

스마트폰 사용자의 감정분석 알고리즘 연구

백주연*, 신혜승**, 원은지*, 윤예슬**

*숙명여자대학교 IT공학전공

**숙명여자대학교 컴퓨터과학전공

19j@sookmyung.ac.kr, hyeseung2000@gmail.com, ji8460@sookmyung.ac.kr,
yoonyesle@sookmyung.ac.kr

A Study on the Emotional Analysis Algorithm of Smartphone Users

Ju-Yeon Baeck*, Hye-Seung Shin**, Eun-Ji Won*, Ye-Seul Yoon**

*Dept. of Information Technology Engineering, Sookmyung Women's University

**Dept. of Computer Science, Sookmyung Women's University

요 약

현대 사회의 스트레스 문제가 심각해짐에 따라 각종 스트레스 관리 서비스가 꾸준히 개발되고 있으나, 해당 서비스들은 정서 상태 판단을 사용자가 직접 입력하는 데이터에만 의존하기 때문에 분석 결과를 완벽히 신뢰하기 어렵다.

본 연구에서 개발한 앱 S-detector는 스마트폰 사용 시간 및 빈도 정보를 자동으로 수집하고, 사용자가 작성한 일기 데이터에서는 감정 단어를 추출하여 스마트폰 사용 데이터와 일기 데이터를 각각 분석, 종합적으로 판단하는 알고리즘을 가지고 있다. 따라서 사용자가 심리·정신적 문제 가능성을 쉽게 인지하는 데 도움을 주는 앱으로서 해당 문제를 예방하거나 조기에 해결함을 목표로 한다.

1. 서론

스트레스란 개인에게 부담을 주는 정신적, 육체적 자극과 그에 대한 반응을 의미한다. OECD 통계에 따르면 한국의 우울증 유병률은 인구의 36.8%인 1,000만 명(2020년 기준)이나 실제 우울증으로 진료받은 환자는 101만 6,000여 명에 불과하다(2020년, 국민건강보험공단). 즉, 과도한 스트레스에 노출되어도 자신의 스트레스 정도를 인지하지 못하거나 인정하지 않아 전문 의료 기관에 방문하지 않는 상황이 발생하는 것이다. 이는 치료 시기를 놓쳐 정신 질환으로 이어지는 결과를 초래할 수 있다.

이러한 분위기 속에서 각종 스트레스 관리 서비스가 꾸준히 개발되고 있으나, 기존의 서비스들은 정서 상태 판단을 사용자 스스로 직접 입력하는 데이터에만 의존하기 때문에 분석 결과의 신뢰도가 높다고 볼 수 없다.

본 연구는 정신건강 관리에 실질적인 도움을 주기 위해, 서비스 사용자의 정서 상태 분석 정확도를 높이는 것에 집중하였다. 스마트폰을 이용해 스마트폰 사용 시간, 빈도 정보 및 작성된 일기의 숨은 감정을 자동 수집, 분석하는 알고리즘을 개발하여 앱 S-detector에 적용하였다.

기존의 정서 관리 서비스와의 차이점은 크게 두 가지로 볼 수 있다.

첫째, 사용자의 조작 없이 스마트폰 앱 사용 시간 및 빈도를 수집한다는 점이다. 유사 생활 패턴의 다른 사용자 및 평균적인 데이터와 비교하여 정서 판

단의 기준으로 활용한다.

둘째, 사용자가 앱에서 기록한 일기 속에서 숨은 감정을 찾아내 추출한다는 점이다. 추출한 감정 단어의 빈도를 분석해 일기 작성 당시 사용자의 실제 감정을 판단한다.

위의 과정을 통해 얻은 스마트폰 사용 분석 데이터와 일기 분석 데이터를 종합적으로 분석하여 사용자의 정서 상태 및 지각하지 못했던 스트레스를 정확하게 찾아내는 것이 알고리즘의 핵심이다. 또, 사용자 정서 분석 내용을 바탕으로 스트레스를 해소할 수 있는 활동을 추천해줌으로써 보다 적극적으로 사용자의 심리적 안정을 돕는다.

