

Character-Net을 이용한 주요배역 추출

Major Character Extraction using Character-Net

박 승 보* 김 유 원** 조 근 식***
Seung-Bo Park Yoo-Won Kim Geun-Sik Jo

요 약

본 논문에서는 동영상의 등장인물 간의 상황을 기초로 배역간의 관계를 정의한 Character-Net을 구축하는 방법과 이를 이용하여 동영상으로부터 주요배역을 추출하는 방법을 제안한다. 인터넷의 발전과 함께 디지털화된 동영상의 수가 기하급수적으로 증가하여 왔고 원하는 동영상을 검색하거나 축약하기 위해 동영상으로부터 의미정보를 추출하려는 다양한 시도가 있어 왔다. 상업용 영화나 TV 드라마와 같이 이야기 구조를 가진 대부분의 동영상은 그 속에 존재하는 등장인물들에 의해 이야기 전개가 이루어지게 되므로, 동영상 분석을 위해 인물 간의 관계와 상황을 체계적으로 정리하고 주요배역을 추출하여 동영상 검색이나 축약을 위한 정보로 활용할 필요가 있다. Character-Net은 영상의 그룹 단위에 등장하는 인물들을 찾아 화자와 청자를 분류하여 등장인물 기반의 그래프로 표현하고 이 그래프를 누적하여 전체 동영상의 등장인물들 간의 관계를 묘사한 네트워크다. 그리고 이 네트워크에서 연결정도 중심성 분석을 통해 주요배역을 추출할 수 있다. 이를 위해 본 논문에서는 Character-Net을 구축하고 주요배역을 추출하는 실험을 진행 하였다.

ABSTRACT

In this paper, we propose a novel method of analyzing video and representing the relationship among characters based on their contexts in the video sequences, namely Character-Net. As a huge amount of video contents is generated even in a single day, the searching and summarizing technologies of the contents have also been issued. Thereby, a number of researches have been proposed related to extracting semantic information of video or scenes. Generally stories of video, such as TV serial or commercial movies, are made progress with characters. Accordingly, the relationship between the characters and their contexts should be identified to summarize video. To deal with these issues, we propose Character-Net supporting the extraction of major characters in video. We first identify characters appeared in a group of video shots and subsequently extract the speaker and listeners in the shots. Finally, the characters are represented by a form of a network with graphs presenting the relationship among them. We present empirical experiments to demonstrate Character-Net and evaluate performance of extracting major characters.

☞ KeyWords : 캐릭터네트(Character-Net), 상황정보(Context), 얼굴인식(Face recognition), 주요배역(Major character), 사회 연결망(Social network)

1. 서 론

본 논문은 등장인물(Character) 간의 상황정보를 기초로 하여 동영상(Video) 내의 배역간의 관계를

정의할 수 있는 Character-Net의 개념을 소개하고, Character-Net을 구축하는 방법과 Character-Net으로부터 주요배역을 추출하는 방법을 설명한다. 인터넷과 디지털 영상장비 시장의 성장과 더불어 인터넷상에는 다양하고 무수히 많은 동영상이 존재하게 되었다. 또한 지금도 상당한 양의 새로운 동영상이 제작되고 있으며 대중매체의 영화와 TV 드라마들이 인터넷을 통해 제공되고 공유되고 있어 방대한 양의 콘텐츠들로부터 원하는 동영상이나 장면을 검색하는 것은 쉽지 않은 작업이다. 이러한 동영상 검색과 축약 및 브라우징을

* 정 회 원 : 인하대학교 정보공학과 박사과정
molaal@eslab.inha.ac.kr

** 준 회 원 : 인하대학교 정보공학과 박사과정
yoowon@eslab.inha.ac.kr

*** 정 회 원 : 인하대학교 컴퓨터정보공학과 교수
gsjo@inha.ac.kr

[2009/05/15 투고 - 2009/05/24 심사 - 2009/07/31 심사완료]

위해 동영상이나 장면에 대해 사물이나 인물에 대한 정보를 가지고 있고 사물이나 인물에 대한 질의를 처리할 수 있는 시스템이 필요하다. 이를 위해 장면 내용을 의미 기반으로 표현하는 연구가 필요하며 다양한 연구들이 진행되고 있다.

