

빅데이터 분석 기반 자동 야구기사작성 로봇 개발

채병주*, 박경철**, 권태정*, 김상환*, 김영준*, 이동규*, 김병규*

*신문대학교 컴퓨터공학과

**금오공과대학 전자공학부

e-mail: bg.kim@mpcl.sunmoon.ac.kr

Development of Automatic Robot for Writing Baseball Articles Based on BigData Analysis

Byeong-Ju Chae, Kyung-Chul Park, Tae-Jeong Kwon, Sang-Hwan Kim, Young-Ju Kim, Dong-Gyu Lee, Byung-Gyu Kim

요 약

세계적으로 자동기사 작성 시스템을 적용한 사례가 증가하는 가운데 미국의 LA 타임스가 2013년에 지진 속보 기사를 완성하면서 주목 받기도 했으며 국내에서도 비슷한 사례들의 개발이 지속적으로 진행되고 있다. 본 논문에서는 로봇저널리즘 시대에 맞춰 빅데이터 기반의 자동기사작성 로봇 개발방법을 제안하고, 데이터를 수집해 기사를 추출해내는 빅데이터 기반의 자동기사작성 로봇을 객관적인 데이터를 수집할 수 있는 스포츠분야에 적용하여 개발하였다. 본 기술로 인해 객관적인 내용을 표현할 수 있는 기사를 실시간으로 작성 및 온라인으로 배포 할 수 있으며, 확장을 통해 온라인 시청자들에게 다양한 실시간 정보제공이 가능할 수 있을 것으로 예상된다.

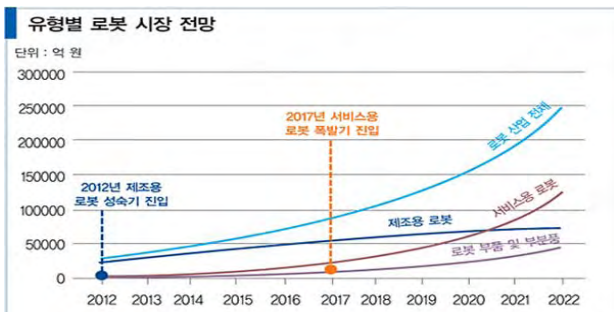
1. 서론

자동 야구기사작성로봇은 대형 포털사이트에서 파싱해온 데이터를 분석·가공하여 경기가 종료되면 경기에 대한 내용을 자동으로 기사를 작성해주는 로봇을 말한다[1].

그림 1은 한국로봇 산업협회 유형별 로봇 시장 전망 지표를 나타낸 것이다. 2017년 서비스용 로봇 성장기를 예상하고 이후 서비스용 로봇이 기타 로봇 산업과 비교했을 때 변화폭이 큰 것을 확인할 수 있다. 이처럼 서비스용 로봇 중 로봇이 자동으로 기사를 작성하는 로봇저널리즘시대에 다가가기 위한 개발이 전 세계적으로 시도되고 있다[2].

세계적으로 자동기사 작성 시스템을 적용한 사례가 증가하고 대표적으로 미국의 LA 타임스가 2013년에 지진 속보 기사를 완성하면서 주목 받기도 했으며 국내에서도 비슷한 사례들의 개발이 지속적으로 진행되고 있다.

본 논문에 소개된 프로젝트는 향후 거대한 경쟁자원이 될 것이라 예상되는 서비스용 로봇을 개발하여 로봇저널리즘 시대에 적용하고자 하여 여러 분야 중 객관적인 데이터를 대형포털에서 쉽게 수집 할 수 있는 스포츠 분야 중 야구를 선택해 빅데이터 분석 기반의 자동 야구기사 작성로봇을 개발하였다 [3],[4],[5],[6].



(그림 1) 서비스 로봇 시장 전망

2. 개발된 야구기사작성 로봇 흐름도

그림 2는 개발된 야구작성 로봇의 흐름도를 나타낸 것이며, 대형포털 스포츠 중계사이트에서 데이터를 얻어와 SBBA_Server에서 수집한 데이터를 분석·가공하여 경기에 대한 기사를 추출하여 데이터베이스에 저장하고 저장된 기사를 웹으로 서비스 할 수 있도록 개발하였다.



