JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 20, No. 2, pp. 335-342, Feb. 2019

게임 시청자 실시간 한글 채팅의 언어 자질에 기반한 하이라이트 탐지

최 승 혁 · 박 상 준 · 한 재 현 · 이 현 아*
금오공과대학교 컴퓨터소프트웨어공학과

Highlight Detection using Linguistic Features of Real-time Korean Game Audience Chat

Seung-Hyuk Choi · Sang Jun Park · Jae-Hyun Han · Hyun Ah Lee*
Department of Computer Software Engineering, Kumoh National Institute of Technology, Gumi, 39177, Korea

[요 약]

멀티미디어 콘텐츠의 생성과 공유가 활발해지면서 동영상 콘텐츠에서 자동으로 하이라이트를 추출할 필요성이 커지고 있다. 근래 디지털 방송에서는 시청자의 실시간 채팅이 일반화되고 있으며 이 채팅은 하이라이트 자동 추출에 좋은 정보원이지만 한글 채팅의 특징을 반영한 연구는 찾기 힘들다. 본 논문에서는 게임 동영상의 시청자 실시간 한글 채팅의 특징을 분석하여 하이라이트 추출에 기여할 수 있는 학습 자질을 선택하고, 하이라이트가 수동 태깅된 채팅을 활용하여 기계 학습을 통해 하이라이트를 탐지하고자 한다. 구축된 시스템에 대한 평가에서는 한글 채팅의 다양한 언어적 정보를 사용하여 하이라이트 추출 성능이 증가하는 것을 확인할 수 있었으며, 하이라이트 구간의 약 87%를 자동으로 탐지하여 하이라이트 제작의 효율성을 높일 수 있었다.

[Abstract]

As creating and sharing multimedia contents become easy and popular, the need for automatic highlight extraction from video contents is increasing. In recent digital broadcasting, real-time chat of audiences that can be a good source of automatic highlighting is becoming common, but highlight detecting based on characteristics of Korean chat is hard to find. In this paper, we analyze characteristics of real-time chat in game videos, select features that can contribute to the highlight extraction, and detect highlights through machine learning using the manually tagged chat. In evaluation, using various Korean linguistic features showed better results than only using utterance frequency. And about 87% of highlights were automatically detected so the efficiency of highlight production was improved.

색인어 : 실시간 시청자 채팅, 한글 채팅 특성, 언어 자질, 영상 하이라이트 탐지

Key word : Audience Live Chat, Characteristics of Korean Chat, Linguistic Features, Video Highlight Detection

http://dx.doi.org/10.9728/dcs.2019.20.2.335

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 22 January 2019; **Revised** 31 January 2019
Accepted 20 February 2019

***Corresponding Author; Hyun Ah Lee**

Tel: +82-54-478-7546
E-mail: halee@kumoh.ac.kr

Ⅰ. 서 론

IT기술의 발달로 공중파 방송에서부터 1인 방송의 개인 저작물까지 방대한 규모와 양의 멀티미디어 콘텐츠가 생성되고 공유되고 있다. 이로 인해 콘텐츠로부터 중요한 부분들만 요약하여 제공하는 하이라이트(highlight)에 대한 사용자 요구도 높아지고 있다. 하지만 하이라이트 추출을 위해 콘텐츠를 수동으로 재가공하는 작업은 시간과 비용의 부담이 매우 크다.

멀티미디어 동영상 콘텐츠로부터 자동으로 하이라이트를 추출하고자 하는 다양한 연구가 지속적으로 시도되어 왔다. 동영상의 하이라이트 추출이나 탐지의 초기 연구는 스포츠 중계의 아나운서 오디오에 기반한 연구[1]나 중계 화면의 자막 구성에 기반한 연구[2]와 같이 공중파 방송 중계에서 제공되는 고비용의 추가 정보에 대부분 의존한다. 이후 소셜 미디어의 등장으로 트위터(twitter)을 활용한 연구들[3,4]이 진행되었으나 트윗(tweet)의 발생 빈도와 같은 단편적인 정보만을 사용하였으며, 대량의 트윗이 발생하는 공중파 방송 중계 등에만 적용할 수 있는 한계가 있다.

근래 딥러닝의 발전으로 비디오 영상 정보로부터 직접적으로 하이라이트를 추출하는 연구들[5,6]과 함께 영상 시청자들의 실시간 채팅 정보를 활용한 연구들[6,7,8]이 진행되고 있다. 디지털 기반의 중계와 방송의 일반화로 대부분의 동영상 콘텐츠와 함께 제공되는 시청자의 실시간 채팅은 하이라이트 추출의 좋은 정보원이다. [6]에서는 게임 영상과 시청자 채팅 정보를 결합한 하이라이트 탐지 방식을 제안하였으며, 영어와 중국어 시청자간 채팅에서 다수 발생하는 오타와 약자, 이모티콘의 처리를 위해 단어 기반이 아닌 글자 기반으로 LSTM (Long Short Term Memory) 모델을 사용하여 하이라이트를 추출한다. [7]에서는 게임 전문 방송국에서 제작한 태깅된 하이라이트와 영어 채팅을 활용하여 하이라이트를 추출하였으나, 초당 최소 1.61개, 평균 4.35개의 채팅을 가지는 고빈도 영어 채팅에서 채팅 트래픽의 양적 피크 위치를 사용하는 단순한 방법으로 하이라이트를 결정한다. [8]에서는 디시인사이드, 네이버, 유튜브, twitch TV등에서 얻은 한글과 영어 채팅으로부터 드라마, 예능, 스포츠, 게임을 포함한 인터넷 방송과 실시간 채팅을 자동으로 수집하고 Word2Vec을 활용하여 하이라이트 구간의 제목을 자동으로 생성하는 효과적인 시스템을 제시하고 있다. 유일하게 한글 채팅을 대상으로 하고 있지만, 하이라이트 추출에서는 기존 연구와 유사하게 채팅 양만을 활용하며, 하이라이트 추출 연구에 대한 객관적인 평가 결과를 제시하지 못 하고 있다.