현재 의미 기반의 표현 방식은, 사물이나 인물 등의 개체 특징에 대한 표현 방법[1,2,3]과 개체간의 관계를 표현하는 방법[4] 그리고 개체의 특징과 관계를 혼합해서 표현하는 방법[5,6] 이렇게 세 가지 분야에 걸쳐서 다양한 형태의 연구가 이루어지고 있다. 하지만 개체의 특징만을 표현하거나 관계만을 표현하는 방법보다는 두 가지를 혼합하여 장면의 내용을 표현하는 방법이, 동영상에서 원하는 정보를 추출할 때 더 좋은 효율을 보이게 될 것이다. 이를 위해 본 논문에서는 동영상의 등장인물들을 추출하고 등장인물들 간의 대화 형태를 파악하여 대화 형태를 관계로 표현하는 방법을 제안한다. 이와 유사한 연구로 Roth의 방법을 뽑을 수 있으나 Roth의 연구는 장면 단위로 시맨틱(Semantic)을 표현하며 동영상 전체를 누적하여 표현할 경우 개체 간의 관계 표현만 가능한 방법이다[6]. 이 방법은 개체 간의 관계 중요도를 표현하는 것과 장면의 다양한 개체와 개체의 상황을 모두 표현하고자 할 경우에 자동화하기 어려운 방법이다.

따라서 본 논문에서는 동영상에 포함된 등장인물을 자동으로 인식한 후 등장인물 간의 대화 형태를 화자 인식 기술을 이용하여 자동으로 파악하여 사회연결망(Social Network) 형태로 표현하는 Character-Net 방법론을 제안한다. 대부분의 동영상 이야기 전개는 중심축은 인물과 인물들의 관계로 표현된다. 얼굴인식과 화자 인식기술 및 자막-대본 매칭 기술[3]을 통하여 동영상에 등장하는 인물과 배역명이 무엇인지 파악할 수 있고 배역들 간의 대화 상황 또한 파악할 수 있다. 이렇게 파악된 배역명과 배역들 간의 대화 상황을 노드(Node)와 라인(Line)으로 조합된 Character-Net으로 표현하여 동영상의 주요 의미를 표현할 수 있

다. 또한 Character-Net으로 표현된 동영상에서 연결정도 분석을 통해 주연 배우를 추출할 수 있다. 주요배역 정보와 Character-Net 표현 정보는 주요 장면을 추출하는 데 사용될 수 있으며 나아가 동영상 검색과 축약 및 동영상 브라우징 등의 분야에 활용될 수 있다.

위에서 제안한 방법론을 설명하기 위해 본 논문을 다음과 같이 구성 하였다. 관련 연구들을 통해 Character-Net의 필요성을 설명하였으며, 본론에서 Character-Net의 개념과 구축 방법 및 주요배역 추출 방법을 기술하였다. 다음으로 동영상에서 Character-Net의 구현과 분석을 통해 동영상의 주요배역을 추출하는 실험을 진행하여 Character-Net이 효과적인 방법임을 보였다. 마지막으로 Character-Net의 의미를 결론에서 서술하고 향후 연구 과제에 대해 제시하였다.

2. 관련 연구

2.1 의미 기반 표현 방식

동영상의 내용을 표현하는 의미 기반 표현 방식에 대한 다양한 연구들이 제시되었고 진행되어 왔다. 내용을 표현하기 위해선 영상에 나오는 사물이나 인물과 같은 개체를 검출 혹은 인식할 수 있어야 한다. 그리고 검출 혹은 인식된 개체간의 상황을 파악할 필요가 있다. 이에 대한 연구 방향은 크게 세 가지 분야로 나타나고 있다. 영상에 나타나는 인물과 사물에 대한 개체를 파악하려는 분야와 개체들 간의 상황을 파악하려는 연구 분야가 있다. 그리고 개체 인식과 개체들 간의 상황을 함께 파악하여 표현하려는 분야가 있다.

