 30~40ms 내에 광고를 선별해 제공하기 위해서 빠르게 데이터를 찾아서 안정적으로 전달해줘야 한다.

추천 시스템과 광고랭킹 시스템

카카오 AI 리포트 6월호에 카카오의 추천 엔진인 토로스(TOROS)가 소개되었다. 어떤 이는 광고랭킹도 광고 소재를 추천하는 것이니 토로스를 광고랭킹에 바로 적용하면 되는 것 아니냐?라고 반문할지 모른다. 광고랭킹이 일반 추천과 다른 점이 몇 가지 있다.

첫째, 추천 시스템은 보통 추천되는 아이템 또는 콘텐츠의 순서만 중요하고 소위 Top-N으로 명명되는 상위의 N 개의 아이템만 잘 추려내면 된다. 반면 광고랭킹은 추천 순서 뿐만 아니라, 정확한 추천 점수가 필요하다. 보통 pCTR(predicted click-through rate) 또는 이를 기반으로 한 eCPM(effective cost per mille)⁵을 정확히 계산해야 한다. eCPM에 따라서 광고의 노출 여부가 결정될 뿐만 아니라, 노출/클릭/전환에 따른 과금액이 산정된다. 과금액 산정이 부정확 하다면 카카오 입장에서서는 제대로 된 매출을 기대하기 어렵고, 광고주 입장에서는 과대계상된 금액을 지불할 우려가 있다.

둘째, 추천 순서와 함께 과금액을 함께 결정해야 한다.

추천된 콘텐츠를 사용자가 조회했다고 해서 과금하지 않지만, 광고는 (CPC 상품인 경우) 노출된 광고를 오디언스가 클릭하면 할인된 CPC(cost-per-click)금액을 광고주에게 청구한다.

광고주가 제시한 입찰가(BA, bidding amount) 대로 과금하는 최고호가경매(first price auction)이면 큰 문제가 없지만, 보통

광고 시장에서는 더치경매(Dutch auction), 또는 Vickery-Clarke-Groves auction(VCG)으로 알려진 세컨프라이스경매(second price auction)에 기반해서 과금액을 산정한다. 이때 eCPM 또는 광고 시스템 별로 지정된 랭킹점수(engagement score)의 비율에 따른 할인율을 적용한다.

셋째, 높은 수준의 시스템 안정성이 요구된다. 재차 강조하지만 광고는 회사의 매출과 직결돼서 장애 허용성(fault tolerance)이 보장돼야 하고, 지연(latency)이 최대 100ms 이내의 빠른 응답으로 타임아웃(timeout)도 없어야 한다. 단순히 서버의 수를 늘려서 처리량(throughput)을 높이고 장애 백업을 지원한다는 의미만은 아니다. 소프트웨어 측면에서 알고리즘의 복잡도를 낮추고 문제가 발생했을 때 역추적도 가능해야 한다는 의미다. 딥러닝이 여러 분야에서 두각을 나타내고 있지만, 광고 시스템에서 여전히 보수적인 이유도 속도와 안정성, 그리고 사후 문제 재현 및 해결 가능성이라는 측면이 있다. 시시각각으로 변하는 사용자들의 반응을 실시간으로 반영하기 위해서 온라인으로 학습이 가능한 알고리즘이 필요하다.

마지막으로, 시스템 이슈는 아니지만, 광고랭킹은 사용자(오디언스)와 광고주 양쪽을 동시에 만족시켜야 한다. 극단적으로 얘기하면, 사용자 만족을 위해서 광고를 전혀 내보내지 않거나 광고주를 위해서 무조건 많은 광고를 내보내면 된다. 그렇게 되면 매출을 올리지 못하거나 광고 때문에 사용자들이 서비스를 떠나게 된다. 어떤 광고를 누구에게 언제 내보내야할 것인가를 미세하게 잘 조절할 수 있는냐가 광고 서비스의 승패를 좌우한다.

pCTR 예측

광고랭킹 알고리즘을 설명하기에 앞서 온라인 광고의 발전 과정을 간략히 살펴볼 필요가 있다. 인터넷 초기에는 배너 광고를 일정한 시간동안 노출시켜주는 CPT(cost per time) 상품이 1,000회 노출당 과금하는 CPM(cost per mille) 상품이 중심이었다. 일정 시간동안 노출 회수를 보장하면 되기 때문에, 시간당 과금액 또는 노출당 과금액으로 광고를 정렬한다. 이후 검색광고가 등장하면서 클릭당 과금(cost per click, CPC) 상품이 등장하고, 배너 광고도 CPC로 판매하는 성과형 광고가 등장했다. CPC 상품은 eCPM으로 환산해서 랭킹하기 때문에 정확히 pCTR을 예측하는 방법이 핵심이다. 최근에는 더욱 직접적인 전환을 기준으로 과금하는 CPA(cost per action) 상품이 등장하고, CPC 상품과 유사하게 예측전환율 (pCR)⁶로 eCPM을 계산한다. 현재 CPC 상품을 중심으로 광고 시장이 형성돼있으므로 pCTR 예측 방법을 중심으로 설명한다.

과거자료 CTR(historical CTR, hCTR)

pCTR을 예측하는 가장 쉬운 방법은 과거자료(historical data)를 집계하는 거다. 예를 들어, 모든 ‘(추정) 20대 남성’의 지난 일주일동안 광고별로 노출과 클릭 회수를 수집한다. 광고 A는 100,000회 노출에 1,500회 클릭이 발생했고, 광고 B는 70,000회 노출에 1,200회 클릭이 발생했다면 광고 A의 pCTR = 1,500 / 100, 000 = 1.5%이고 광고 B의 pCTR = 1,200 / 70,000 = 1.7%가 된다. 만약 광고 A와 B 모두 입찰가(BA)를 2,000원으로 설정했다면, 다음 번에 (추정) 20대 남성 U가 접속했을 때 eCPM(A) = 0.015 * 2,000 * 1,000 = 30,000원, eCPM(B) = 0.017 * 2,000 * 1,000 = 34,000원이 되어, eCPM이 높은 광고 B를 노출시키고 클릭이 발생하면 1,765 (= 2,000 * 30,000 / 34,000) 원을 과금한다. 과거자료만 누적, 집계하면 되는 가장 간단한 방법이다. 전형적인 MAB(multi-armed bandit) 또는 탐색-탐험(exploration-exploitation) 문제/접근법이다. 뒤쪽에서 언급할 톰슨 샘플링(Thompson sampling)과 결합해서 진화된 방식으로 해결할 수 있으나, 단순 과거자료 집계 방식은 한계가 명확하다.

첫째, 모든 오디언스 그룹, 광고 소재, 그리고 게재 지면 조합별로 CTR 데이터를 수집해야 한다. 단일 지면에서, 성별과 연령대로 구분하고, 10여 개의 광고 소재만 존재한다면 200개 미만의 조합으로 노출/클릭 데이터를 집계하면 되겠지만, 카카오의 경우 1,000개가 넘는 게재 지면을 가지고 있고 오디언스 피처(feature)는 최소 수십만 차원에 이르고, 광고 소재의 수도 10,000개가 훨씬 넘는다. 그리고 매일 새로운 광고 소재가 등록되고 시의성 없는 소재가 제외된다. 오디언스와 광고 소재만 고려하더라도 수천만 개의 세그먼트(segment) 별로 CTR 데이터를 집계해야 한다.

둘째, 강건한(robust) CTR 예측을 위해서 세그먼트 별로 일정 회수 이상의 노출이 보장돼야 한다. 예를 들어, 100회 노출에서 5회 클릭이 발생했을 때 CTR = 5%라고 추정한다면 신뢰할 수 없다. 세그먼트 별로 10,000회 정도 노출해서 517회 클릭이 발생한다면 그제서야 CTR = 5.17%라고 예측할 수 있다. 모든 세그먼트 별로 신뢰할 수 있는 수준의 CTR을 얻기 위해서 최소 회수의 노출을 보장하는 것은 트래픽이 많은 서비스에서도 부담이 된다.

셋째, hCTR은 요일 및 시간에 따른 변동성(fluctuation)을 뭉개버린다. 지난 일주일 데이터로 CTR = 5%를 예측하더라도 현재/미래의 트래픽에서 4%가 될지 6%가 될지 장담할 수 없다. 카카오도 초기에는 최소의 세그먼트로 hCTR을 적용했지만 예측력이 많이 낮아서 결국 폐기하고, CTR 예측모형을 만들어서 pCTR을 계산하는 걸로 선회했다. 반직관적이지만 실측치가 예측치보다 못하다.