이처럼 기존 연구들은 대량의 채팅이 발생하는 인기 영상의 영어 채팅을 주로 대상으로 하고 있으며, 한글 채팅의 특징을 반영한 연구는 찾기 힘들다. 본 논문에서는 게임 동영상 실시간 시청자의 한글 채팅의 언어적 특징을 활용하여 하이라이트를 추출하는 방법을 제안한다. 시스템 구축을 위하여 온라인에서 게임 동영상과 시청자 채팅을 수집하고, 하이라이트 구간을 수동으로 태깅하였으며, 게임 시청자의 한국어 채팅의 특징을 분석하여 하이라이트 구간 탐지에 적합한 자질을 제안한다. 제안한 시스템은 다양한 언어 정보를 활용하여 대량의 채팅이 없는 경우에도 적용할 수 있다.

Ⅱ. 게임 시청자 실시간 채팅 데이터

2-1 데이터 수집

사용자 채팅으로부터 하이라이트를 추출하기 위해 Twitch TV[9]에서 임의로 선택한 게임 동영상 10개에 대한 시청자의 실시간 채팅을 수집하였다. 아래 그림 1은 수집된 채팅의 일부분이다. 각 줄은 채팅의 발생 시점을 동영상의 시작 시간 기준으로 나타낸 [시간:분:초] 정보와 채팅자 id, 채팅 발화(utterance) 내용으로 구성된다. 동영상에 대한 하이라이트는 수동 태깅으로 추출하였으며, 그림 오른쪽 아래 상자와 같이 “00:32:00~ 00:32:40”의 형태로 시작 시간과 끝 시간으로 하이라이트 구간을 나타낸다.

그림 1에서 상단의 00:10:05~00:10:10 구간에서 발생한 채팅은 하이라이트가 아닌 비(非)하이라이트 구간이며 하단의 00:32:00~00:32:13 구간의 채팅은 하이라이트 구간에 해당한다. 예제에서는 채팅당 어절이나 음절의 길이가 하이라이트 구간에서 짧아지며 각 구간에서 나타나는 단어들에도 차이가 있는 것을 관찰할 수 있어, 이러한 채팅의 언어적 특성이 하이라이트 추출에 기여할 것으로 예상된다.

```
[00:10:05.389] 세실: 이거 삼성 모니터 기존 뭔지아시는분?
[00:10:07.469] shadyzunba: ㅋㅋㅋㅋㅋ
[00:10:07.803] 하도르: 히 익
[00:10:09.307] inhoch0331: ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ
[00:10:09.907] 쳔슥: 논란되서 사실이면 실격시키면되지멀
[00:10:09.913] 아루선생: 방플은 모르겠고 판다 티밍은 의심 강력함
[00:10:10.398] 펠프스입니다: 뒤에 심판있는거안보이나 그런거하나
   신경안쓰것냐
[00:10:10.549] woe2aw: 1인칭은 방플 가능하다 한자리에 10분씩 짱
   박히고 하는데
[00:10:10.672] 준성이: ㅋㅋ
....
[00:32:00.087] 한국제록스: 퍼스트블러드
[00:32:00.911] fvwowen: 지가 쏠라고 클릭하는겨
[00:32:02.739] jjla95: 샷건계의 awm
[00:32:03.444] 춘샷건: 옵저버 고의적인가?
[00:32:06.003] 아러피: 일인칭은 ㄹㅇ공포겜이야
[00:32:07.405] skinhoney: 해외 옵저버 보니까 인벤갓맞군 ㅋㅋㅋㅋ
   ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ
[00:32:11.157] 4sanagi: 와
[00:32:12.360] 는븝를: 와
[00:32:12.582] 앞산: 오
[00:32:13.281] 소루마키: 와
[00:32:13.797] 김도아조씨따라미국옴: 와
```

```
00:32:00~00:32:40
00:38:35~00:39:00
00:40:55~00:41:05
00:44:31~00:44:50
00:46:28~00:48:40
.....
```

그림 1. 게임 시청자 실시간 채팅의 예
Fig. 1. Example of Live Chat of Game Audience

2-2 데이터 분석

1) 게임 동영상과 시청자 한글 채팅의 특징

표 1은 수집된 데이터의 정보를 요약한다. 영상의 길이(Video Length)는 최소 1시간 24분에서 최장 11시간 48분까지 다양하게 수집되었다. 게임 동영상은 경기 시간이 정해져 있는 일반 스포츠 중계에 비해 플레이 시간이 길어 하이라이트에 대한 사용자 요구가 더 높을 것으로 예상된다. 채팅의 개수(Chat Count)는 천여 개에서 9만여 개, 채팅 참여자수(Audience Count)도 90여명에서 만여 명으로 다양하게 나타났다. 초당 평균 채팅의 개수는 1.76개로 동영상 #6이 가장 낮은 0.19개, 동영상 #2가 5.14개로 나타냈다.

게임 시청자 실시간 채팅의 언어적 특징을 파악하기 위해 한 채팅의 평균 음절과 어절, 한글 낱글자의 채팅 내 평균 개수를 분석했다. 음절(Syllable)은 하나의 종합된 음의 느낌을 주는 말소리의 단위로 텍스트에서는 한 글자를 의미한다. 어절(Word)은 띄어쓰기를 기준으로 나뉘는 문장 성분의 최소 단위이며, 낱글자(Single Char.)는 'ㅋ'나 'ㅜ'와 같이 다른 자소와 조합되지 않고 발생한 한글 자소를 의미한다. 예문 "일인칭은 ㄹㅇ공포겜이야"는 11개의 음절, 2개의 어절, 2개의 낱글자를 가진다. 표 2에서는 비교 분석을 위해 세종말뭉치[10]에서 실제 대화를 전사한 구어체 말뭉치 일부를 분석하여 얻은 통계를 보인다. 말뭉치로는 게임 채팅과 유사할 것으로 기대되는 학생 대화자들의 일상 잡담(Student Chat), 놀이문화(Playing), 웹사이트(Web sites), PC방(PC room)을 선택했다.