S-detector의 활용 가능성을 알아보기와 전문가와 인터뷰를 진행했다. 숙명여자대학교 교육학부 상담심리전공 교수는 실제 상담 및 심리치료, 약물치료 이전에 본인의 상태를 자기 모니터링을 통해 비교적 정확하게 확인할 수 있다는 점을 긍정적으로 판단했다. 특히 스마트폰을 많이 사용하는 경우, 스스로 관리하는 측면에서 심리 상담 외 보조적인 도움이 될 수 있다는 평가를 통해 이번 연구에서 개발한 알고리즘과 앱의 가치를 확인할 수 있었다.

2. 감정서비스 구성도 및 주요 기능

2.1 연구 배경

현대인의 스트레스 인지 관련 실태를 파악, 인식을 조사하고자 설문조사(2021.08.20. ~ 2021.08.27. 162명

응답)를 실시하였다. 통계를 위해 참여자의 성별, 연령, 직업 정보를 수집, 스마트폰 사용 유형 관련 자체 문항과 스트레스 자가 진단 척도 문항(BEPSI-K)을 활용하여 설문 내용을 구성하였다.

통계·분석 결과, 일 평균 스마트폰 사용 시간은 '3시간 이상 6시간 미만'이 42.5%로, 자주 사용하는 스마트폰 앱의 평균 사용 시간은 '1시간 이상 3시간 미만'이 46.3%로 가장 많았다.

스트레스 해소 방법 문항은 '운동, 여행 등 활발한 활동'이 전체 24.7%로 가장 많았으나, 연령대별 분석 결과에서는 30대는 음식 섭취, 40대는 수면이 가장 많이 선택되었다.

또한 본 연구에서는 본인의 정서 상태의 인지를 스트레스 관리에서 중요한 요인으로 보았기 때문에, 질병관리청에서 실시한 스트레스 인지율(일상 중 스트레스를 대단히 많이/많이 느낀다고 답변한 비율)에 관한 설문 조사를 연구 지표로 활용하였다. 질병관리청에서 제공한 통계에 따르면, 한국의 만 19세 이상 인구의 전체 평균 스트레스 인지율은 30.8%이고, 남자 30대(38.8%), 여자 20대(42.3%)에서 가장 높았다. [3]

설문 결과는 앱 개선 및 개발 방향을 설정하는 데 활용되었다. 특히, 스트레스 인지율과 관련한 설문 결과 중 성별, 연령별, 소득 수준별 차이를 바탕으로 비슷한 집단을 구성해 데이터를 분석하여, 보다 정확한 기준을 성립하였다.

(1) 스마트폰 사용 시간 판단 기준

설문조사 결과에 따라 일평균 스마트폰 사용 시간을 6시간으로 설정하였다. 따라서 주평균 스마트폰 사용 시간은 42시간(±10시간)이 안정 범위이다.

수집한 스마트폰 사용 시간 로그가 안정 범위 이상으로 판단되면 마음 정리 보고서에서 경고를, 범위 이하인 경우 건강한 상태임을 알려준다.

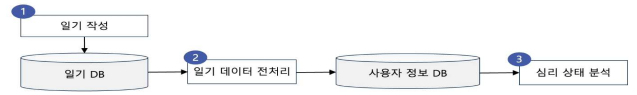
(2) 스트레스 해소 방법 제안

스트레스 해소 방법 문항에서는 '운동, 여행 등 활발한 활동', '수면', '책, 영화 등 정적인 활동'이 순서대로 높은 응답률을 보였다.

사용자가 과도한 스트레스를 받았다고 평가되면 해당 주의 마음 정리 보고서에서 위의 응답 결과에 기반하여 스트레스 해소 방법을 제안한다.

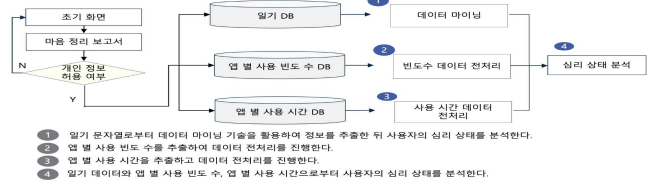
2.2 서비스 구성도

앱 S-detector는 사용자의 앱 사용 시간 및 빈도수 로그 정보를 수집하여 가공한 것과 사용자가 작성한 감정 일기 데이터를 수집하여 텍스트 마이닝을 통한 분석의 과정을 DB에서 관리한다. 이에 대한 데이터들과 심리 진단 지표 정보를 바탕으로 해당 주 사용자의 진단 결과를 생성한 후, 그 결과를 앱 화면에 나타낸다.