(그림 2) 로봇 시스템 구성도

3. 기사 작성 알고리즘

대형포털 스포츠 중계사이트에서 수집한 데이터와 SBBA서버에서 분석·가공된 데이터에서 타자, 아웃 카운트, 헛스윙, 홈런, 안타, 스코어 등 진행 중인 경기의 데이터를 수집 및 데이터베이스에 저장한다. 대형포털에서 얻어온 데이터 값 중 ‘경기종료’ 문자열이 나오면 데이터 수집을 중단하고 수집된 데이터를 기반으로 키플레이어 선정과 경기결과 분석이 이루어진다.

키플레이어를 선정하는 과정에서 키플레이어 선정 로봇은 데이터베이스에 저장된 데이터 값들을 집 퍼테이트가 고안해낸 XR(추정득점) 계산식 (1)을 참고하여 키플레이어를 선정한다. 계산식을 통해 알 수 있듯이 표 1의 값에 특정한 가중치를 곱해서 더한 값이다.

타석에 선 타자마다 XR값이 계산되어 저장되며 이들 중 가장 큰 수치를 가진 선수가 해당 경기의 키플레이어로 선정된다.

<표 1> XR 계산을 위한 데이터 필드

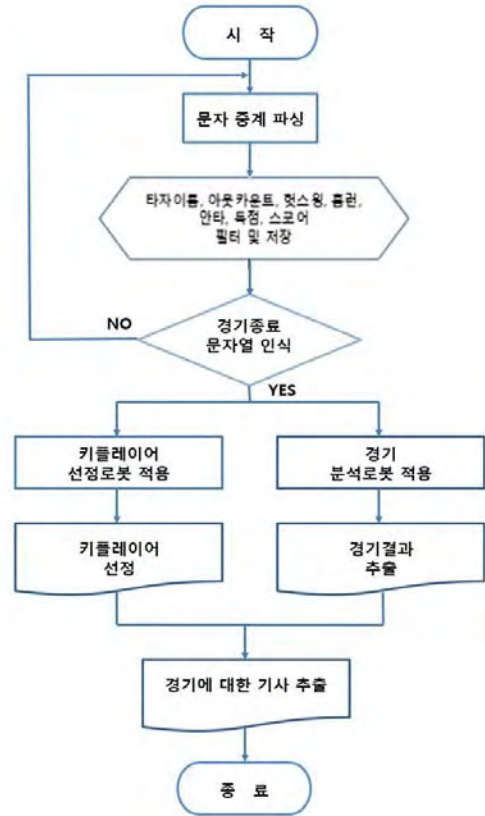
1B	2B	3B	HR
1루타	2루타	3루타	홈런
HBP	IBB	SB	CS
사고	볼넷	도루성공	도루실패
K	GIDP	SF	SH
삼진	병살	희생플레이	희생번트

$$\begin{aligned}
 XR = & (0.5*1B) + (0.72*2B) + (1.04*3B) + (1.44*HR) + (0.34* \\
 & (HBP+TBB-IBB)) + (0.25*IBB) + (0.18*SB) - (0.32*CS) \\
 & - (0.09*(AB-H-K)) - (0.098*K) - (0.37*GIDP) + (0.37*S \\
 & F) + (0.04*SH)
 \end{aligned}
 \tag{1}$$

식(1)에서 가중치는 다양한 데이터를 분석하여 설정한 값이며, 경기 분석 단계에서는 SBBA 서버에서 각 팀의 선발투수, 경기 중 팀별 스코어에 대한 데

이터를 수집하여 데이터베이스에 저장된다.

그림 3과 같이 키플레이어 선정로봇과 경기분석 로봇에서 추출한 결과 값을 기반으로 경기에 대한 기사를 추출한다.