비교 결과에서는, 온라인에서의 한글 채팅이 실제 대화보다 한 발화가 짧을 것으로 예상하였으나, 표 1의 게임 채팅의 음절 수가 8.04~12.47, 어절 수가 1.72~3.23개로 나타나 표 2의 구어체 말뭉치의 11.79~14.74와 2.82~3.52에 비해 그 격차가 크지 않은 것으로 나타났다. 또한 온라인에서의 한글 채팅에서는 띄어쓰기가 잘 이루어지지 않을 것으로 예상하였으나, 음절수 대비 평균 어절 수가 4.6으로 말뭉치의 4.35보다 크게 나타났다. 이는 동영상 시청자의 실시간 채팅이 모바일이 아닌 PC 환경에서 이루어져서 키보드 입력에서 습관화된 띄어쓰기가 긍정적으로 발현된 것으로 보인다. 그림 1에서도 한 발화를 제외하고는 충분한 띄어쓰기가 이루어졌으며 비문의 비율도 높지 않은 것을 알 수 있으며, 이로부터 채팅에 대한 언어 분석으로 유효한 자질들을 추출할 수 있을 것으로 기대된다.

표 1. 하이라이트 태그된 게임 영상과 채팅 통계
Table 1. Statistics of highlighted game videos and chats

#	Video Length (H:M:S)	Chat Count	Audience Count	Avg Count per Chat			Highlight Length (M:S)
				Syllable	Single Char.	Word	
1	1:24:38	4129	1131	11.92	1.07	3.23	05:20
2	5:03:44	93583	11162	8.70	2.27	2.24	19:06
3	4:29:37	4848	439	8.04	1.22	2.10	07:20
4	6:08:04	24110	1709	8.83	3.24	1.88	13:00
5	11:16:53	9581	827	10.25	1.14	2.70	19:42
6	1:42:27	1174	94	10.04	0.94	2.46	06:31
7	3:07:42	44517	3291	9.69	4.84	1.72	09:55
8	7:03:21	25900	2892	12.47	2.14	2.28	22:46
9	11:48:56	29419	3249	9.52	2.61	1.99	21:17
10	4:28:00	68158	8358	8.85	2.17	2.37	15:49

표 2. 한국어 구어체 말뭉치의 통계
Table 2. Statistics of Korean colloquial corpus

Corpus	Chat Count	Audi-ence Count	Avg Count per Chat		
			Syllable	Single Char.	Word
2 Student Chat-1	273	2	13.76	-	3.52
2 Student Chat-2	303	2	11.79	-	3.09
3 Student Chat	592	3	14.28	-	3.32
4 Student Chat	1833	4	14.05	-	2.82
Conversation about Playing	2580	3	13.07	-	3.38
Conversation about Websites	1001	3	14.74	-	2.86
Topic Conversation about PC-room	702	4	12.07	-	2.89

2) 게임 시청자 채팅의 이모티콘 특징

그림 1에서 볼 수 있듯이 한글 시청자 채팅에서는 'ㅋㅋ'나 'ㅜㅜ'와 같은 한글 낱글자열이 많이 나타난다. 표 1에서는 한 채팅 발화가 평균 0.94~4.84개의 한글 낱글자를 포함하여, 음절수 대비 낱글자수 비율이 30.26%로 높게 나타났다. 특히 7번 동영상의 경우 음절 중 49.97%가 낱글자였으며, 가장 긴 경우에는 'ㅋ'이 122번 연속하여 나타난 경우도 있었다. 이에 반하여 구어체 말뭉치의 경우 녹음 전사의 특성상 한글 낱글자가 한 번도 발생하지 않았다.

게임 실시간 채팅에서 이모티콘에 사용되는 특수문자의 평균 개수는 채팅 당 0.23개에서 0.39개로 나타났다. 전체 채팅에서 한글 단일자소의 연속인 'ㅋㅋ'와 그 반복열은 71135번, 'ㅡㅡ' 반복열은 3769번 발생한 것에 비하여, 이모티콘 ':)'와 ';)'는 전체 채팅에서 33번 발생하였으며 모바일에서 자주 사용되는 '^^' 형태는 한 번도 발생하지 않았다. 이러한 데이터로부터 인터넷 사용자들이 한글 낱글자를 사용한 줄임말과 감정 표현을 선호하는 것을 수치적으로 확인할 수 있었다.

3) 하이라이트 구간의 한글 채팅의 특징

하이라이트가 태깅된 사용자 채팅으로부터 하이라이트 구간의 채팅과 비하이라이트 구간의 채팅의 특성을 비교 분석하였다. 분석에서는 표 1의 동영상 #1을 활용하였다. 표 3의 Avg. Count per 1 sec는 하이라이트 구간과 비하이라이트 구간에서

표 3. 하이라이트와 비하이라이트의 채팅 통계 비교
Table 3. Comparison of chat statistics between highlights and non-highlights

	Avg. Count per 1 sec				Avg. Count per Chat		
	Chat	Syllable	Single Char.	Word	Syllable	Single Char.	Word
highlight	1.23	9.38	1.49	2.71	7.61	1.21	2.21
nonhighlight	0.78	9.71	0.83	2.62	12.38	1.06	3.34

의 1초당 평균 채팅 개수와 음절, 어절, 한글 낱글자 수를 보인다. 하이라이트 구간에서는 초당 평균 1.23개의 채팅 발화가 나타났으나 비하이라이트 구간에서는 0.78개가 발생하여 하이라이트 구간에 더 많은 채팅이 발생했다. 초당 어절 수와 음절 수는 큰 차이가 나지 않았으나 낱글자 수는 하이라이트 구간에서의 값이 크게 나타났다.