처음의 동영상에 나타나는 개체를 파악하려는 분야에는 인물(얼굴, 보행자, 운동선수 등)[7-9], 로고[1], 수화[10] 등의 다양한 개체를 식별하려는 시도가 있다. 동영상에 포함된 개체들을 식별하여 동영상의 종류가 무엇인지 알아내기도 하며 동영상의 개체에 대한 정보를 주석처리 하려는 시도도 있다. 개체를 파악하려는 연구는 의미 기반의

표현을 위한 가장 초석이 되는 연구라 할 수 있다. 개체 정보로부터 영상을 분류하는 직접적인 정보를 추출할 수도 있지만 내용 표현의 기본 단위로 활용할 수도 있다.

두 번째로 언급된 개체들 간의 상황을 파악하려는 연구 분야에는 인식된 개체들 간에 발생하는 이벤트(Event)를 정의하여 상황을 표현하려는 연구[4]들이 있다. 이벤트는 동영상에 나타나는 차나 사람과 같은 개체들의 움직임이 발생시키는 상황을 지칭하며, Liang Lin 등은 그들의 연구[4]에서 이벤트를 And-Or 그래프로 표현하였다. 그래프를 통해 개체들이 stay, stop, move, death, birth와 같은 상황에 있는 것을 표현할 수 있다.

세 번째의 개체인식과 개체들 간의 상황을 함께 파악하여 표현하려는 분야에는 영상에 나타나는 개체를 인식하고 개체들이 조합되어 발생시키는 상황을 함께 고려하는 연구가 있다[5,6]. 동영상에 등장하는 인물들 간의 대화 상황을 파악하기 위하여 얼굴의 입이 움직이는 인물을 화자로 선택하고 나머지 사람을 청자로 선택하여 인물들 간의 대화 상황을 파악하는 연구가 있고[5], 장면에 나타나는 개체와 개체들의 상황을 시맨틱 네트워크로 표현하려는 연구도 있다[6]. 동영상에 나타나는 인물과 사물 및 폭발 장면과 같은 특수 효과 등을 찾은 후에 인물들이 춤을 춘다든지 기기를 동작 시킨다는 상황을 찾아내어 장면을 시맨틱 네트워크로 표현하였다. 하지만 개체와 개체 간의 상황을 수동으로 사용자가 지정하여야 하는 문제점이 있다. Roth의 연구는 동영상 검색을 위한 장면 표현 방식과 구조만을 제안한 것으로 장면들이 누적된 동영상 전체에 대한 묘사 방식은 언급되지 않았다.

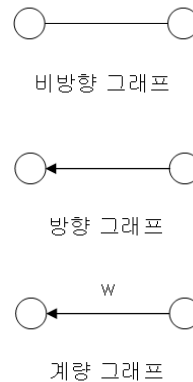
또한, 멀티미디어 콘텐츠를 효율적으로 색인하고 검색하기 위하여 여러 가지 메타 데이터 정의와 디스크립션(Description) 절차를 표준화시킨 멀티미디어 콘텐츠 디스크립션 인터페이스(Multimedia Contents Description Interface)인 MPEG-7[15]과 텍스트 기반의 멀티미디어 데이터

검색에 필요한 질의(Query)와 도메인에서 엔티티의 관계에 대한 문법적 표현에 이용되는 시맨틱 그래머(Semantic Grammar)의 관련 연구[16,17]를 찾아 볼 수 있다.

2.2 사회 연결망(Social network)

사회연결망은 행위자 간의 관계를 노드(Node)와 라인(Line, Tie)으로 표현하여 네트워크 형태의 그래프로 표현한 것이다[11]. 본 연구에서도 동영상의 의미를 표현하기 위해 사회 연결망 표현을 이용하였다.