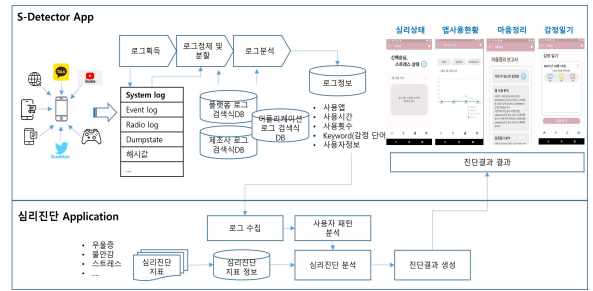
- 1 사용자 일기를 작성하고 저장하면 일기 DB에 데이터가 저장된다.
- 2 일기 데이터 전처리를 진행한다.
- 3 일기 데이터로부터 사용자의 심리 상태를 분석한 뒤 DB에 저장한다.

[그림 1] 일기 분석 기능 흐름도



- 1 일기 문자열로부터 데이터 마이닝 기술을 활용하여 정보를 추출한 뒤 사용자의 심리 상태를 분석한다.
- 2 앱 사용 빈도 수를 추출하여 데이터 전처리를 진행한다.
- 3 앱 별 사용 시간을 추출하고 데이터 전처리를 진행한다.
- 4 일기 데이터와 앱 별 사용 빈도 수, 앱 별 사용 시간으로부터 사용자의 심리 상태를 분석한다.

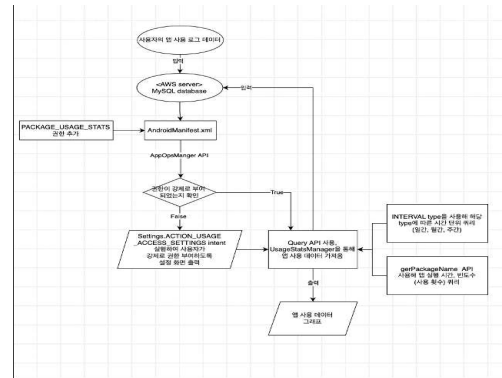
[그림 2] 심리 상태 분석 기능 흐름도



[그림 3] 서비스 흐름도

2.3 주요 기능 및 특징

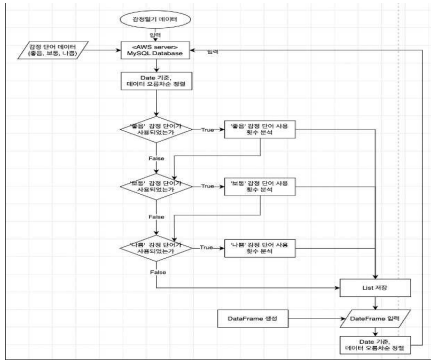
(1) 스마트폰 앱 사용 데이터 자동 수집 및 분석
회원가입 시 사용자 접근권한 동의 후에는, 별도의 조작을 하지 않아도 사용자의 스마트폰 앱 사용 시간 및 빈도 데이터를 자동 수집한다. 수집된 데이터를 일반적인 평균 데이터 또는 유사 생활 패턴 그룹의 다른 사용자 지표와 비교·분석해 심리 분석 결과를 제공한다.



[그림 4] 앱 사용 분석 기능 흐름도

(2) 감정 일기 데이터 분석

앱 내의 감정 일기 기능은 사용자가 작성한 일기에서 감정 단어를 추출하여 숨은 감정을 알아낸다. 사용자의 기록을 통해 객관적으로 심리상태를 판단한다는 점에서, 사용자의 주관적 판단에 의존하던 기존 정서 관리 서비스(앱)과 큰 차별성을 갖는다.



[그림 5] 일기 분석 알고리즘 흐름도

(3) 사용자 심리상태 분석 및 결과 제공

자체 알고리즘을 사용해 스마트폰 앱 사용 데이터 및 감정 일기 데이터의 분석 결과로부터 사용자의 심리상태를 종합적으로 분석한다. 분석 결과는 주 단위로 사용자에게 제공된다.