그림 1에서는 하이라이트 구간과 비하이라이트 구간의 채팅에서 발화 길이의 차이가 관찰되었다. Avg. Count per 1 sec의 초당 어절수와 음절수는 채팅 개수가 많을수록 증가하므로, 표 3의 Avg. Count per Chat에서는 채팅 당 음절, 어절, 낱글자 개수를 분석하였다. 결과에서는 비하이라이트에서 채팅 당 음절수와 단어수가 확연히 높게 나타났다. 낱글자 수는 초당 개수와 채팅 당 개수가 모두 하이라이트 구간에서 높게 나타났다. 이로부터 하이라이트 구간에서 낱글자를 포함한 짧은 발화가 많이 발생함을 알 수 있었으며, 이러한 특징에서 음절, 낱글자, 어절 수가 하이라이트 분별에 의미있는 정보로 사용될 수 있음을 기대할 수 있다.

게임 시청자 채팅에서 하이라이트 구간과 비하이라이트 구간에서는 사용되는 단어에 차이가 있을 것으로 예상된다. 예를 들어 '지겹다'라는 단어가 발생하는 구간은 비하이라이트 구간에 해당할 가능성이 크고, '대박'이 발생하는 구간은 하이라이트일 가능성이 크다. 실제 수집된 동영상 #1에서 형용사 '지겹다'는 비하이라이트 구간에만 발생하였으며, 감탄사 '오오'는 하이라이트 구간에서만 발생하였다. 이처럼 하이라이트 구간과 비하이라이트 구간에서 발생 양상에 큰 차이를 보이는 단어는 하이라이트 탐지에 긍정적인 역할을 할 것으로 기대된다.

Ⅲ. 시청자 채팅을 이용한 하이라이트 탐지

3-1 구간 윈도우에 기반한 학습 자질

시청자 채팅으로부터 하이라이트를 탐지하기 위해 채팅 발화 각각이 하이라이트에 맞는 채팅인지 아닌지를 판단한다면 추출할 수 있는 특징이 부족하여 정확한 결과를 얻기 어렵다. 초 단위 구간에 발생하는 채팅을 모아서 사용하는 방식도 데이터 부족 문제에 시달리기 쉽다. 이러한 문제를 해결하기 위해 본 연구에서는 채팅에 구간 윈도우를 적용하여 학습 데이터를 추출한다. 본 논문에서는 3초와 5초 길이의 두 가지 크기의 윈도우를 실험에서 사용한다. 구간 윈도우 추출에서는 윈도우 구간을 1초 간격으로 겹치게 이동하면서 자질을 얻는다. 게임 시청자 채팅은 동영상 시청 후 반응이 발생하므로 시작 시간을 기준으로 구간 윈도우를 설정한다. 예를 들어 0초에서 3초 직전까지의 채팅을 0초의 정보로, 1초에서 4초 직전까지의 채팅을 1초의 정보로 사용한다.

하이라이트가 태깅된 채팅은 기계 학습(machine learning)에 좋은 학습 데이터가 될 수 있으나, 채팅의 단어나 문장을 기계 학습의 입력으로 그대로 사용하면 자료 희소성(data sparseness) 문제로 올바른 결과를 얻기 어렵다. 본 논문에서는 구간 윈도우의 자질을 수치 벡터로 변환하여 기계 학습에 사용하며, 2-2의 3)의 분석에 기반하여 아래 목록과 같이 ① 구간 윈도우에 발생하는 채팅의 단순 빈도 정보와 ② 언어적 빈도 정보, ③ 내용 정보로 하이라이트 추출을 위한 학습 자질을 구성한다. 아래에서는 각각에 대해서 설명한다.

① 구간 내 채팅의 기본 빈도 정보(onlyFreq)
- ①-1. 구간 내 채팅 수
- ①-2. 전체 구간에서 현재 윈도우의 시간적 위치 비율
- ①-3. 구간 내 채팅 참여자 수 비율(중복 id제외)

② 구간 내 채팅의 언어적 빈도 정보(withLang)
- ②-1. 구간 내 특수문자 총 개수
- ②-2. 구간 내 한글 낱글자 총 개수
- ②-3. 구간 내 총 음절수
- ②-4. 구간 내 총 어절수
- ②-5. 구간 내 한 채팅의 평균 음절수
- ②-6. 구간 내 한 채팅의 평균 어절수
- ②-7. 구간 내 어절의 평균 음절길이

③ 구간 내 채팅의 내용 정보
- ③-1. 구간 내 형태소의 긍정 점수합(Pscore)
- ③-2. 구간 내 형태소의 부정 점수합(Nscore)
- ③-3. 구간 내 형태소열에 대한 Doc2Vec 벡터

① 기본 빈도 정보(onlyFreq)

자질 ①-1은 기존 연구에서 사용하던 구간내 채팅 수를 값으로 가진다. 본 연구에서는 구간내 채팅수와 함께 전체 동영상에서 해당 채팅의 시간적 위치를 [0,1] 범위의 상대값으로 표현한 자질 ①-2와 참여자 수의 자질 ①-3을 기본 빈도 정보로 사용한다. 자질 ①-2는 구간의 위치 비율로 전체 1000초인 동영상에서 20초에 시작하는 구간 윈도우는 0.02를 값으로 가지며, 동영상의 위치에 따른 하이라이트 가능성을 반영한다. 자질 ①-3은 전체 동영상의 전체 채팅 참여자 수를 현재 구간의 채팅 참여자 수로 나눈 값으로, 하이라이트 구간일수록 다양한 참여자가 나타날 가능성을 반영한다.