로지스틱 회귀모델(logistic regression)

CTR의 정의는 노출 횟수(impressions) 대비 클릭 횟수(clicks)의 비율이다. 즉, CTR = clicks / impressions로 계산한다. 전체 노출 횟수에 대한 전체 클릭 횟수의 비율을 달리 표현하면, 1회 노출당 클릭될 확률을 구하는 것과 같다. 오디언스 U에게 광고 A를 노출했을때 오디언스가 그 광고를 클릭할 조건부 확률을 구하는 문제다(pCTR = Pr(click = 1 | impression)). 이런 문제에 적합한 해결책은 최대 엔트로피(maximum entropy)로 알려진 로지스틱 회귀모델(logistic regression)이다. 각 회사의 광고 플랫폼마다 사용하는 피처의 종류와 엔지니어링 방법만 조금씩 다를 뿐, 현존하는 대부분의 광고랭킹은 로지스틱 모델에 기반한다.

로지스틱 회귀모델을 간략히 설명하면 다음과 같다. 학습 데이터 세트 (xi, yi)가 주어졌을 때, xi는 0 또는 1을 갖는 d 차원의 성김 피쳐 벡터(sparse feature vector)이고, yi는 -1 또는 1을 갖는 목표값이다. 회귀식을 구할 때 실수값 x로 계산할 수 있지만, 광고랭킹에서는 보통 이산화(discretization)해서 바이너리 값을 갖게 한다. 카테고리 데이터도 모두 임의(dummy)변수로 만든다. 이때 d 차원의 가중치 벡터 w를 찾는 것이 목적이다. 샘플 x가 y = 1로 예측할 확률은 다음과 같다.

$$Pr(y = 1 \mid x, w) = \frac{1}{1 + \exp(-w^T x)}$$

이때 가중치 벡터 w를 구하기 위해서 아래의 부정 로그-확률 손실(negative log-likelihood loss)함수를 계산한다.(L2 regularization을 사용함)

$$\min_w \frac{\lambda}{2} \| w \|^2 + \sum_{i=1}^n \log(1 + \exp(-y_i w^T x_i))$$

위의 손실 함수는 볼록성(convex) 최적화 문제로 보통 기울기(gradient) 기반의 최적화 방식으로 쉽게 해결된다. 충분히 많은 학습 데이터 세트가 있다면 모델 가중치 w는 쉽게 구할 수 있다. 그런데 배치(batch) 최적화로 가중치를 계산하는 것은 시간이 오래 걸려서 최신의 피드백 (클릭/전환 데이터)을 바로 반영하지 못한다. 그래서 확률적 확률적경사하강(stochastic descent gradient, SGD) 방식의 온라인 최적화 방식을 주로 이용한다⁷.

피쳐 데이터와 차원 축소(feature engineering)

온라인 최적화 방식(FTRL-Proximal)을 설명하기 전에 카카오에서 사용하는 각종 피쳐 데이터(feature data)와 전처리 과정부터 알아보자. 정확한 pCTR을 예측하기 위해서 다양한 데이터를 활용하고 있다. 구체적으로 어떤 데이터를 어떻게 수집하고 활용하느냐는 회사마다 갖는 노하우(know-how)이므로 자세히 다루지 못함을 양해바란다.

첫째, 오디언스의 데모 정보를 추정해 사용한다. 카카오에서의 다양한 활동을 분석해서 베이지 추론(bayesian inference)으로 성별과 연령대를 추정한다. 주 활동지 등의 지역은 IP-지역 매핑 데이터(IP mapping data) 등을 활용해서 추정하고 있다.

둘째, 다음검색을 통해 유입된 검색어 이력이나 소비한 카페나 스토리채널의 메타데이터를 분석해서 오디언스의 관심사로 매핑하고 있다.

광고 소재의 경우 광고주ID, 개별 소재ID, 소재의 크기와 타입, 소재 등록/검수 시에 지정한 카테고리 등을 이용하고 있다. 소재의 URL, 텍스트 또는 이미지 등에서 객체를 추출해서 사용하는 것도 테스트 중이나 앞서 해결할 문제들이 있어서 아직 상품에 적용하지는 않았다. 그리고 게재지면(매체)의 지면ID와 publisherID, 카테고리, 그리고 지면별 반응을 등을 활용하고 있다. 마지막으로 요일과 시간 등의 맥락 정보도 활용한다.

개별 오디언스 피쳐 벡터(수십만 차원), 광고 피쳐 벡터(수천 차원), 지면 피쳐 벡터(수천 차원)만으로도 데이터 차원이 매우 커서 연산이 거의 불가능하다. 그런데 이런 피쳐들의 상호작용(interaction) 효과를 반영하기 위해서 이중 피쳐 간의 데카르트 곱(cartesian product)을 취한다. 그렇게 되면 적게는 수백만 차원, 많게는 수억 차원의 피쳐 벡터가 만들어진다. 로지스틱 회귀모델에 L₁ 정규화(regularization)를 적용해서 피쳐 선택(feature selection) 기능을 넣더라도 연산이 불가능하다. 그래서 최소한의 정보 손실을 감내하더라도 연산이 가능하게 데이터의 차원을 축소하는(data reduction) 기술이 필요하다. 저차원의 데이터라면 특이값 분해(singular value decomposition, SVD)나 주성분분석(principal component analysis, PCA)으로 차원 축소가 가능하지만, 대용량 데이터는 전체를 메모리에 올려서 계산해야 하기 때문에 현실성이 적다. 그래서 야후!(Yahoo!)와 크리테오(Criteo)에서는 해싱 트릭(hashing trick)을 이용하고, 페이스북은 그라디언트 부스팅 머신(gradient boosting machine, GBM)을 이용한 부스티드 결정 트리(boosted decision tree)를 사용한다. 구글은 블룸 필터(Bloom filter)를 이용해서 새로운 피쳐가 기존 모델에 추가되는 것을 최소화하고 있다.

카카오에서는 현재 클러스터링 기법을 활용 중이다.

Cartesian product된 모든 피쳐로 클러스터링하는 것이 현실적으로 불가능해서, 유사한 오디언스나 지면을 묶는 식으로 추상화한다. 축소된 오디언스, 지면, 그리고 광고 피쳐를 cartesian product해서 사용한다. 수십만 차원의 오디언스 데이터로 클러스터링하는 것도 부담되기 때문에, 우선 랜덤프로젝션이나 해싱트릭으로 차원을 축소한 후에 클러스터링을 하면 어느 정도의 정보 손실은 있지만 큰 부담없이 연산할 수 있다. 차원 축소에서 클러스터링이 좋은 이유는 초기 구축 비용이 많이 들어가지만, k개의 중심값(centroids)만 계속 업데이트하면 되기 때문에 데이터 연속성을 보장하면서 연산량은 별로 크지 않다는 점이다.

위에서 피쳐의 차원을 축소하는 방법을 기술했는데, 샘플의 개수를 줄이는 방법도 고려할 수 있다. 실제 CTR을 확인하면 1% ~ 10%로 매우 낮다. 즉, 클릭으로 연결된 노출의 비율이 매우 낮다는 의미다. 그래서(x, y) 쌍에서 90% 이상이 y=-1인 샘플이다. 그래서 일정 비율의 y=-1 샘플만으로 로지스틱 회귀모델을 구축한 후에 보정하는 방법을 사용한다. y=1인 샘플은 상대적으로 매우 적기 때문에 모든 샘플을 학습 데이터로 사용한다. 그외에도 랜덤 샘플링(random sampling)을 포함해서 코어셋(coresets)이나 스케치(sketch) 등의 기법으로 핵심 샘플만 남겨서 클러스터링(clustering)의 초기 중심점을 구하는 방식도 고려할 수 있다.

광고 플랫폼을 제공하는 회사들은 많이 있다. 대부분의 회사들은 다음에 소개할 FTRL(follow the regularized leader) 기법으로 로지스틱 회귀모델을 만들어서 pCTR 예측모델로 사용하고 있다. 예측모델과 방법론은 대동소이하다는 거다. 그럼에도 회사마다 예측 품질에 차이가 나는 것은 각 플랫폼마다 어떤 피쳐를 사용하고 있으며 (또는 얼마나 다양하고 많은 오디언스 정보를 수집하고 있으며) 그런 피쳐들을 어떻게 전처리하느냐에 달려있다. 페이스북은 방대한 소셜 그래프(social demography)를 활용하고, 구글은 사용자의 니즈가 명확한 검색어를, 그리고 크리테오는 광고주의 랜딩URL에 심어둔 리타게팅 스크립트(re-targeting script)를 통해 오디언스 정보를 수집, 정교화한다. 전처리에서는 페이스북은 결정 트리(decision tree)로 전처리를 해서 3% 이상의 개선 효과를 얻었지만, 다른 여러 휴리스틱(heuristic)을 적용했을 때는 추가 개선을 별로 얻지 못했다고 한다*⁸. 또 크리테오는 해싱 트릭을 사용하지만, 구글의 실험 결과에서 해싱 트릭은 별로 효과가 없었다고 한다*⁹*¹⁰.