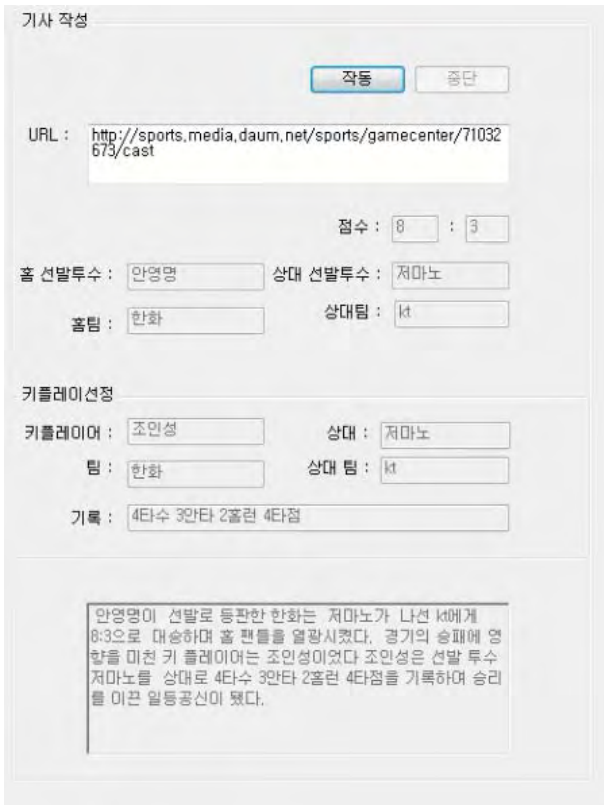


(그림 3) 기사 추출 알고리즘

4. 시스템 개발 결과

실제 구현을 위해서 윈도우 7 운영체제에서 Visual Studio 12와 SQL 서버를 활용하였다. 기사작성 로봇의 구성은 그림 4와 같으며, .Net Framework 4.0 과 WinForm을 기반으로 실시간으로 야구경기 데이터를 얻어와 진행 중인 경기를 분석해주는 Game Analysis Robot, 승리한 팀에서 해당 경기에 가장 기여도가 높은 선수를 추출하는 Choosing Key Player Robot, 분석한 데이터를 기사로 시각화해주는 Automation News Robot으로 구성되어 있다.

그림 4에서 볼 수 있듯이 실제 Daum 포털의 문자 중계에 의한 데이터를 실시간으로 분석 및 추출하여 주요 정보를 저장하게 된다. 추출된 정보를 바탕으로 ‘경기종료’ 문자열 값을 얻어 오면 경기에 대한 내용을 분석과 키플레이어를 추출하여 기사로 시각화하고 실시간 데이터 파싱이 중지된다.



(그림 4) 기사작성로봇을 통한 작성 실행 화면

미래.” 한국정보과학회지 32권, 1호 pp. 8-17, 2014.
 [4] 안창식, 오상엽, “의미 분석과 형태소 분석을 이용한 핵심어 인식 시스템,” 한국멀티미디어학회 논문지, 13권, 11호, pp.1586-1593, 2014.
 [5] Daum스포츠 야구 문자중계 서비스
<http://sports.media.daum.net/sports/gamecenter/>
 [6] Jim Furtado, “The 1999 Big Bad Baseball Annual: The Book Baseball Deserves,” Mad Aztec Press, (Kirkwood, MO), 1999.

5. 결론

본 논문에서는 대형포털의 스포츠 야구문자중계 서비스를 통해 얻어온 데이터를 분석·가공하여 자동 야구기사작성로봇을 개발하였다. 제안된 로봇은 수집된 데이터를 통해 자동으로 경기결과 분석과 선수들의 기록을 분석해 키 플레이어를 추출하여 뉴스기사로 작성해주는 기능을 지원할 수 있으며 다양한 분야로 확장하여 실시간 정보 제공 시스템으로 활용이 가능할 것이다.

ACKNOWLEDGEMENT

본 논문의 내용은 선문비트교육센터 전문가과정 29기 프로젝트 결과물을 토대로 하고 있으며, 지도 교수님과 선문비트교육센터 관계자 여러분의 협조에 감사드립니다.

참고문헌

[1] 강정수, “로봇 저널리즘” ICT인문사회융합 동향, pp.14-25, 2015.
 [2] 정만태, “로봇 산업의 국내외 시장전망 및 발전 방향,” Machinery Indusry 특집호, pp. 28-33, 2008.
 [3] 김상락, 강만모, “빅데이터 분석 기술의 오늘과