② 언어적 빈도 정보(withLang)

언어적 빈도 정보인 자질 ②의 ②-1~②-7은 음절과 어절, 특수문자, 낱글자의 구간 내 총 개수나 채팅 당 평균 개수를 자질로 가진다. 자질 ②-1은 한글이나 영어 문자가 아닌 이모티콘에 해당하는 특수 문자의 구간 윈도우에서의 발생 수, ②-2는 'ㅋ'이나 'ㅜ'와 같은 한글 낱글자의 구간 윈도우에서의 발생 수다. ②-3은 음절의 개수, ②-4는 어절의 개수를 값으로 가진다.

②-3와 ②-4의 음절과 어절의 수는 구간 윈도우에서 많은 채팅이 발생할수록 큰 값을 가지므로, 2-2절의 3)에서 본 바와 같이 하이라이트 구간일수록 채팅의 길이가 짧아지는 특성을 반영하기 어렵다. 자질 ②-5와 ②-6은 구간 윈도우에서의 채팅 발화별 평균 음절수와 어절수로, 해당 구간의 각 채팅이 많은 단어를 가지는 긴 발화인지 적은 수의 단어를 가지는 짧은 발화인지를 하이라이트 추출에 반영한다. ②-7은 한 어절이 몇 개의 음절로 구성되는지의 평균값을 가져, 하이라이트 구간에서 짧

은 어절이 자주 발생하는 특징을 반영한다.

③ 내용 정보 (Positive Score, Negative Score and Doc2Vec)

자질 ③은 하이라이트 탐지에서 채팅 내 어휘가 미치는 영향을 반영하기 위한 내용 정보에 해당한다. 2-2절의 3)에서 '지겹다'와 '오오'의 예제에서와 같이 하이라이트 구간과 비하이라이트 구간에서 발생 비율에 큰 차이를 보이는 단어들은 하이라이트 탐지에 기여할 것으로 기대된다. 이 때 그림 1의 하이라이트 구간에 많이 발생하는 '와'는 한글에서 감탄사로 사용될 수도 있지만, 'x와 y'에서처럼 조사로 사용될 수 있다. 이러한 품사에 따른 단어의 다른 의미를 반영하기 위해서 채팅을 형태소 분석한다. 게임 채팅에서는 사전에 등재되지 않은 다양한 표현들이 하이라이트 탐지에 기여하므로, 불용어(stop word)에 해당하는 어미와 조사, 보조용언을 제외하고, 명사와 동사, 형용사, 부사, 기호, 낱글자, 외래어 등의 대부분의 품사 어휘를 추출한다. 형태소 분석기에 따라 각기 다르게 판별되는 품사는 명사(N), 동사(V), 관형사(M), 부사(Adv), 형용사(Adj), 감탄사(I)의 품사 최상위 분류로 변환하고, 이를 제외한 채팅에서 자주 발생하는 외래어나 미등록어 등의 품사는 모두 POS로 통일한다. 시스템에서는 '지겹다/Adj'로 형태소와 품사를 결합한 형태를 한 어휘로 간주한다. 형태소 분석기로 KoNLPy[11]에서 제공하는 여러 형태소 분석기 중 구어체에 적합하다고 알려진 twitter 분석기가 미세하지만 가장 높은 정확성을 보여 구현된 시스템에서는 이를 사용하였다. 추가적인 후처리로 'ㅋㅋㅋㅋㅋㅋㅋㅋ'나 'ㅂㄷㅂㄷㅂㄷ'과 같은 자소열의 다양한 반복 패턴을 동일 낱글자의 3번 이상의 반복은 2개로 줄이고(ex. 'ㅋㅋ') 두 낱글자의 6번 이상의 반복은 4개로 줄여서(ex. 'ㅂㄷㅂㄷ') 한글 채팅의 특성을 반영하고 자료 부족 문제를 감소시킨다.

내용 정보 자질의 ③-1의 긍정 점수합 Pscore(Positive Score)는 구간 내 채팅들이 하이라이트에 자주 나타나는 단어들을 얼마나 포함하는지, 부정 점수합 Nscore(Negative Score)은 비하이라이트에 자주 나타나는 단어들을 얼마나 포함하는지를 반영한다. 수식 (1)~(3)은 긍정 점수 계산을 위한 수식을 보인다. 수식 (1)은 단어 *w*의 긍정 점수 *Pscore(w)*를 구한다. 식에서 집합 *H*는 하이라이트 구간으로 태깅된 시간대에 발생한 모든 채팅을, 집합 $\overline{H}$는 비하이라이트 구간으로 태깅된 시간대에 발생한 모든 채팅을 의미한다. 단어 *w*의 하이라이트에서의 비율 *Ratio(w, H)*가 비하이라이트에서의 비율 *Ratio(w,* $\overline{H}$*)*보다 크면, 단어 *w*는 하이라이트에서 더 자주 발생하는 단어이다(수식에서 *else*의 경우). 이 단어들에 대해서 *Ratio(w,* $\overline{H}$*)/Ratio (w,*

$$Pscore(w) = \begin{cases} 0\ , & \text{if } Ratio(w,H) \le Ratio(w,\overline{H}) \\ 1 - \dfrac{Ratio(w,\overline{H})}{Ratio(w,H)}, & else \end{cases} \quad (1)$$

$$Ratio(w,S) = \frac{freq(w,S)}{\sum_{w_i \in S} freq(w_i,S)} \quad (2)$$

$$Pscore(W) = \frac{\sum_{w_i \in W \text{ and } Pscore(w_i) \ge 0.3} Pscore(w_i)}{\sum_{w_j \in W} freq(w_j, W)} \quad (3)$$