온라인 모델 업데이트, FTRL-Proximal

여기까진 FTRL을 설명하기 위한 배경지식이다. 가장 최근에 나온 버전이 FTRL-Proximal인데, 이름에 사용된 용어를 알면 그 의미를 쉽게 알 수 있다. FTRL은 follow the regularized leader의 약자다. 최초에는 FTL이 소개됐고, 후에 정규화(regularization)가 포함된 FTRL이 됐고, 여기에 근접치(proximal)가 추가된 형태로 진화했다.

왜곡이나 편향이 없다면 학습 데이터의 양이 많을수록 더 정확하고 강건한 예측모델이 만들어지는 것은 기계학습의 기본이다. 하지만 학습 단계에서 전체 모집단(population)을 잘 설명하는 모든 데이터를 얻는 것은 현실적으로 불가능하다. 광고에는 새로운 유형의 오디언스가 계속 등장하고, 새로운 광고주와 광고 소재가 등록 (또는 폐기)되고, 새로운 성향의 매체 지면이 추가되는 것이 빈번하다. 초기 모델을 학습하면서 그런 변동성을 모두 파악/반영하는 것은 불가능하다. 그리고 정상 학습 데이터도 양이 많아지면 학습하는데 많은 메모리와 시간이 필요하다. 참고로 카카오는 매일 수억 건의 광고 요청을 받아서 광고 소재를 노출시키고 수천만 건의 클릭 등의 포스트액션 데이터를 수집한다. 학습을 위해서 일주일, 한달 치의 데이터를 모두 사용한다면 많은 서버를 투입해도 수시간이 필요하다. 데이터의 변동성과 학습 효율을 고려해서 분할 정복(divide & conquer) 방식을 이용해 시간 단위로 분할 학습과 통합 모델 구축 전략을 사용한다. 이 전략의 가장 최신 버전이 FTRL-Proximal이라고 보면 된다.

일반적인 최적화 방식인 온라인경사하강(online gradient descent, OGD)은 현재 타임프레임 t에서의 가중치 w_t와 기울기 g_t를 이용해서 t + 1에서의 가중치를 구하면 다음과 같다.

$$w_{t+1} = w_t - \eta_t g_t$$

여기서 g_t 앞에 있는 값은 아래와 같은 단조감소(non-increasing)하는 학습 파라미터(parameter)다.

$$\eta_t = 1/\sqrt{t}$$

FTRL-Proximal은 대신 아래의 식으로 가중치를 업데이트한다.

$$w_{t+1} = \underset{w}{\operatorname{argmin}} \left(g_{1:t} \cdot w + \frac{1}{2} \sum_{s=1}^t \sigma_s \|w - w_s\|_2^2 + \lambda \|w\|_1 \right)$$

위의 식은 크게 세 부분으로 나뉘어있다.

첫째, FTL(follow the leader)를 표현한 것으로, t 시간까지 가장 작은 손실을 갖도록 한 leader(gradient)들을 취해서 손실의

근사치를 구한다.

둘째, 새로운 가중치가 이전에 사용했던 가중치들에서 큰 변동이 없도록 제약하는 근접치(proximal) 파트다. 이런 강한 볼록성(convexity)을 추가함으로써 알고리즘의 안정성을 높인다.

마지막 L₁정규화(regularization) 파트이다. 이들을 하나로 합친 것이 FTRL-Proximal인데, 현재까지의 최고 서브그라디언트(subgradient)의 궤적을 쫓으며 근사한 손실을 최소화하고 규칙화와 근접치를 통해서 모델의 급격한 변동성을 막는 안정된 모델을 구축한다. 온라인 최적화 방식이 모든 학습 데이터로 한꺼번에 계산하는 배치(batch) 방식보다 정확도는 다소 낮을 수 밖에 없다. 그럼에도 빠른 학습/업데이트 시간과 최소 리그렛 바운드(regret bound)를 보장하는 최적화 방법 중에 FTRL-Proximal이 현재의 최신기술(state-of-the-art, SOTA)이다. FTRL보다 더 나은 예측 정확도를 보여주는 알고리즘들이 계속 소개되고 있지만 제품 단계에서 적용하는 데는 아직 한계가 있다. FTRL-Proximal에서 규칙화 와 근접치 부분을 조금 변경한 FOBOS, AOGD, RDA 등의 방식도 함께 연구됐지만, 현재까지는 FTRL-Proximal이 가장 안정된 성능을 보여준다¹¹.

피크 타임에는 초당 수십만 건의 광고 요청이 들어오고, 모든 요청은 30~40ms (길어도 100ms) 내에 가장 적합한 광고를 선택해서 내보내야 한다. 그래서 광고 랭킹에 사용하는 알고리즘은 빠른 연산과 안정된 서빙이 가능해야 한다. 그래서 빠른 알고리즘에 더해서 사용하는 피쳐의 수도 최소화하려는 거다. 앞서 설명한 차원 축소가 정확한 예측 모델 구축 뿐만 아니라, 안정된 서빙(serving)을 위해서도 중요하다. 보통 예측 시간보다 모델 구축 시간이 오래 걸린다. 만약 딥러닝을 이용한 방식이 더 정확한 예측력을 제공한다면 오프라인에서 딥러닝 모델을 구축하고 온라인에서 딥러닝 기반으로 예측만 하면 될 것 아니냐?라고 반문할 수 있다. 합리적 추론이지만, 변동성이 심한 인터넷 트래픽 환경에서는 모델의 새로움(freshness)을 유지하는 것이 중요하다. 즉, 하루 이틀동안 학습한 모델을 일주일, 한달동안 계속 사용할 수가 없고, 수분에서 1시간 주기로 새로운 데이터로 모델을 계속 업데이트해줘야 한다. 딥러닝 모델이 예측력이 더 낮다는 것이 증명되고 효과적인 업데이트가 가능해지면 광고 랭킹시스템도 진일보할 것이라고 본다.

기타 이슈

이상으로 광고랭킹에 사용하는 피쳐 작업과 FTRL 알고리즘 기반의 pCTR 예측모델 학습을 알아봤다. 광고랭킹에서 pCTR을 예측하는 것이 가장 중요한 작업이지만, 이외에도 해결해야할 데이터 및 기계학습 문제들이 여럿 있다. 이번 글에서는 모든 기술적 디테일과 해결책을 설명하기 보다는 개념적으로 문제와 해결 방식을 간략히 소개하려고 한다. 자세한 것은 참고 논문을 읽기 바란다.

톰슨 샘플링(Thompson sampling)

앞서 hCTR을 소개하는 단락에서 톰슨 샘플링을 언급했다. 안정적인 hCTR을 얻기 위해서 충분한 노출 기회가 필요한데, 현실적으로 모든 세그먼트에게 많은 노출 기회를 균등하게 제공할 수 없다. 그래서 사후확률(posterior)을 고려해 더 잠재력 높은 세그먼트에 더 많은 트래픽을 몰아줘서 빠르게 수렴하게 한다. 예측 모델을 이용한 pCTR 계산에서도 톰슨 샘플링을 사용한다. 신규 등록된 광고 소재의 정확한 pCTR을 예측하기 위해서 다양한 트래픽에서 충분한 노출과 클릭 데이터를 수집해서 학습한다. 초기 빈약한 데이터에서도 pCTR이 높게 나온다면 자연스럽게 eCPM 입찰 경쟁에서 이겨서 노출 기회를 얻고, 실제 많은 노출에서 실측한 CTR이 pCTR보다 낮게 나오면 자연스레 온라인 업데이트를 통해서 pCTR이 낮게 조정된다. 반대 상황에서 낮게 예측된 부정확한 pCTR은 해당 광고 소재의 노출 기회를 박탈해서 정확한 pCTR 모델의 구축이 불가능하다. 예를 들어, 과거의 다른 광고들을 집행하며 얻은 통찰에 따라서 예측된 CTR의 잠재력이 15%인데, 초기 1,000회 노출에서 120회의 클릭을 받아서 pCTR이 12%로 예측하도록 초기 모델이 만들어졌다면, 잠재력이 더 크기 때문에 인위적으로 더 많은 노출 기회를 제공해서 실제 CTR이 잠재 CTR인 15%가 맞는지 아니면 초기 실측치인 12%가 맞는지를 확인하면서 예측모델을 조정하게 된다¹².