사회 연결망은 그래프 유형에 의해 그림 1처럼 비방향 그래프(undirected graph), 방향그래프(directed graph), 계량 그래프(valued graph)의 3가지 유형으로 나누어진다[11].



(그림 1) 그래프의 3가지 유형

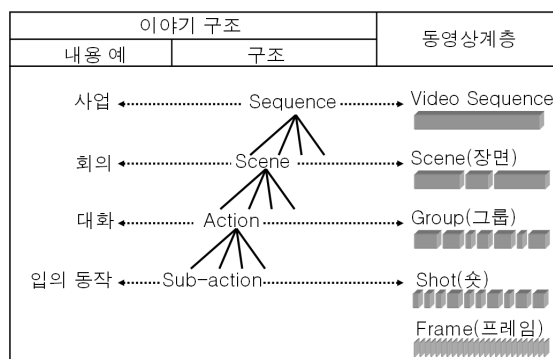
그래프 유형의 선택은 분석 대상의 관계의 형태에 따라 이루어질 수 있다. 관계의 유무에 초점을 맞추면 라인의 방향성이 없는 비방향 그래프를 선택하고, 관계행위의 방향에 초점을 맞추면 라인의 방향성이 있는 방향 그래프를 선택하고, 관계 행위를 정량화하여 표현하는 경우에는 라인에 가중치가 부여된 계량 그래프를 선택할 수 있다. 본 연구에서 제안하는 Character-Net의 경우 대화의 양에 따라 관계가 정의 되므로 3번째의 계량

그래프를 선택 하였다.

사회 연결망에서 중요한 정보를 추출하기 위해 서 아래와 같은 기본 개념들이 있다[11].

- 연결정도(Degree) : 이웃인 모든 노드들의 합
- 인디그리(In-degree) : 어떤 노드가 다른 노드 에게 화살표를 받는 관계의 정도
- 아웃디그리(Out-degree) : 어떤 노드가 다른 노 드에게 화살표를 주는 관계의 정도
- 밀도(Density) : 노드들 간의 연결된 정도로 네트 워크에서 노드들 간 라인의 많고 적음의 정도
- 중심성(Centrality) : 네트워크에서 중심에 위치 하는 정도
- 서브네트워크(Sub-network 혹은 Sub-group) : 네트워크 내 부분집단

2.3 이야기 구조와 동영상 계층 구조



(그림 2) 이야기 구조와 동영상 계층

동영상의 구성은 그림 2의 우측과 같이 5가지 계층으로 이루어진다. 이 계층들은 가장 기본 단 위인 정지된 사진과 같은 필름 한 장에 해당하는 프레임(frame)과 일련의 프레임의 연합으로 장소 나 시간의 끊어짐 없이 촬영된 숏(shot)과 장면의 의미적 기본 구성이 되는 그룹(group)과 동영상의 이야기 의미 단위가 되는 장면(scene)으로 구성된 다. 장면은 시간적으로 이웃된 숏들의 모임으로 정의할 수 있다[12]. 그리고 장면들이 모여서 하나 의 동영상(video sequence)을 이루게 된다.