*H)*으로 [0,1]의 범위를 가지는 값을 얻고 이를 다시 1에서 빼서 하이라이트에서 비율이 높을수록 큰 값을 가지는 범위 [0,1]의 값을 얻는다. 이를 단어 *w*의 긍정 점수 *Pscore(w)*로 표현한다. 비하이라이트에서의 발생 비율이 더 큰 단어의 긍정 점수는 0으로 한다. 수식 (2)는 *Ratio(w, H)*와 *Ratio(w,* $\overline{H}$*)*를 구하는 방법을 보인다. 수식 (2)에서 *freq(w, S)*는 단어 *w*가 채팅 집합 *S*에서의 몇 번 나타났는지의 빈도를 의미한다. 이 빈도를 그대로 수식 (1)에 사용한다면 하이라이트 구간에 비해 비하이라이트 구간이 월등히 길고 많은 단어를 포함하므로 공정한 긍정 점수나 부정 점수를 얻을 수 없다. 공정한 값을 얻기 위해 *w*의 빈도를 채팅 집합에 포함된 모든 단어의 빈도합으로 나누어 단어 *w*의 채팅 집합 *S*에서의 발생 비율인 *Ratio(w, S)*을 구하여 *Pscore(w)*계산에 사용한다. 수식 (3)은 구간 윈도우 *W*에서 발생하는 단어들의 *Pscore*를 합산하여 구간 윈도우의 긍정 점수인 *Pscore(W)*를 얻는다. 합산에서는 긍정 점수가 일정 수준(0.3) 이상인 단어만을 고려하여 노이즈에 의한 오류를 방지한다. 또한 구간 윈도우에 발생하는 채팅이나 단어수가 많아 구간 전체가 높은 긍정 점수를 받는 경우를 방지하기 위해, 구간 윈도우에 포함되는 단어 개수로 긍정 점수의 합을 나누어 *Pscore(W)*를 [0,1]의 범위로 정규화한다. 부정 점수 *Nscore(W)*는 수식 (1)과 (3)에서 *H*와 $\overline{H}$를 맞바꾸어 구한다. 이렇게 구한 긍정 점수와 부정 점수의 두 개 값이 자질 ③-1과 ③-2에 해당한다.

기계 학습에서 단어나 문장을 그대로 사용하는 경우 발생할 수 있는 자료 희소성 문제를 해결하는 방안으로 Word2Vec[12]이나 Doc2Vec[13] 모델이 등장하였다. 자질 ③-3은 Doc2Vec에 기반하여 구간 윈도우에서 발생하는 문장의 특징을 반영하고자 한다. 구간 윈도우에 포함되는 여러 채팅을 하나의 문서로 취급하여 학습 데이터로부터 Doc2Vec을 구성하고 각 구간 윈도우의 채팅에 대한 수치 벡터를 얻어 학습에 사용한다.

3-2 학습 모델

시스템 구축을 위한 기계 학습에서는 시계열 데이터에 좋은 성능을 내는 LSTM[14]을 Tensorflow 환경에서 사용한다. 표 1에서 볼 수 있듯이 전체 동영상 대비 하이라이트 구간의 비율이 낮아, 학습에서 태깅된 하이라이트 구간을 그대로 적용하면 모든 데이터를 비하이라이트로 판단하더라도 90% 이상의 정확률을 얻을 수 있어 편향된 결과를 얻게 된다. 시스템에서는 학습 데이터의 하이라이트 구간의 비율이 6.4% 이하인 점을 감안하여 하이라이트 윈도우의 가중치를 비하이라이트 구간의 10배로 부여한다. LSTM를 이용한 학습에서는 sequence_length를 5로 사용하였다.

3-3 하이라이트 추출 시스템

그림 2는 최종 시스템의 사용자 인터페이스를 보인다. 시스템에서는 게임 동영상과 함께 자동으로 탐지한 하이라이트 구간을 관리자에게 제공한다. 관리자는 자동 탐지된 하이라이트 구간의 동영상을 시청하여 적합하지 않는 하이라이트는 제거하고, 여러 하이라이트를 연결하여 하나로 병합하거나 하이라이트 전후에 추가하거나 삭제해야 하는 구간을 반영하여, 최종 하이라이트 영상을 생성한다. 이러한 인터페이스에서는 정답 하이라이트 구간에 포함되는 여러 구간 윈도우 중 하나 이상의 구간 윈도우가 하이라이트로 탐지되면 편집을 통해 쉽고 빠르게 하이라이트 영상을 제작할 수 있다. 이 점에 기반하여 다음 장에서는 추출정확률을 기준으로 시스템을 평가한다.

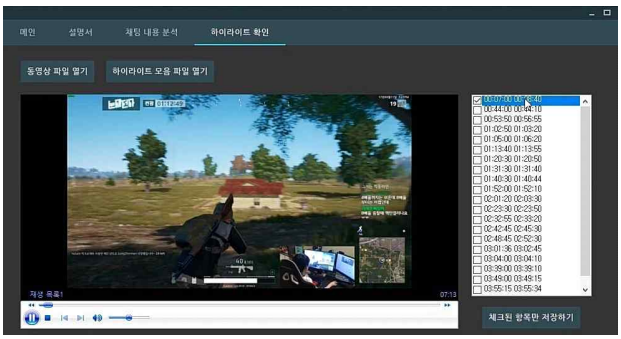

그림 2. 하이라이트 추출 시스템의 사용자인터페이스
Fig. 2. User Interface of Highlight Extraction System

Ⅳ. 실험 및 평가

제안한 하이라이트 추출 시스템의 성능을 평가하였다. 평가에서는 2.1에서 수집한 데이터에 대하여 leave-one-out 방식을 사용하여, 표 1의 열 개 영상 중 하이라이트를 추출하려는 하나의 영상을 제외한 나머지 9개 영상을 학습 데이터로 사용하여 대상 영상의 하이라이트를 탐지하고, 10개 동영상에 대한 결과의 평균을 시스템 성능값으로 사용한다.