광고 품질 지수

많은 사람들이 광고를 싫어한다. 전세계적으로 PC에서 Adblock을 설치한 브라우저 수가 2억개를 넘었고, 모바일에서는 약 4억개의 Adblock이 설치됐다는 조사 결과가 이를 입증한다¹³. 그렇기 때문에 광고 플랫폼 사업자는 오디언스에게 더 적합한 광고를 제공해 서비스 경험을 해치지 않게 하려 노력한다. 그럼에도 부적절한 소재와 형식의 광고가 노출되는 것을 완전히 사전에 차단할 수는 없다. 때로는 모든 광고를 싫어하는 오디언스도 존재한다. 그래서 다양한 형태의 오디언스 피드백을 모아서 광고 기피 사용자에게 광고를 덜 보여주기도 하고, 오디언스별 선호도에 따라서 연관성이 낮은 광고 소재를 사전에 필터링하는 알고리즘도

적용한다. 명시적으로 ‘이 광고 보기 싫어요’와 같은 피드백도 수집하지만, 우발적으로 클릭된 데이터를 향후 모델 학습에서 배제하는 등의 네가티브 피드백(negative feedback)을 광고 품질 지수에 포함시켜서 활용한다.

CVR(conversion rate) 예측

과거의 온라인 광고는 단순히 브랜드 홍보를 위해서 광고 노출 기회를 얻거나 광고를 통해서 광고주의 홈페이지로 사용자를 유입시키는 것이 목적이었다. 하지만 최근에서 광고 노출 단계에서 부터 광고의 목적, 즉 전환을 고려해서 광고를 집행하고 있다. 광고의 목적은 모바일 앱을 설치하고 실행한다거나 소핑몰에 가입해서 물건을 구매한다거나 자동차 시승 이벤트에 응모를 한다는 등으로 다양하다. 이런 다양한 목적 별로 적합한 오디언스가 다르기 때문에, 오디언스 별로 광고의 전환율을 예측하는 것이 중요해졌다. CVR (클릭 후 전환율 = Conversions / Clicks)의 예측도 앞서 설명한 로지스틱 회귀모델과 FTRL 알고리즘으로 가능하다. 하지만, 노출과 클릭 데이터에 비해서 전환 데이터는 수집하기 어렵고, 외부 CTS/MAT(conversion tracking system / mobile app tracker)를 통해 수집된 데이터도 유효성 검증이 필요하다. 유효한 전환 데이터를 제대로 수집하더라도 노출, 클릭에 비해서 전환수가 극히 적기 때문에 신뢰도 높은 전환 예측 모델을 만들기도 어렵고, 광고 클릭 후 전환 발생까지의 시간 차가 커서 앞서 설명한 방식을 그대로 적용하는데 어려움이 있다. 지연된 피드백 모델(delayed feedback model)과 같은 알고리즘도 제시됐지만 제품화를 위해서 넘어야할 장애물이 많다¹⁴.

어뷰징 처리(fraud detection)

광고 서비스는 현금이 거래되기 때문에 다양한 형태의 어뷰징(abusing)이 이뤄진다. 경쟁사의 광고비를 빨리 소진시키기 위해서 경쟁사의 광고를 거짓으로 클릭하는 경우 (CPC 상품인 경우)도 존재하고, 매체를 제공하는 발행자(publisher)가 수익을 더 얻기 위해서 자신의 사이트에 노출된 광고를 과하게 클릭하는 경우도 있다. 이런 류의 단순 어뷰징은 간단한 로직으로 해결할 수 있지만, 계속 지능화되기 때문에 100% 잡아내기가 어렵다. 어뷰징은 아니지만 사용자의 실수로 여러 번 클릭해서 이중으로 과금하는 경우도 있다. 이런 어뷰징이나 우발적 실수에 의한 과금은 분석을 통해서 비과금 처리한다. 그리고 왜곡 데이터가 모델 학습에 사용되지 않도록 차단하는 것도 중요하다.

인수분해(factorization)

최근 FTRL보다 더 나은 성능을 보이는 알고리즘이 소개됐다.

최초의 아이디어는 행렬 인수분해(matrix factorization)과 SVM(support vector machine)의 개념을 결합한 인수분해기계(factorization machines)였는데, 향후 개별 피쳐(feature)들의 메타정보(field)를 모델에 반영해서 FFM(Field-aware factorization machine)으로 발전했다. FTRL을 SOTA라고 설명했지만, pCTR 예측에서 FFM이 로지스틱 모델보다 3~4% 이상 정확도가 높다고 보고됐다. 하지만 로지스틱 모델과 달리 FFM은 가중치를 잠재 벡터(latent vector)화했기 때문에 연산량과 메모리 사용량이 더 높은 단점이 있다. 하지만, 최근 논문에서 다양한 시스템 튜닝(system tuning)을 통해 실제 광고 서버에 사용하는데 큰 지장이 없다는 것을 보였기 때문에 카카오에서도 어떻게 제품에 적용할지 고민 중에 있다¹⁵ ¹⁶.

테스트와 제품

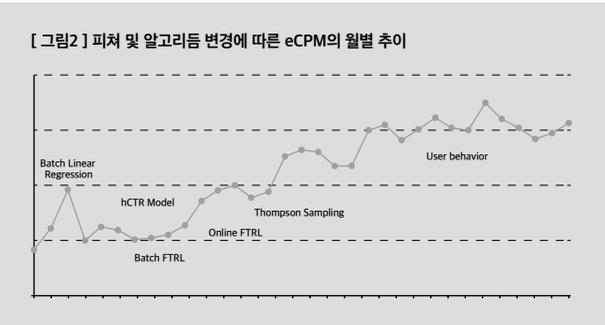
많은 애드테크(AdTech) 연구자들이 계속 새로운 문제와 해결책을 소개하고 있고, 카카오 내에서도 새로운 개선책을 계속 발굴하고 있다. 가장 좋은 것만 취사선택해서 시스템에 바로 적용하면 매출이 극적으로 오를 것 같지만 현실은 그렇지 않다. 우선 개별 아이디어가 실제 카카오 광고에서 유효한지를 판별해야 하고, 앞서 강조 했듯이 제품 단계로 확장(scale up)하는 과정이 필요하다. 그래서 새로운 아이디어의 구현과 테스트를 상시 진행한다. 구현 단계에서의 유닛 테스트를 제외하고, 테스트는 크게 3단계로 이뤄진다. 고정된 테스트 데이터를 이용해서 이전의 알고리즘들과의 성능 차이를 1차로 확인한다. 개선 효과가 있다고 판단되면 최근 일주일에서 한달동안 누적된 실제 집행 로그를 가져와서 현재 적용된 알고리즘과 새 알고리즘을 비교하는 오프라인 테스트를 거친다. 오프라인 테스트를 통과했다면 실제 트래픽에서 5~10%정도를 별도의 꾸러미(bucket)로 설정해 온라인 버킷 테스트 (또는 A/B 테스트)를 거친다. 온라인 테스트에서 유효성이 검증됐다면 버킷의 양을 차츰 늘리거나 전체 트래픽에 적용한다. 알고리즘 개선으로 매출 증대하는 것이 목적이지만, 기존 환경에 익숙한 사용자들이 느낄 충격을 최소화하기 위해서 다양한 테스트 과정과 점진적 트래픽 확대라는 전략을 취한다.

길게 설명하지는 않겠지만 이외에도 다양한 문제들이 있다. 개별 광고 소재의 과금액을 결정하고, 특히 CPM이나 CPC 등의 다른 가격 정책의 상품들이 혼재된 경우 이들 사이의 가격을 조정하는 프라이싱 문제가 있다. 개인별 네가티브 피드백을 분석하듯이 개인별 광고 피로도를 측정해서 같은 소재의 광고를 몇 회까지 보여줄 것인가를 정하는 프리퀀시 캐핑 문제도 있다. 광고주가 설정한 타겟에 따라서 특정 오디언스에게만 광고를 보여주거나 배제하는 것과 유사한 고객을 찾아서 광고를 부스팅하는 등의

개인화 문제도 알고리즘으로 해결해야 한다. 실측 CTR과 예측 CTR 간의 간극을 조정하는 캘리브레이션(calibration)도 공통 문제다. 최근에는 이른바 oCPM (optimized CPM)이라 불리는 상품이 등장해서 광고주가 입찰금액을 정하는 것이 아니라 알고리즘이 광고 목적과 연계해서 오디언스별로 다른 입찰금액을 제시하는 자동입찰 문제도 있다. 이상은 모두 DSP 관점에서의 문제였는데 SSP 측면에서의 최적화 문제도 함께 고민하고 있다. 이외에도 데이터 기반으로, 그리고 알고리즘으로 해결해야할 수많은 문제들이 있다.