그림 2의 이야기 구조의 내용 예에서 ‘사업’이

라는 이야기는 구조의 ‘Sequence’에 해당하며 한 개의 동영상으로 표현된다. ‘Sequence’는 따라서 구성이 완성된 하나의 이야기라고 할 수 있다. 구조상의 ‘Sequence’는 다수의 ‘Scene’으로 구성되는데 이야기 내용 예에서 ‘사업’을 위한 여러 번의 ‘회의’ 중에 한 번의 ‘회의’와 관련된 내용이 ‘Scene’에 해당된다. 내용 예의 ‘회의’를 표현하는 동영상의 ‘장면’은 동일 공간에서 회의 장면을 표현하게 되며 시간과 공간적으로 연속적인 배치를 보이게 된다. 동영상에서 ‘회의’을 표현하기 위해 여러 등장인물들의 대화를 촬영하게 되고 등장인 물의 감정과 의도를 표현하기 위해 세밀한 미장센(Mise-end-scene)과 적절한 몽타주(Montage)를 사용할 수도 있다. 이렇듯 동영상의 장면은 등장 인물과 회의 분위기를 표현하기 위해 여러 번의 촬영을 하게 된다. 이렇듯 그림 2에서 표현되는 한 개의 ‘장면’은 등장인물과 등장인물의 행동 및 감정 상태를 이용하여 표현된다[13]. 등장인물은 ‘숏’만으로도 표현이 가능하지만 행동이나 감정 상태는 등장인물 간의 관계 속에서 표현되므로 롱 컷(long cut)이나 ‘숏’들의 ‘그룹’으로 표현하는 것이 적합하다. 따라서 영화나 드라마의 제작이나 편집 시에 특히 ‘대화’와 같은 인물의 행동을 여 러 개의 ‘숏’으로 나누어 표현한다. 따라서 영상의 편집이 아닌 영상의 분석에 있어서도 의미적인 표현을 위해 ‘숏’들의 연계인 ‘그룹’으로 묶은 후 에 분석할 필요가 있다. 본 논문에서는 등장인물 간의 대화 상황을 파악하기 위해 ‘그룹’에 대한 정의를 화자의 말하기의 시작과 끝 시점으로 경 계 지어 구분하였다.

3. Character-Net에 의한 의미표현 과 주요배역 추출

Character-Net은 장면의 인물 정보와 인물간의 대화상황을 표현하는 방식으로 장면을 누적하여 전체 동영상에 대한 등장인물 간의 관계 또한 표 현할 수 있는 방식이다. Character-Net은 다음에 기

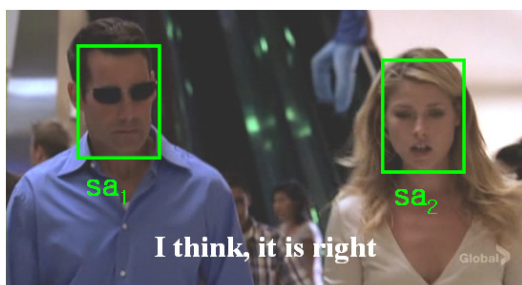
술된 과정을 거쳐서 생성된다.

3.1 등장인물 표현

3.1.1 인물 표현식

췌에 나타나는 등장인물은 이야기 구조에서 가장 기본이 되는 요소이다. 췌에 나타나는 등장인물은 아래 식 1처럼 표현할 수 있다.

$$Shot_a = \{sa_1, sa_2, \dots, sa_n\} \quad (1)$$



(그림 3) 등장인물이 있는 췌

예를 들어 그림3과 같은 췌가 있다고 하면 등장인물이 2 명이므로 췌의 등장인물 표현식은 식 2와 같이 표현된다.

$$Shot_1 = \{sa_1, sa_2\} \quad (2)$$

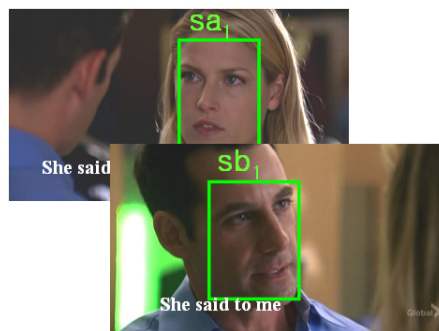
그림 2의 이야기 내용의 대화를 췌 1개로 완결하여 표현하는 경우는 드물다. 췌은 그림 2에 표현된 것처럼 대화라는 ‘Action’을 이루는 얼굴 표정이나 손동작, 몸의 움직임과 같은 Sub-action을 표현하는 수단으로 사용된다. 즉 1개의 대화를 구성하기 위해 1 - 3개의 췌를 사용한다. 대화와 같은 하나의 ‘Action’을 구성하는 여러 개의 췌의 연합을 그룹(group)이라 한다.