그림 1에서와 같이 하나의 동영상은 여러 개의 하이라이트 구역을 가진다. 한 하이라이트 구역은 여러 개의 구간 윈도우로 구성된다. 본 논문에서는 정답 하이라이트에 포함되는 구간 윈도우 중에서 하나 이상을 시스템에서 하이라이트로 판별하면 해당 하이라이트 구간을 탐지했다고 보고, 동영상의 N개의 하이라이트 구역 중 n개를 시스템에서 탐지한 경우에 값 n/N을 추출정확률(Correctly Detected Ratio, CDratio)으로 정의하고 이를 사용하여 시스템의 성능을 평가한다.

표 4는 실험 결과를 보인다. 표에서는 구간 윈도우의 길이 |W|로 3초와 5초를 사용한 경우 각각에 대해서, (A) 채팅의 단순 빈도 정보(onlyFreq, 3-1절의 자질 ①)만 사용하는 경우, (B) (A)에 언어 빈도 정보(withLang, 3-1절의 자질 ②)를 추가로 사용하는 경우, (C) (B)에 긍정 점수와 부정 점수(P/N scores, 3-1절의 자질 ③-1과 ③-2)를 추가로 사용하는 경우, (D) (B)에 문서 벡터(Doc2Vec, 3-1절의 자질 ③-3)를 추가하여 사용하는 경우를 비교한다. 결과에서는 기존 방식과 같이 채팅 빈도만을 이용하는 (A)에서는 약 57%의 추출정확률(CDratio)을 보였다. 언어정보를 동시에 이용한 (B)와 (C), (D)에서 75% 이상의 추출정확률을 얻어, 언어 정보를 사용하여 성능을 크게 향상시킬 수 있음을 확인하였다. 결과에서는 윈도우 길이가 짧은 3초 길이의 구간 윈도우에서 채팅 빈도, 언어 빈도와 함께 긍정 부정 점수를 사용한 (C)에서 87.44%의 가장 높은 추출정확률을 얻었다. 5초 크기의 구간 윈도우는 관련성이 떨어지는 정보가 포함되어 성능이 떨어지는 것으로 보인다. 문장집합의 추상화된 벡터를 반영하는 Doc2Vec은 사용한 (D)는 성능 향상에 기여하지 못하는 것으로 나타나, 추가 연구가 필요한 것으로 나타났다.

표 4. 하이라이트로 추출된 동영상 길이 비율과 추출정확률
Table 4. Ratio of highlighted time and correctly detected

	|W|=3 sec.		|W|=5 sec.	
	CDratio	HTratio	CDratio	HTratio
(A) onlyFreq	57.85%	13.89%	56.05%	14.32%
(B) (A)+withLang	82.06%	16.61%	81.61%	18.67%
(C) (B)+P/N scores	87.44%	18.20%	79.82%	16.64%
(D) (B)+Doc2Vec	75.78%	16.92%	76.23%	17.09%

추출정확률은 모든 구간을 하이라이트로 판별하면 100%의 값을 가지는 문제가 있다. 표 4에서 HTratio 는 동영상 전체길이 대비 시스템에서 하이라이트로 탐지한 구간의 길이 비율(Highlighted Time Ratio)를 나타낸다. 이 값은 많은 구간 윈도우를 하이라이트로 판정할수록 큰 값을 가지게 되는데, 이 경우 많은 양의 동영상을 검토해야 하므로 시스템 관리자의 업무 효율이 떨어진다. 따라서 길이 비율은 낮고 추출정확률은 높을수록 하이라이트 자동 탐지가 효율적이다. 결과에서는 (C)가 (A)에 비해 추출정확률을 30% 가량 향상시킨 것에 비해 탐지된 길이 비율 HTratio는 5%만 증가하여 제안한 방식이 효과적으로 하이라이트를 추출한 것으로 나타났다.

동영상의 특징에 따른 추출정확률을 분석하기 위해 동영상별 결과를 표 5에서 보인다. 결과에서는 표 1의 10개 동영상 각

표 5. 각 동영상에 따른 추출정확률과 하이라이트 시간 비율
Table 5. Ratio of correctly detected and highlighted time each Video

#	Avg. Chat Count	Highlight		CDratio	HTratio
		Ratio	Count		
1	2.44	6.30%	11	63.64%	5.67%
2	15.41	6.29%	55	100.00%	23.02%
3	0.90	2.72%	17	58.82%	9.02%
4	3.28	3.53%	20	95.00%	20.35%
5	0.71	2.91%	21	57.14%	11.20%
6	0.57	6.36%	8	87.50%	3.33%
7	11.86	5.28%	13	100.00%	27.60%
8	3.06	5.38%	20	80.00%	14.12%
9	2.07	3.00%	21	90.48%	13.38%
10	12.72	5.90%	37	100.00%	55.07%

각에 대해 3초 구간 윈도우에 학습 자질로 onlyFreq+withLang+P/N scores의 (C)의 방식을 적용했을 때의 추출정확률과 길이 비율을 보인다. 표에서 Avg. Chat Count는 3초 구간 윈도우의 평균 채팅의 개수, Highlight의 Ratio는 전체 동영상 중 하이라이트로 태그된 구간의 시간 비율, Count는 정답 하이라이트 구간의 개수를 나타낸다. 채팅 개수가 많은 동영상일수록 하이라이트 추출에 높은 성능을 보였으며, 채팅 개수가 작은 동영상 #3, #5, #6에 대해서도 12%이하의 동영상을 검토하여 57%이상의 하이라이트를 찾을 수 있는 것으로 나타나, 대용량의 채팅 정보가 제공되지 않더라도 제안한 방식으로 절반 이상의 하이라이트를 검출할 수 있었다.