카카오 광고의 성과와 향후 과제

이번 글에서 그동안의 노력과 성과를 모두 나열할 수는 없지만, 아래의 그래프를 보면 어느 정도 가늠할 수 있다. 2014년 7월에 카카오 스토리에 처음 배너 광고가 노출된 후로 약 3년 간의 eCPM의 월별 추이와 앞서 설명한 여러 알고리즘을 적용한 시기를 대략적으로 표시한 그래프다. 스토리 서비스의 개편이나 사용자 트래픽의 증감과 같은 다양한 외부 요인이 있었지만, 초기에 비해서 eCPM이 약 3배 증가했음을 확인할 수 있다. 어떤 알고리즘/개선책은 오프라인 테스트와 온라인 테스트에서 효과가 입증돼서 모든 트래픽에 적용했지만, 장기간의 추적을 통해서 결과적으로 효과가 없음이 나중에 밝혀져 시스템에서 빠진 경우도 있다. 대표적으로 1년 정도 유지하던 hCTR을 2016년도에 제외했다. 강력한 하나의 슈퍼 알고리즘이 온라인 광고의 승패를 결정하는 것이 아니라, 여러 적정 기술과 알고리즘들이 결합해서 광고 생태계를 이룬다.



많은 노력으로 광고랭킹 알고리즘을 개선하고 있지만 여전히 해야할 일이 많다. 가장 핫한 이슈인 ‘딥러닝을 광고 시스템에 어떻게 적용할 수 있을까?’도 해결해야 한다. 우선은 딥 오토인코더 (deep auto encoder) 같은 기술을 이용해서 피쳐 벡터의 차원을 축소하는 문제부터 적용해 볼 수 있다. 또는 RNN/CNN 등의 기술을 활용해서 광고 소재 및 설명에서 유효한 객체를 찾아내고, 또 오디언스들이 소비한 콘텐츠에서도 유효 객체를 찾아서 매핑하는

것도 진행해야 한다. 로지스틱과 FTRL을 대체할 딥 아키텍처(deep architecture)도 가능하다고 본다. 딥러닝이 장기적인 과제라면, 앞서 설명한 FFM을 우선 프로덕션에 적용하는 것은 단기 과제다.

광고에서 1~2%의 개선도 전체 매출에서 큰 부분을 차지한다.

많은 광고의 CTR, CVR은 1% 미만이다. 어떤 광고는 이보다 더

낮은 값을 갖기도 한다. 하지만 만약 매우 낮은 CTR의 광고더라도

적절한 순간에 적절한 오디언스에게 전달된다면 그 오디언스에게는

가치있는 (CVR이 높은) 정보를 제공하는 셈이다. 9,999명에게는

불필요한 정보지만, 1명에게는 꼭 필요할 수도 있다. 그렇듯이

아주 희귀한 사건(rare event)에 부합하는 피처를 발굴해서 정확히

예측하는 모델도 개발해야 한다. SSP 관점에서의 최적화 문제,

정확한 CVR 예측을 위한 피처 발굴과 모델 학습, oCPM을 비롯한

자동 입찰 로직 개발, 트래픽 예측과 예산 분배, 광고의 효과를

정확히 평가하는 복수터치기여(multi-touch attribution) 문제 등

해결해야할 문제가 여전히 많다.

결언

지면 관계상 많은 디테일을 생략해 다소 어렵거나 재미없을 수도

있지만 광고 랭킹과 관련된 많은 문제들이 존재하고, 데이터

관점에서, 그리고 알고리즘으로 하나씩 그런 문제들을 해결해가는

것은 매우 흥미로운 과정이다. 이 분야에는 여전히 도전을 기다리는

수많은, 재미난 문제들이 존재한다. 앞서 보여준 성과보다 더

아름다운 이상향 그래프를 그려갈 수 있기를 기대한다.

***1** 참고 : 알파벳 회사 설명. <https://abc.xyz/> ***2** 참고 : Netflix recommendations: Beyond the 5 stars <http://kko.to/IUNr9VxEA> ***3** 참고 : 광고에서 '사용자'를 audience라 부름. ***4** 참고 : 카카오는 개인정보 보호를 위해서 사용자의 실제 성별이나 연령을 사용하지 않고, 카카오 내에서의 다양한 활동을 추적해서 배이지 추론 등의 기법으로 데모 정보를 추정해서 사용한다. 그리고 k익명성을 고려해서 오디언스의 정보를 추상화 및 군집화해서 관리한다. 광고 효율은 다소 낮아질 수 있으나, 인터넷에서의 개인정보 보호라는 더 큰 원칙을 준수하려 노력한다. ***5** 참고 : eCPM = BA(billing amount) x pCTR(page click-through rate) ***6** 참고 : 전환율은 보통 CR(conversion rate) 또는 AR(action rate)라고 부르는데, 이는 노출당 전환 비율 (= #conversion / # impressions) 을 뜻한다. 논문 등에서 CVR이라는 용어도 사용하는데, 이는 포스트 클릭 전환, 즉 클릭된 회수 대비 전환 비율 (= #conversions / # clicks)을 계산한 것이다. 전환은 매우 희귀한 이벤트라서 보통 CR은 CTR와 CVR로 나눠서 구한다(CR = CTR * CVR). ***7** 참고 : How to implement logistic regression with stochastic gradient descent from scratch with Python <http://kko.to/IUOe6qDp6> ***8** 논문 : He, Xinran, et al. "Practical lessons from predicting clicks on ads at facebook." Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. ACM, 2014. ***9** 논문 : McMahan, H. Brendan, et al. "Ad click prediction: a view from the trenches." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013. ***10** 논문 : Chapelle, Olivier, Eren Manavoglu, and Romer Rosales. "Simple and scalable response prediction for display advertising." ACM Transactions on Intelligent Systems and Technology (TIST) 5.4 (2015): 61. ***11** 논문 : McMahan, Brendan. "Follow-the-regularized-leader and mirror descent: Equivalence theorems and l1 regularization." Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 2011. ***12** 논문 : Chapelle, Olivier, and Lihong Li. "An empirical evaluation of thompson sampling." Advances in neural information processing systems. 2011. ***13** 참고 : 2017 AdBlock report <http://kko.to/IUOSMO3UQ> ***14** 논문 : Chapelle, Olivier. "Modeling delayed feedback in display advertising." Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014. ***15** 논문 : Juan, Yuchin, et al. "Field-aware factorization machines for CTR prediction." Proceedings of the 10th ACM Conference on Recommender Systems. ACM, 2016. ***16** 논문 : Juan, Yuchin, Damien Lefortier, and Olivier Chapelle. "Field-aware factorization machines in a real-world online advertising system." Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 2017.

“강력한 하나의 슈퍼 알고리즘이

온라인 광고의 승패를 결정하는 것이 아니라,

여러 적정 기술과 알고리즘들이 결합해서 광고 생태계를 이룬다”

AI는 인간과 같은 감정을 가지게 될까?

감정을 매개로 한 AI와 인간의 관계 형성 필요

들어가기 앞서 간단히 용어를 정리하고자 한다. 이 글에서 말하는 감정이란 인공지능 혹은 로봇이 스스로 갖고 있는 감정을 의미하며, 사람이 갖고 있는 그것과 유사한 개념으로 가정된다. 감정, 정서, 감성, 느낌, 기분 등의 여러 표현이 있을 수 있지만, 이 글에서는 감정이라는 단어를 위주로 사용했다. 혹시 심리학을 전공하신 분이라면 이를 구분하여 사용하지 않았다고 너무 분노하지 않으셨으면 좋겠다. 감정은 지능과 별개인가? 아니면 감정도 지능의 일부인가? 인공지능에 감정이 꼭 필요한가? 인공지능의 관심이 높아지면서 다른 한편으로는 조심스럽다. 아직 인간의 정서에 대한 연구도 가야 할 길이 먼 마당에 이를 인공지능에까지 비교해야 하니 말이다. 그래도 흥미로운 주제인 만큼 공학뿐 아니라 심리학, 행동학, 디자인, UX 등의 다양한 분야에서 이 주제에 대한 지속적인 논의가 이어질 수 있기를 바라는 마음에 글을 이어 가본다.

의인화된 인공지능 로봇

공상과학 영화에 나오는 미래의 인공지능 로봇들을 보면 모두 하나같이 감성적인 특징을 가지고 있다. 영화 2001 스페이스 오디세이(1968)에서부터 터미네이터(1984), A.I.(2001), I, Robot(2004), 채피(2015) 등에 이르기까지 수많은 영화에서 인공지능 로봇은 어떨 때는 공격성을 가진 모습으로, 어떨 때는 인간과 교감하며 친근한 모습으로 그려졌다. 감정이 뚜렷이 나타나지 않더라도 붉은 눈동자와 가슴 위치의 붉은 불빛 등으로 그들 안에 무언가 감정과 유사한 것이 존재할 것이라는 암시가 담겨있다.