췌의 등장인물의 인식은 얼굴인식 기술을 이용하여 판정되며 등장인물의 배역명은 자막이나 대본 등을 통해 알아낼 수 있다[2,3,14]. 등장인물의 배역명이 인식되면 등장인물 중에 화자와 청자를

구별하여야 한다. 등장인물 중에 화자는 말을 하는 사람을 지칭하며 화자의 말을 듣는 사람이 청자가 된다. 췌에서 화자는 입의 움직임을 통해 알아낼 수 있다. 췌 내에서 음성이나 자막이 존재하면서 검출된 얼굴영역의 입 부위의 움직임이 감지된다면 검출된 얼굴은 화자라고 판정된다[3,5]. 화자 검출을 위한 입의 움직임은 프레임 간의 입 주변 영역의 히스토그램 값의 변화[5]나 입의 특징점의 위치 변화[3]를 통해 감지할 수 있다.

그룹에 등장하는 화자와 청자가 1 개의 췌에 동시에 나오기도 하며 화자와 청자를 번갈아가며 여러 개의 췌으로 분할하여 표현하기도 한다. 따라서 그룹 내의 췌들은 등장인물들이 중복적으로 나오기도 하므로 그룹으로 췌의 등장인물을 연합시킬 경우 단순하게 합을 할 수는 없다. 중복적인 인물은 한 명으로 통합하고 등장인물을 합(\oplus)하여 표현한다. 그룹의 등장인물을 식으로 표현하면 식 3과 같다.

$$Shot_a \oplus Shot_b = Group_i = \{a_1, a_2, \dots, a_n\} \quad (3)$$



(그림 4) 췌의 연합인 그룹

예를 들어 그림 4와 같은 경우 1개의 대화가 2 개의 췌으로 구성되었으면 따라서 2개의 췌으로 구성된 그룹이 된다. 뒤쪽의 췌의 경우 등장인물이 sa_1 한명이 되며 앞쪽의 췌은 sb_1 한명이 된다. 따라서 그룹의 등장인물 표현은 각 췌들의 등장인물의 연합인 sa_1, sb_1 이 되며 식 4와 같이 표현

된다. 이때 sa_1 은 a_1 으로 sb_1 은 a_2 로 그룹에서 표현 된다.

$$Group_1 = \{a_1, a_2\} \quad (4)$$

그룹의 화자와 청자는 숫에서 인식한 화자와 청자를 그대로 사용하게 된다. 즉 입이 움직인 등장인물을 화자로 하고 그렇지 않은 등장인물을 청자로 인식한다. 그리고 여러 숫에서 반복적으로 나타나는 동일한 등장인물은 얼굴인식 기술을 이용하여 숫의 연합 시에 1명으로 줄여서 취급한다.

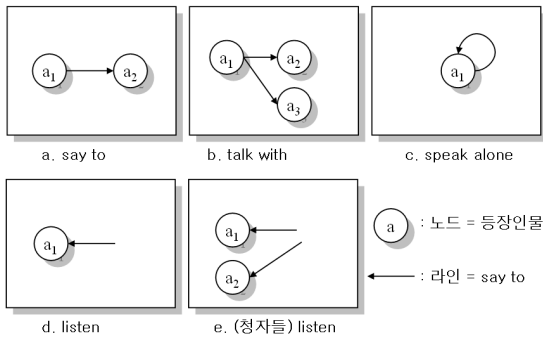
3.1.2 등장인물 기반의 그래프

그룹의 등장인물들이 정해지고 화자와 청자가 결정되면 대화라는 행동이 정의될 수 있다. 대화 행동은 등장인물과 화자와 청자의 판정 결과에 따라 표 1과 같이 5가지로 분류할 수 있다.