(4) 데이터 시각화

수집한 앱 사용 데이터를 MPAndroidChart 오픈 소스 라이브러리를 사용해 안드로이드 앱 내 그래프로 시각화하여 전달의 효율을 높였다. [1]

일주일 단위의 사용자 앱 사용 시간과 빈도수 데이터를 그래프로 나타내기 위해 HorizibtalBarChart를 사용하였다. 앱 사용 시간 데이터를 자세하게 제공하기 위해 지난 4일간, 4주간, 4개월간 데이터는 LineChart를 활용하였다.

3. 감정분석 알고리즘 개발

3.1 스마트폰 사용 데이터 수집 및 분석

(1) 앱 사용 로그 분석

UsageStatsManager는 장치 사용내역 및 통계에 대한 액세스를 제공하는 클래스이다.

UsageStatsManager의 Query API인 queryUsageStats는 인자로 전달한 범위 동안을 Interval type 단위로 묶어서 응용 프로그램 사용량 통계를 제공한다.

```

@RequiresApi(api = Build.VERSION_CODES.LOLLIPOP)
private String[] get_apps_name(int term) {
    if (!checkPermission())
        startActivity(new Intent(Settings.ACTION_USAGE_ACCESS_SETTINGS));
    String[] ret = new String[0];
    String packageName = "Nothing";
    long timeInterval = 500;
    int minutes = 59, seconds = 59, hours = 59;
    UsageStatsManager usageStatsManager = (UsageStatsManager) getContext().getSystemService(USAGE_STATS_SERVICE);
    Calendar cal = Calendar.getInstance();
    cal.add(Calendar.DAY_OF_MONTH, term);
    long cur_time = System.currentTimeMillis(), begin_time = cal.getTimeInMillis();
    List<UsageStats> stats = usageStatsManager.queryUsageStats(UsageStatsManager.INTERVAL_BEST, begin_time, cur_time);
    }
    
```

[사진 1] 스마트폰 사용 로그 수집(queryUsageStats)

(2) 앱 사용 로그 분석 알고리즘

S-detector 내에서는 queryUsageStats의 Interval type은 DAILY로, 분석 기간은 매일 0시를 기준으로 7일전으로 설정하여 일 단위로 지난 7일 동안 사용한 ‘패키지 이름’과 ‘실행된 시간’, ‘실행된 횟수’을 List<UsageStats> 객체로 리턴하고, ‘실행된 시간’과 ‘실행된 횟수’ 상위 4개의 응용 프로그램을 출력하여 분석 서비스를 제공한다.

(3) 감정 일기 분석 알고리즘

일기 데이터 분석을 위해 Python Pandas 라이브러리를 사용하였다. ‘한국어 감정표현단어의 추출과 범주화(손선주 외, 2012)’의 단어 범주를 참고하여 감정 단어를 좋음, 보통, 나쁨으로 분류하였다. [2]

프로그램을 실행하면 서버에 등록된 사용자의 일기 데이터를 자동으로 불러와 감정 단어를 추출, 각 범주별 감정 단어의 빈도수를 수집한다. 알고리즘 실행 후, 일기 분석 데이터는 다시 서버에 저장되어 앱 내 마음정리 보고서(레포트 기능)에서 제공되며, 앱 사용 로그 분석 데이터와 함께 사용자의 심리상태를 종합적으로 판단하는 기준이 된다.

```

for i in range(1):
    #일기 데이터
    for list in ListArray:
        #일기 데이터
        dfFilterDate = df[df['DATE'] == dateArray[k]]
        totalCount_good = 0
        totalCount_bad = 0
        # keyword_good in keywords_good:
        dfFilterKey_good = dfFilterDate[dfFilterDate['content'].str.contains(keyword_good)]
        countKey_good = dfFilterKey_good['content'].count()
        totalCount_good = totalCount_good + countKey_good
        # keyword_bad in keywords_bad:
        dfFilterKey_bad = dfFilterDate[dfFilterDate['content'].str.contains(keyword_bad)]
        countKey_bad = dfFilterKey_bad['content'].count()
        totalCount_bad = totalCount_bad + countKey_bad
        data_list.append([ListArray[k][0], ListArray[k][1], totalCount_good, totalCount_normal, totalCount_bad])
        k += 1
    
```

[사진 2] 일기 데이터 분석(Pandas)

3.2 감정 매칭 알고리즘

감정 매칭 알고리즘은 한 주 동안의 앱 사용 시간과 감정 일기 분석 결과에 근거하여 해당 주의 사용자의 총 감정을 평가한다.