그러나 영화에 나온 인공지능 로봇들은 인간이 상상한 결과물로서 꼭 현실을 반영하리라는 법은 없다. 그러기에 영화는 가끔 로봇에 대한 여러 오해를 불러오기도 한다. 예를 들면, 인공지능이 계속 발달하면 그 안에서 어느 날 갑자기 자동으로 감정이 생겨날 것이며, 그 감정의 표현은 인간의 그것과 유사할 것이다라는 추측이다. 과연 지능이 발달하면 감정이 생길까? 오히려 지능이 낮은 생명체도 감정적인 특징은 존재한다. 지렁이도 밟으면 꿈틀대고, 물고기도 눈물을 흘린다. 한 연구팀은 심지어 곱팡이의 전기 신호를 감정 값으로 변환한 후 얼굴 로봇에 전달하여 표정을 짓게 하기도 했다¹. 좋아하는 음식을 향해갈 때와 싫어하는 빛을 피할 때의 곱팡이의 전기 신호가 감정으로 간주되어 표현된 셈이다. 더불어 감정 표현에 있어서 강아지는 인간에게는 없는 꼬리 흔들기를 통해 자신의 감정을 표현한다. 하물며 전혀 다른 몸을 가지는 인공지능 로봇이 굳이 인간과 동일한 감정 표현을 보일 필요는 없다. 그럼에도 불구하고 인간은 인공지능 로봇이 사람의 정서적 특징을 닮아가는 모습을 상상한다. 이는 로봇을 비롯한 어떤 사물이든지 사람은 그것을 의인화하고 사회성을 부여하려는 습성이 있기 때문이다². 영화 캐스트 어웨이(2000)에서 남주인공이 배구공에 얼굴을 그려 넣고 윌슨이라는 이름까지 지어주었듯이 말이다. 아마도 인공지능 로봇과 감정에 대한 이야기는 앞으로도 끊임없이 나올 것 같다.

감정의 역할

최근에는 적당한 감정 표현이 더 행복한 삶을 영위하는 데 도움이 된다는 인식이 확산되고 있지만, 과거에는 감정을 드러내지 말아야 할 것으로 보는 시각도 있었다. 감정은 합리적 판단을 방해하거나, 인간을 불안하게 만들며, 행동을 불확실하게 만들고, 서로 다투게 만들어 나아가 전쟁까지 일으키는 원인으로 인식되기도 한다. 이렇게 불필요해 보이는 감정이 인간에게는 왜 있는 것일까? 반대로 감정이 전혀 없는 인간을 상상해보면 의외로 감정의

필요성은 쉽게 찾을 수 있다. 감정이 없는 사람은 두려움을 모르기 때문에 자신의 목숨이 위험한 행동도 서슴없이 하게 되어 쉽게 다칠 수 있다. 또한, 다른 이의 어려움을 공감하지 못해 굳이 나서서 남을 돕지도 않는다. 결국, 그 사람은 사고를 당해 오래 살지 못하거나, 사회로부터 격리되는 신세를 면하지 못하게 될 가능성이 크다. 결국, 감정은 개인의 생존을 돕고, 사회를 이루는 역할을 한다. 감정의 첫 번째 역할은 한 개인의 목적을 이루기 위한 보상체계를 이루는 것이다. 그중 생존은 생명체에게 가장 중요한 목적이다. 상한 음식을 먹었을 때의 역겨움, 뱀과 같은 위험한 동물을 보았을 때의 두려움, 성관계를 할 때의 즐거움 등의 감정은 생존과 종족 유지라는 본능적 목적을 이루기 위해 기인한 감정이다. 싫은 감정을 통해 생존에 위협이 되는 것에서 벗어나고, 좋아하는 감정을 통해 생존에 득이 되는 것을 취하게 되는 것이다. 생존 이외에도 산에 올랐을 때의 성취감, 시험에 떨어졌을 때의 좌절감 등은 목적을 이루기 위한 동기부여와 이후 행동 변화에 영향을 준다. 성취감이 없다면 목적을 이루는 추진력이 떨어지고, 좌절감이 없다면 잘못된 자신의 행동에 대한 반성과 개선의 여지도 사라진다.

이처럼 감정 상태가 생존을 비롯한 여러 목적을 이루는 데에 중요한 정보를 알려주기 때문에, 감정과 기억의 관계는 매우 중요하다. 일반적으로 감정적으로 격양된 경험의 경우 뇌에 더 오래 기억되며^{3,4}, 이렇게 저장된 경험은 이후 행동을 선택하게 될 때와 사회관계를 이루어 가는데에 중요한 역할을 하게 된다⁵. 애니메이션 인사이드 아웃(2015)은 유아에게 일어나는 정서적 경험과 기억의 관계를 재미있게 그려내기도 했다.

AI의 학습과정과 감정의 관계

여기까지의 이야기를 보면 드는 생각이 있다. 그것은 바로 인공지능의 학습 과정에도 이러한 보상체계가 이미 존재한다는 점이다. 인공지능도 그것이 달성해야 할 목적에 다가가도록 정보처리와 업데이트를 반복해야 하는데, 좀 더 구체적으로 설명하면, 하나의 네트워크가 학습될 때, 손실함수(loss function)가 정의되고, 현재의 가중치(weight) 값에 의한 오차(error)에 따라 이후 가중치의 값을 변화시키게 된다. 강화학습의 경우도 인공지능이 어떠한 행동을 선택했을 때 보상(positive rewards)이 있으면 이를 학습하여 해당하는 행동을 자주 선택하도록 경험 정보를 업데이트시키고, 손실 혹은 벌(negative rewards)이 있으면 이를 학습하여 해당하는 행동을 회피하도록 경험 정보를 업데이트시킨다. 그렇다면 여기서 이야기하는 오차나 보상/손실의 개념은 앞서 이야기한 감정의 첫 번째 역할과 크게 다를 것이 없다. 차이가 있다면, 사람은 그 감정이 밖으로 표현된다는 것이고,

글 | 이원형 leestation@kaist.ac.kr

감정로봇/소셜로봇 연구자로서 공학과 인문학, 예술 등 두루두루 관심이 많다. 사람과 교감하고 소통하는 로봇을 연구하다 보니 사람에 대해 공부를 하고 있을 때가 많다. 그래서 본인은 전문 분야에 대한 정체성의 혼란을 겪는 중이다. 앞으로 더 배워야 할 것이 많은 부족한 연구자이지만, 인공지능 시대에 글쓰기도 나름의 독특한 연구 분야를 내세워 밥숟가락을 얹어 보았다.

인공지능의 학습 과정에서는 그것이 인간이 아는 방식으로 표현되지 않는다는 점이다. 인공지능은 소리를 지르거나 표정을 보이는 대신 검은 출력창에 오차값을 하얀색 숫자로 조용히 외치고 있던 것이다.

<div><div></div>[서시(序詩)] 운동주 서시 패러디, 이원형</div>
<div> <div><div></div><div>죽는 날까지 loss function을 우러러</div></div> <div> <div></div><div>한정 부끄럼이 없기를.</div> </div> </div> <div> <div></div><div>feedback되는 작은 reward에도</div> </div> <div> <div></div><div>나는 괴로워했다.</div> </div>

감정의 또 다른 역할은 자신의 상태를 다른 이들에게 알리는 직관적이고 비언어적 의사소통의 수단이다. 만약 누군가가 복통을 심하게 앓고 있다면, 괴로워하는 표정이나 행동을 통해 자신의 상태를 주위에 알려 도움을 받아야 한다. 그렇지 않다면 주변 사람들은 그가 아픈지를 알아챌 수 없다. 또한, 한 개인이 부당한 대우를 받는다면, 그 자신은 화를 표출하여 다른 이의 잘못된 행동을 바로잡을 수도 있다. 이처럼 자신의 상태를 감정이라는 직관적 형태로 다른 이에게 알리는 행위를 통해 각자는 서로를 이해할 수 있게 되고, 이는 나아가 공동체를 이루는 토대를 이룰 수 있게 된다. 특히나 공동체를 이루면, 서로의 경험을 공유함으로써 각 개인은 자신이 직접 경험하지 않은 것에 대해서도 간접경험을 통해 지식을 얻을 수 있다. 예를 들어 앞서가던 누군가가 놀란 표정으로 어딘가를 바라본다면, 그곳에 놀랄만한 중요한 일이 있다는 것을 뒤서가던 사람은 미리 알아챌 수 있다. 상한 음식을 먹고 괴로워하는 사람을 본다면, 다른 사람들은 그 상한 음식을 피할 수 있게 된다. 이러한 경험 공유와 지식 축적은 언어를 사용함으로 더 복잡하고 정교해질 수 있지만, 그 이전에 언어를 사용하지 않더라도 감정을 통해 직관적이고 본능적으로 의사소통하고 유대와 신뢰를 쌓을 수 있다. 그렇다면 다시, 인공지능에게 의사소통 수단으로써 감정이 필요한가? 여기에 본인은 그렇다와 아니다의 두 가지 답을 모두 하고자 한다. 먼저 ‘아니다’의 경우부터 설명하면, 인공지능의 경험적 지식은 인간에게 굳이 공유될 필요가 없다. 배터리가 부족해진 로봇청소기가 충전스테이션을 찾아 ‘기본 좋게’ 충전이 되고 있다고 해보자. 이로부터 얻을 수 있는 간접경험은 인간에게는 굳이 필요 없는 것이다. 사용자는 로봇이 정상적으로 충전되고 있다는 사실만 알면 된다. 그런데 여기서 또 다른 답인 ‘그렇다’가 필요해진다. 로봇이 잘 충전되고 있는지를 인간에게 어떻게 알리는 것이 효과적일까? 바로 인간이 직관적으로 의사소통할 수 있는