(표 1) 대화 행동의 종류

인물 수	화자 수	청자 수	대화행동
2	1	1	(화자) say to (청자)
≥ 3	1	≥ 2	(화자) talk with (청자들)
1	1	0	(화자) speak alone
1	0	1	(청자) listen
≥ 2	0	≥ 2	(청자들) listen

이렇게 분류된 표 1의 대화행동을 등장인물 간의 그래프로 그려보면 그림 5와 같다.



(그림 5) 등장인물기반의 그래프

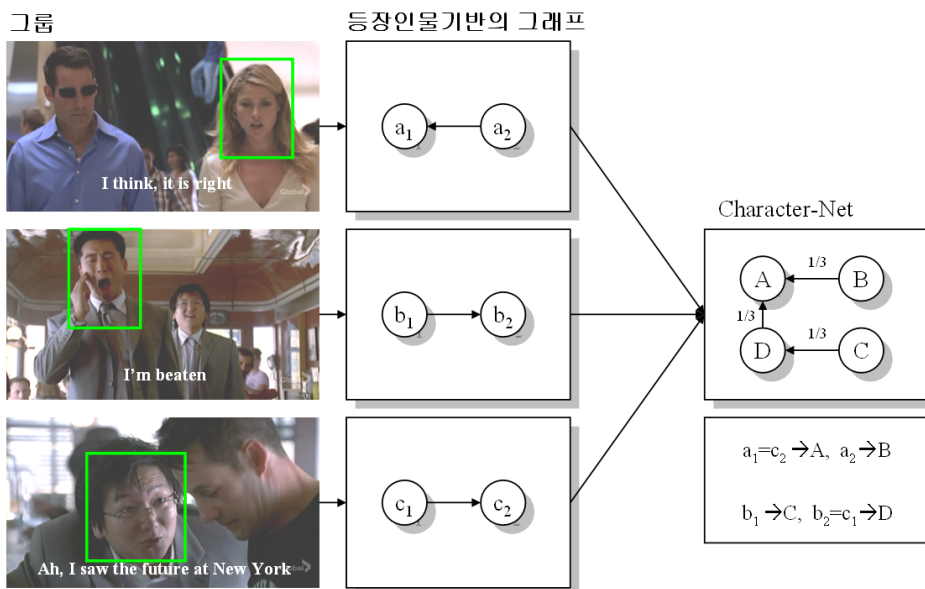
그림 5는 표 1의 대화행동의 종류를 토대로 한 등장인물기반의 그래프로 Character-Net의 가장 기본이 되는 그래프다. 그래프의 동그라미는 노드 (node)를 나타내며 그룹에서 검출된 등장인물이 노드로 표현된다. 노드간의 라인은 대화(say to)를 의미하며 그림 1의 그래프 유형 중에 방향성 그래프를 사용하였다. 방향성 그래프를 선택한 이유는 대화의 방향이 화자에서 청자로 향하기 때문이다. 따라서 라인의 시작 노드가 화자가 되며 화살표가 도달하는 노드가 청자가 된다. 그림 5의 a(say to)는 화자(a_1)와 청자(a_2)가 각각 1명씩 검출된 상황으로 화자가 청자에게 말하는 상황이다. 식 4의 경우가 여기에 해당한다. b(talk with)는 화자 1명 (a_1)과 청자 다수(a_2, a_3)가 검출된 상황으로 화자가 화자 이외의 다수에게 말하는 상황으로 연설장면이나 회의 장면이 여기에 해당한다. c(speak alone)는 화자 1인(a_1)만 검출된 상황으로 독백의 경우가 여기에 해당한다. d(listen)는 화자 검출 없이 청자 (a_1)만 나타난 경우로 문밖에서 기다리던 사람이나 전화를 받고 있는 사람이 여기에 해당한다. e ((청자들) listen)는 화자 검출 없이 다수의 청자(a_1, a_2)가 화면에 나타나는 경우로 연설을 듣는 상황 등이 여기에 해당한다.