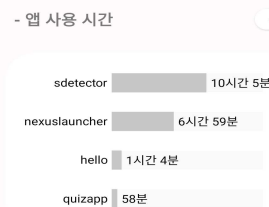
한 주 동안의 앱 사용 시간을 사용자의 평균과 비교하여 5% 이상 증가했을 때에는 감정 상태를 나쁨(-1)으로, 5% 이상 감소했을 때에는 좋음(1)으로, 그 외에는 보통(0)으로 책정하여 30%의 비중을 준다.

해당 주의 감정 일기에서 각 일기마다 좋음(1), 보통(0), 나쁨(-1)으로 감정을 추출하여 일별로 10%의 비중을 준다. 일기를 작성하지 않아 추출되지 않은 부분은 보통(0)으로 책정한다.

위와 같이 해당주의 감정을 모두 측정하여, 총 합산이 0.3 초과일 경우 좋음 감정을, -0.3미만일 경우 과도한 스트레스를 느낀 것으로 설정한다.

3.3 감정분석 평가

감정 매칭 알고리즘의 평가를 위해, 앱 사용 로그 분석 알고리즘과 감정 일기 분석 알고리즘의 정확도를 분석해보았다.



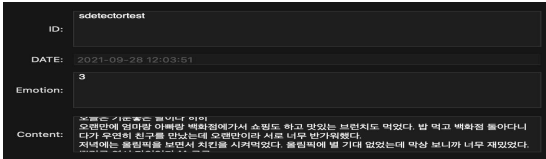
[사진 3] 앱 내 사용 시간

위의 [사진 3]은 사용자의 지난 7일 동안 상위 4개의 응용 프로그램 사용 시간을 S-detector 내에 그래프를 통해 확인한 것이다(9월 29일 기준, 7일간).

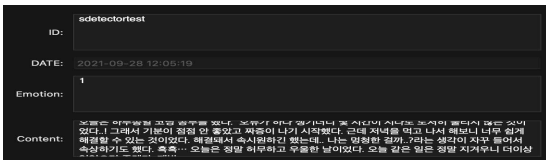


[사진 4] 설정 내 S-detector 앱 사용 사진

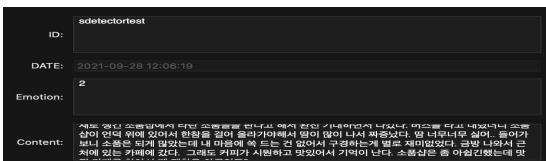
지난 7일 중 사용 시간이 가장 긴 S-detector 앱의 실제 사용 시간을 안드로이드 내 설정을 통해 9.22~9.28 기간 동안 확인하여 합친 결과, 이는 [사진 3]에서의 그래프에서와 같은 결과를 얻었다.



[사진 5] 사용자 작성 감정일기(감정 좋음 선택)



[사진 6] 사용자 작성 감정일기(감정 나쁨 선택)



[사진 7] 사용자 작성 감정일기(감정 보통 선택)

위의 세 개 화면은 사용자가 작성한 감정일기를 서버에서 확인한 것이다. 순서대로 일기 작성 시 감정을 좋음, 나쁨, 보통을 선택하였다. 분석을 위해 각각의 일기 데이터에는 좋음-3, 보통-2, 나쁨-1의 가중치를 부여하였다.