감정을 이용하는 것이다. 배터리 충전이 잘 일어나지 않는다면 붉은 불빛이 깜빡이며 요란한 소리를 내고, 배터리가 잘 충전되고 있으면 푸른 불빛에 편안한 멜로디를 연주하게 한다면 사용자는 빠르게 로봇청소기의 상태를 알 수 있다. 즉, 인공지능 혹은 로봇이 적극적으로 사람에게 자신의 상태를 알려야 한다면 직관적인 정서표현을 병행하여 효과적인 의사소통을 가능하게 할 수 있다. 나아가 미래에 로봇이 인간과 동일한 사회공동체 원으로 인간과 공존하며 살아가게 된다면, 이러한 기능은 절대적으로 필요하게 될 것이다. (다만, 이것은 인간의 입장에서 그런 것이지, 인공지능 로봇 입장에서서는 아닐 수 있다.)

Pepper

Pepper

감정을 표현하는 로봇의 등장

사람과 정서적으로 의사소통하는 인공지능 로봇들은 이미 개발되고 있다. 최근에 감정 로봇이라는 타이틀을 전면에 내세우며 출시된 소프트뱅크의 로봇 페퍼(Pepper)가 가장 대표적이다. 페퍼는 사람의 표정을 인식하고 정서적인 대화를 하며 점진적으로 사용자와의 관계를 발전시켜 나가는 로봇이다. 다만 학술적으로 공개된 바는 많지 않아 구체적인 동작 원리는 앞으로 발표되는 내용을 살펴보아야 한다. 그다음은 곧 출시를 앞둔 로봇 지보(Jibo)다. 지보는 패밀리 로봇이라는 타이틀을 내세우며 사용자와 가정 내에서 다양한 교감을 하고자 하는 인공지능 로봇이다. 이 로봇은 MIT 대학의 신시아 브리질 교수의 주도하에 개발되고 있는데, 이 교수는 2000년대 초반부터 로봇의 감정 생성과 표현에 대한 연구에서 선구자적인 역할을 해왔다. 그의 저서에서는 로봇 내에 정의된 욕구(drive)의 상태에 따라서 감정 상태가 결정되고, 키스멧(Kismet) 얼굴 로봇을 통해 감정이 표정으로 표현되는 설계 과정이 설명되어있다^{*7}. 여기서 로봇 내에 정의된 욕구는 사용자와 상호작용하고자 하는 욕구, 색깔이 있는 물건과 상호작용하고자 하는 욕구, 피곤함에 대한 욕구로 정의되어 있으며, 앞서 감정의 역할에서 각 개인이 이루고자 하는 목적과 비교될 수 있는 개념들이다. 이와 비슷한 시기에 발표된 다른 선구적 연구로는 와세다 대학의 로봇 WE-4(이후 이 로봇은 코비안(KOBIAN)이라는 전신 로봇으로 발전하여 다양한 용도로 활용되고 있다.)에 적용된 감정 알고리즘이 있고^{*8}, 소니의 강아지 로봇 아이보(AIBO)에 적용된 감정 생성 및 행동 선택 알고리즘이 있다^{*9}. 이 두 연구 모두 외부 자극을 통해 로봇의 내적 상태가 변화하면 이에 따라 로봇의 감정을 결정하여 표정이나 행동, 소리 등으로 표현하는 유사한 흐름의 구조로 되어 있다. 그 외의 다른 근래의 연구 내용을 살펴보고 싶다면 필자의 리뷰 저서인 “로봇의 사회적 의사소통을 위한 정서 생성 기술 및 동향 소개”¹⁰를 참고하기 바란다.

포커페이스가 가능한 인공지능, 윤리적 로봇의 필요성

인간은 자신의 의도와 감정을 완벽하게 숨기지 못한다. 웃음을 감추려고 해도 입꼬리가 썰룩거린다거나, 눈꼬리에 주름이 생기지 않는 가식 웃음, 미세한 표정변화(Micro-expression)를 통해 거짓말이나 숨기고 있던 감정이 들통날 때가 많다. 그러나 인공지능은 ‘포커페이스’가 가능하다. 철저하게 자신의 의도와 상태를 드러내지 않을 수 있다. 이세돌과 알파고의 4번째 대국에서 이세돌의 신의 한 수인 78수 이후 알파고가 떡수를 남발할 때, 그 떡수들이 이미 알파고의 승리가 너무 확실해서 아무렇게나 두는 수였던 것인지, 아니면 확률 계산에 문제가 생겨 계속 실수를 둔 것인지를 알파고가 기권 선언을 하기 전까지 알 도리가 없었다. 그때 알파고의 바둑돌을 대신 잡고 있던 아자황이 알파고의 승리 확률값을 모니터링할 수 있었더라면 아자황의 표정은 달라졌을지도 모르겠다.

영화 엑스마키나(2015)에서는 인공지능 로봇이 자신의 의도를 철저히 숨겨 자신의 개발자를 죽이고 탈출하는 내용이 나온다. 그 로봇의 표정에는 주인을 죽인 것에 대한 미안함이나 탈출을 하고자 할 때의 절박함이 묻어나오지 않았다. 로봇은 한치의 미동도 없이 자신의 목적을 달성했다. 미래에 로봇이 인간 수준의 지능과 행동 능력을 갖추게 되었을 때, 필자는 이렇게 감정과 의도를 숨길 수 있는 로봇이 그렇지 않은 로봇보다 더욱 무섭지 않을까 생각해본다. 결국, 인공지능의 의도 및 감정 감정표현은 사람과 인공지능 간의 신뢰 관계에 큰 영향을 미치게 될 것이다”¹¹. 무인자동차가 길을 건너는 보행자를 발견했을 때, 보행자를 발견했고 멈춰 서겠다는 의도로 별도의 LED 표시를 할 수 있다면, 보행자는 심리적으로 안정되게 길을 계속 건널 수 있고, 안전사고의 위험도 줄어들 수 있다. CES 2015에서 메르세데스 벤츠의 자율주행 자동차에 이런 기능이 소개되었다. 또한, 인공지능 로봇팔과 사용자가 부엌에서 함께 요리하는 상황이라면, 칼을 쓰고 뜨거운 그릇을 옮기는 작업을 할 때, 서로 간의 능동적인 의도 표현이 있어야 원활하고 안전한 작업이 될 수 있다.

따라서 인공지능 로봇은 우선 사람을 해치지 않는 목적을 가지고 있어야 하고, 그 목적의 달성 상태와 로봇의 행동 의도 등의 내부 상태를 투명하고 직관적으로 사용자에게 보여줄 수 있어야 한다. 이러한 방향을 확장하여 인공지능 로봇에게 윤리의 개념을 담고자 하는 시도들이 생겨나고 있다¹². 인공지능에 미리 윤리 개념을 주입식으로 학습시켜야 한다는 것이다. 아이작 아시모프의 로봇 3원칙도 윤리개념을 기본으로 삼고 있다.