Ⅴ. 결 론

본 논문에서는 실시간 한글 채팅으로부터 하이라이트를 추출하기 위해 실시간 채팅을 수집하고 분석하여 얻은 언어 자질을 특징을 활용하는 방법을 제안하였다. 한글 채팅 분석을 통해 구어체 말뭉치와의 차별화되는 언어적 특징을 찾을 수 있었으며, 하이라이트 구간에서의 낱글자 사용, 음절 어절 수의 특징을 발견하였다. 이에 기반하여 기존 하이라이트 추출 연구에서 사용하던 단순한 채팅 빈도에 음절, 어절, 한글 낱글자 등의 빈도 정보와 하이라이트와 비하이라이트에서 발생 비율에 차이를 보이는 단어의 내용 정보를 추가적으로 사용할 것을 제안하였다. 기존 연구에서는 대부분 대량의 영어 채팅에 대해 다루고 있어 실험 결과를 직접 비교하기 어렵지만, 한글 채팅에 대한 실험에서는 기존 방식에서 사용하던 채팅량만을 사용하는 경우 얻을 수 있었던 추출정확률 57.85%를 을 87.44%로 향상시켜 제안한 자질의 유용성을 확인할 수 있었다. 또한 본 논문에서 제안한 자질을 기존 연구의 영상 정보나 채팅 빈도의 변동량 자질과 결합한다면 추가적인 성능 향상이 가능할 것으로 기대된다. 더불어 본 논문에서 제안한 다양한 언어자질을 사용한 결과에서 채팅량이 많지 않은 동영상에서도 실용성 있는 결과를 얻어, 국내에서 생산되는 동영상에 대한 하이라이트 추출의 자동화에 기여할 것으로 기대된다. 향후 연구로는 인터넷 채팅에서 나타나는 한글의 특성에 대한 추가적인 분석과 연구를 예정하고 있다.

감사의 글

이 연구는 금오공과대학교 학술연구비로 지원되었음 (2016-104-093)

참고문헌

[1] Jinjun Wang, Changsheng Xu, Engsiong Chng and Qi Tian “Sports highlight detection from keyword sequences using HMM”, *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo*, Taipei, pp. 599-602, 2004.

[2] M.H. Heo , S.Y. Sin, Y.W. Lee, G.H. Lyu, “Creation of Soccer Video Highlight Using The Structural Features of Caption”, *The KIPS Transactions : Part D*, Vol.10, No.4, pp. 671-678, 2003.

[3] D.M. Seo and C.S. Lim, “User Generated Highlight System Based on Real-time Sports Text Scoreboard with Social Media and Live Media”, *Journal Of Korean Institute of Next Generation Computing,* Vol.8, No.6, pp.49-58, 2012.

[4] D.W. Kim, S.Y. Ha and J.H. Lee, “Extraction of Highlights and Search Indexes of Digital Media by Analyzing Online Activity Data”, *Journal of Korea Multimedia Society,* Vol. 19, No. 8, pp.1564-1573, 2016.

[5] Ting Yao, Tao Mei, and Yong Rui. "Highlight detection with pairwise deep ranking for first-person video summarization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 982-990, 2016.

[6] Cheng-Yang Fu, Joon Lee, Mohit Bansal and Alex C. Berg, “Video Highlight Prediction Using Audience Chat Reactions”, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 972–978, 2017.

[7] Eunyul Kim and Gyemin Lee, “Highlight Detection in Personal Broadcasting by Analysing Chat Traffic : Game Contests as a Test Case ”, *Journal of Broadcast Engineering*, Vol.23, No. 2, pp.218-226, 2018.

[8] Hyungyu Nam, “Automatic generation of titled video highlights from mass interaction“, MS thesis. Chungnam National University , 2018.

[9] twitch TV, Available: https://www.twitch.tv/

[10] National Institute of Korean Language, Sejong Corpus. Available: https://ithub.korean.go.kr/user/total/database/corpusManager.do

[11] Python package for Korean natual language processing, Available: https://konlpy-ko.readthedocs.io

[12] Word2Vec, Available : https://en.wikipedia.org/wiki/Word2vec

[13] Quoc Le and Tomas Mikolov. 2014. “Distributed representations of sentences and documents”, *Proceedings of the 31st International Conference on Machine Learning,* pages 1188–1196, Beijing, China. 2014.

[14] Long Short Term Memory, Available : https://en.wikipedia.org/wiki/Long_short-term_memory

최승혁(Seung-Hyuk Choi)

2013년~현재 : 금오공과대학교 컴퓨터소프트웨어공학과(학부생)

※관심분야 : 컴퓨터 통신, 네트워크, 인공 지능 등

박상준(Sang Jun Park)

2013년~현재 : 금오공과대학교 컴퓨터소프트웨어공학과(학부생)

※관심분야 : 게임 클라이언트 개발, 인공 지능 등

한재현(Jae-Hyun Han)

2013년~현재 : 금오공과대학교 컴퓨터소프트웨어공학과(학부생)

※관심분야 : 인공지능, 알고리즘 등

이현아(Hyun Ah Lee)

1998년 : KAIST 전산학과 대학원 (공학석사)
2004년 : KAIST 전산학과 대학원 (공학박사)
2000년~2004년 : ㈜다음소프트 언어처리연구소
2004년~현재 : 금오공과대학교 컴퓨터소프트웨어공학과 교수
※관심분야 : 자연언어처리, 텍스트데이터 마이닝 등