False	ID	DATE	num_good	num_normal	num_bad
7	sdetectorstest	2021-08-17 11:00:35	0	0	0
8	sdetectorstest	2021-08-17 11:32:19	0	0	0
9	sdetectorstest	2021-08-17 11:32:19	0	0	0
10	sdetectorstest	2021-08-17 12:43:12	0	0	0
11	sdetectorstest	2021-08-17 12:43:52	0	0	0
12	sdetectorstest	2021-08-17 12:54:29	0	0	0
13	sdetectorstest	2021-08-17 13:12:43	0	0	0
14	sdetectorstest	2021-08-19 05:44:43	0	0	0
15	sdetectorstest	2021-08-19 14:12:00	0	0	0
16	sdetectorstest	2021-09-21 11:56:56	0	0	0
17	sdetectorstest	2021-09-23 07:14:19	2	0	2
18	sdetectorstest	2021-09-23 07:14:43	3	0	0
19	sdetectorstest	2021-09-23 07:15:17	4	1	0
20	sdetectorstest	2021-09-23 08:35:39	0	0	0
21	sdetectorstest	2021-09-28 12:03:51	5	0	0
22	sdetectorstest	2021-09-28 12:05:19	1	1	5
23	sdetectorstest	2021-09-28 12:06:19	0	0	4
24	sdetectorstest	2021-09-28 13:03:40	0	0	0

[사진 8] 감정일기 분석 결과

위 화면은 감정 일기 분석 알고리즘을 통해 일기 데이터를 분석한 결과를 서버에서 확인한 것이다. 표시된 부분은 순서대로 사진 5, 6, 7의 분석 결과이다.

사용자가 감정-좋음을 선택한 2021-09-28 12:03:51 일기([사진 5]에 해당) 데이터 분석 결과, 감정 단어의 빈도수는 좋음 5개, 보통 0개, 나쁨 0개이다.

사용자가 감정-나쁨을 선택한 2021-09-28 12:05:19 일기([사진 6]에 해당) 데이터는 감정 단어 빈도수 좋음 1개, 보통 1개, 나쁨 5개라는 분석 결과를 얻었

다.

위의 두 경우는 사용자의 선택 감정과 일기 데이터 분석 결과 감정이 일치하므로, 사용자가 자신의 감정을 잘 인지하고 있다고 판단할 수 있다. 따라서 감정 매칭 알고리즘 중 해당 부분에서 긍정적으로 평가된다.

하지만 사용자가 감정-보통을 선택한 2021-09-28 12:06:19 일기([사진 7]에 해당) 데이터를 분석하면 감정 단어 좋음 0개, 보통 0개, 나쁨 4개라는 결과를 얻는다.

이러한 경우에는 사용자가 인지하고 있는 감정과 일기에서 드러나는 사용자 내면의 감정에 차이가 존재하는 것이므로, 감정 매칭 알고리즘 중 심리 상태를 평가하는 과정에서 부정적으로 작용한다. 따라서, 종합적인 심리상태를 사용자에게 마음 정리 보고서(앱 내 레포트)를 통해 알릴 때 해당 내용을 추가하여 심리상태를 정확하게 인지할 수 있도록 도움을 준다.

4. 결론

본 연구에서는 효과적인 감정 분석을 위한 알고리즘 제작에 대해 알아보았다. 사용자 로그 분석 라이브러리와 Python Pandas 라이브러리에 기반한 자체 감정분석 알고리즘은 기존의 사용자 스스로 입력하는 데이터에 기반한 스트레스 측정에서 나아가, 주체적 데이터 분석을 통해 사용자의 지각하지 못했던 스트레스를 측정하고 분석 결과를 제공한다.

S-detector는 본 연구의 감정 분석 알고리즘을 적용한 응용 프로그램으로, 앞으로 프로그램을 사용함에 따라 다량의 데이터를 축적하고, 감정 측정의 집단별, 개인별 더욱 정확한 기준을 마련하길 기대한다. 또한, 위 기준을 바탕으로 사용자 개개인에게 맞춤형이며 객관적인 피드백 기능이 수행되어 심리 상담의 실용적인 보조 도구가 되기를 바란다.

향후 연구에서는 자체 감정 평가 알고리즘을 API로 개발하여 활용 분야를 넓히고, 웨어러블 디바이스들과의 연동을 통한 생체 정보 데이터를 바탕으로 보다 정확한 스트레스 감지 및 평가를 실시할 것이다.

Acknowledgement

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

- [1] 서태설, 최선희, 최희운, “학술정보 시각화 서비스 전략”, 한국과학기술정보연구원, 2014
- [2] 손선주, 박미숙, 박지은, 손진훈, “한국어 감정표현단어의 추출과 범주화”, 감성과학, vol.15, no.1, pp. 105-120
- [3] “국민 건강 통계”, 질병관리청 국민건강영양조사, 2019, pp. 80