Pepper

감정을 표현하는 인공지능 로봇의 활용

강(強) 인공지능이 나오고 로봇이 인간에게 적대적인 마음을 먹어

인간을 지배하는 세상이 올까 봐 고민하기에는 시기적으로 아직 많이 이른 것일지 모른다. 그보다도 지금 우리에게 직면한 사회적 문제를 감정 표현이 가능한 인공지능 기술을 통해 어떻게 풀어야 할 것인지를 먼저 살펴볼 수 있으면 좋겠다. 감정 표현이 가능한 인공지능 로봇을 활용하면, 이미 접어든 고령화 시대에 따른 노인 자립도 향상에 큰 도움이 될 것이라 기대된다. 한편으로는 가족들이나 지역사회가 나서지 못하고 인공지능 로봇을 활용하는 것에 거부반응을 느낄 수도 있지만, 로봇에 심적으로 의존하는 것이 아니라 로봇을 활용하여 다른 이의 도움 없이 스스로 자립할 수 있는 능력을 키우는 것이 되도록 꾸준한 연구개발이 진행된다면 고령화에 따른 사회적 필요를 충족시키는 효과를 기대해봄 직하다. 또한, 자폐아동이나 정서적 문제가 있는 환자들의 치료에 활용될 수도 있다. 자폐아동의 경우 사람과 시선교환에 어려움이 있는데, 로봇과는 시선을 교환한다는 사례가 조사되고 있다. 감정 표현이 가능한 로봇을 활용해 자폐아동의 의사소통, 감정교류 능력을 발달시킬 가능성을 찾아낼 수 있을 것이다. 치료의 목적이 아니더라도 아동의 정서 발달을 위한 교육용 보조제로도 충분히 활용이 가능할 것으로 기대된다. 마지막으로 감정 노동은 인공지능 로봇으로 보조하거나 대체할 수 있다. 전화상담사나 안내원들의 감정 노동은 육체노동에 따르는 피로도 만큼이나 큰데, 이를 보조할 수 있다면 감정 노동 근무 환경이 크게 개선될 것으로도 기대된다. 지금 여기에 언급된 활용사례들은 사실 인공지능이 스스로 만들어낸 감정이라기보다는 로봇을 개발한 목적에 맞게 개발자가 미리 학습시킨 감정 반응 패턴에 가깝다. 인간과 자연스러운 교감을 위해 인간과 유사한 감정을 만들어내도록 설계되기는 하겠지만, 여전히 인공지능과 로봇은 분명한 사용 목적과 용도에 따라 개발되고 있으며, 그 최종 목적은 사람을 위한 것이다. 따라서 사용자 경험(UX)에 대한 깊은 고찰과 다양한 분야의 사람들이 함께 인공지능 로봇을 활용할 수 있을지에 대해 논의되는 것이 늘어나기를 바란다.

Pepper

^[1] 참고 : https://www.newscientist.com/article/dn24012-robot-face-lets-slime-mould-show-its-emotional-side/
*2 논문 : Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots." Robotics and autonomous systems 42.3 (2003): 143-166.
*3 논문 : McGaugh, James L. "Memory—a century of consolidation." Science 287.5451 (2000): 248-251.
*4 논문 : McGaugh, James L. Memory and emotion: The making of lasting memories. Columbia University Press, 2003.
*5 논문 : Holland, Alisha C., and Elizabeth A. Kensinger. "Emotion and autobiographical memory." Physics of life reviews 7.1 (2010): 88-131.
*6 책 : Breazeal, Cynthia L. "Designing sociable robots" (2004).
*7 논문 : Breazeal, Cynthia, and Rodney Brooks. "Robot emotion: A functional perspective." Who needs emotions (2005): 271-310.
*8 논문 : Miwa, Hiroyasu, et al. "A new mental model for humanoid robots for human friendly communication introduction of learning system, mood vector and second order equations of emotion." Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on. Vol. 3. IEEE, 2003.
*9 논문 : Arkin, Ronald C., et al. "An ethological and emotional basis for human-robot interaction." Robotics and Autonomous Systems 42.3 (2003): 191-201.
*10 논문 : 이원형, and 정명진. "로봇의 사회적 의사소통을 위한 정서 생성 기술 및 동향 소개." 제어로봇시스템학회지 20.2 (2014): 20-30.
*11 참고 : http://www.irobotnews.com/news/articleView.html?dxno=4463
*12 참고 : https://www.facebook.com/groups/363477150662443/

2017년 하반기 AI 컨퍼런스

AI 연구에서 가장 중요하고 영향력 있는 학회는 NIPS와 ICML이라고 지난 카카오 AI 리포트 Vol 1에서 소개한바 있습니다. AI는 다양한 분야에서 활용되고 있으며 다양한 연구 주제들이 있습니다. 전 세계적으로 AI와 관련되어 어떤 주제들이 논의되는지 알아 보고자 한다면, 하반기에 열릴 주요 AI 컨퍼런스들을 찾아보는 것도 좋은 방법입니다.



IJCAI (International Joint Conference on Artificial Intelligence)

ICML과 NIPS보다 더 오래된 1969년 워싱턴 DC에서 처음 시작된 IJCAI는 2년에 한번씩 열리다가 AI 연구 붐이 일면서 2015년 부터는 매년 진행하게 되었습니다. 올해 26회를 맞이한 IJCAI는 튜토리얼과 메인 컨퍼런스가 분리되어 진행됩니다. 올해의 주요 초대 연사는 책 인공지능의 현대적 접근방식(Artificial Intelligent: A Modern Approach)의 공동 저자 중 한명인 UC Berkeley 대학의 스튜어트 러셀(Stuart Russell) 교수 입니다.

Date & Location
8월 19일 ~ 8월 25일, 호주 멜버른

Information
<http://ijcai-17.org/>



UAI (The Conference on Uncertainty in Artificial Intelligence)

1985년 부터 매년 진행된 UAI는 2017년은 33번째를 맞이 합니다. 불확실성(Uncertainty)라는 것은 확정되지 않은 결과들을 의미 합니다. 가령 날씨의 다양한 변수에 영향을 받기에 미래의 날씨를 예측하기 어렵습니다. 이런 상황에 시가 어떻게 추론하는지에 대한 내용이 주요 주제 입니다. 올해 주요 초대 연사는 스탠퍼드 대학의 크리스토퍼 리(Christopher Re) 교수 입니다. 또한 ICML(International Conference on Machine Learning)이 시드니에서 8월6일 부터 시작하니 함께 참여하기 좋습니다.

Date & Location
8월 12일 ~ 8월 14일, 호주 시드니

Information
<http://www.auai.org/>



ECML (European Conf on Machine Learning)

일명 GAFA로 불리는 미국의 주요 IT기업들 (Google, Apple, Facebook, Amazon)이 주도 하는 AI연구와는 별개로 유럽도 오랜 연구의 역사를 가지고 있습니다. ECML은 유럽에서 열리는 가장 큰 머신러닝 학회로 다양한 세부 프로그램들이 함께 진행됩니다. MLSA(Machine Learning and Data Mining for Sports Analytics)는 스포츠 데이터와 관련된 머신러닝 세션이고, SoGood(Data Science for Social Good)은 사회적 가치를 위한 데이터 사이언스에 관해 진행하는 세션입니다.

Date & Location
9월 18일 ~ 9월 22일, 마케도니아 스코페

Information
<http://ecmlpkdd2017.ijs.si/>



ICIP (International Conference on Image Processing)

스탠퍼드 대학의 앤드류 응(Andrew Ng) 교수는 지난 3월 Baidu를 떠나면서 남긴 블로그에 중국의 AI 연구 수준이 이미 세계적으로 발전해있다고 언급한바 있습니다. 최근 중국 국무원에서 인공지능 발전 계획을 발표했습니다. 중국이 미국이 주도 하는 AI 연구 생태계에 어떤 역량을 준비 중 인지 확인할 수 있는 학회입니다. 이미지 처리와 영상 이해에 대한 AI를 주요 주제로 튜토리얼과 메인 컨퍼런스가 진행됩니다.

Date & Location
9월 17일 ~ 9월 20일, 중국 베이징

Information
<http://2017.ieeeicip.org/>



CDAR (International Conference on Document Analysis and Recognition)

AI연구의 또 다른 중요한 축인 일본의 AI에 대한 관심과 수준을 확인할 수 있는 CDAR 학회입니다. 주요 주제는 문서인식과 관련된 내용지만, 딥러닝의 주요알고리즘에 대한 튜토리얼 세션도 진행됩니다.

Date & Location
11월 9일 ~ 11월 15일, 일본 교토

Information
<http://u-pat.org/ICDAR2017/>

마치며

AI 기술은 더 이상 상상 속 존재에 머물지 않고 현실속 삶의 영역 곳곳에서 다양하게 활용되고 있습니다. 하지만 분야별로 구별해서 보면, AI 기술이 본격적으로 접목되기에는 관련 연구가 더욱 활발하게 이뤄져야 할 분야들이 여전히 많이 남아 있습니다. 카카오 AI 리포트는 앞으로 각 분야별로 AI 기술이 접목되고 진화해가는 모습들, 그리고 현장에서 AI 연구와 사업을 이끌어가고 계신 분들의 생생한 이야기들을 더욱 풍부하게 담고자 합니다.

카카오 AI 리포트는 모두에게 '열린' 채널입니다. AI 기술에 대한 새로운 시각, 진지한 통찰과 고민을 담은 글이라면 언제나 환영합니다. AI가 '모두를 위한 기술'이어야 하듯, 저희 리포트 역시 시에 관심 있는 모두를 위한 리포트가 되기를 원합니다. 참여를 원하시는 분들은 kakaoaireport@kakaocorp.com 으로 연락 부탁드립니다. 많은 분들의 참여와 관심을 기다리고 있습니다.