3.2 Character-Net

3.2.1 Character-Net의 생성과 확장

그룹의 등장인물 기반의 그래프가 그려지면 이를 누적하여 장면이나 동영상에 대한 그래프를 그리게 된다. 그룹의 누적 시에 다른 그룹에 나타나는 동일 인물은 얼굴 인식을 이용하여 동일한지 아닌지를 판단한 후 동일하다면 한사람으로 취급하게 된다[3,14]. 그림 5의 그룹들을 누적시킨 장면에 대한 표현은 그림 6처럼 표현된다.

그림 6에서 좌측의 이미지에 네모 상자로 표시된 것은 화자를 표시한 것이며, 3.1.2 절에서 설명한 화자와 청자 사이의 대화 상황을 판단하여 그림 6의 중간에 보이는 등장인물 기반의 그래프로

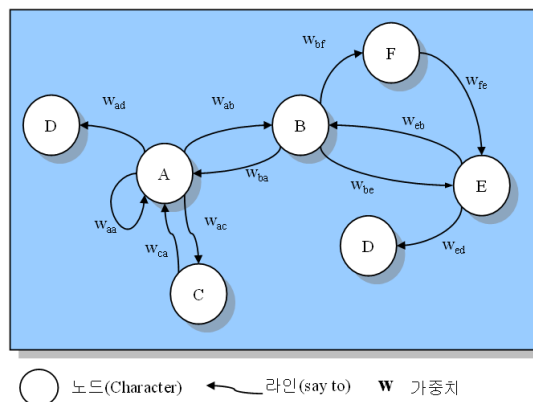


(그림 6) Character-Net의 생성

표현할 수 있다. 이어서 등장인물 기반의 그래프들을 누적하여 우측의 **Character-Net**을 생성할 수 있다. 이때 등장인물 그래프에 표현된 노드들 중에 얼굴 인식을 통해 a_1 과 c_2 , b_2 와 c_1 이 각각 동일 인물을 파악하여 **A**와 **D**로 통합 표현하고 나머지 인물 a_2 는 **B**로 b_1 은 **C**로 표현한다. 그래프들을 누적하면서 대화의 양이 많은 등장인물 간에는 높은 가중치를 라인에 할당한다. 그림 6의 경우 이해를 쉽게 하기 위해 가중치를 대화의 빈도수로 설정하였다. 이외에도 대화의 시간량을 사용해서 설정할 수도 있다. 이 가중치를 설정하는 방법은 다음의 3.2.3 절에서 다룬다. 따라서 **Character-Net**에 사용되는 그래프의 유형은 그림 1의 계량 그래프의 형태가 된다. 새로운 그룹이 추가 되면 그룹을 등장인물 기반의 그래프로 변환 한 후 **Character-Net**에 추가한다. 이런 과정을 거쳐 동영상의 등장인물 간 네트워크를 확장시켜 나가게 된다.

3.2.2 동영상의 Character-Net

3.2.1 절의 **Character-Net**의 생성과 확장 과정을 거쳐서 완성된 **Character-Net**은 그림 7과 같이 표현될 수 있다. 노드의 등장인물 간에 대화 상황이 있는 경우 방향선이 그려지며 대화 상황이 많은 등장인물 간에는 높은 가중치를 갖게 된다.



(그림 7) Character-Net

그림 7에서 등장인물 간의 대화량인 가중치는 w 로 표현되며 원으로 표현된 노드는 등장인물(Character)을 표시한다. 방향선은 등장인물 간의 대화를 의미하며 대화의 방향은 화살표의 방향과 일치한다. 이 네트워크는 등장인물 간의 대화 상황을 누적시킨 사회 연결망의 형태가 된다. 따라서 연결정도(degree)와 밀도(density)와 중심성(centrality) 등의 개념을 도입하여 다양한 분석을 할 수 있다. 주요배역의 경우 다수의 등장인물과 라인으로 연결되어 있을 것이다. 따라서 연결 정도가 높을 것이며 중심성 또한 높게 나타날 것이다.

3.2.3 가중치의 설정

Character-Net의 가중치는 대화의 양을 표현하는 값이다. 따라서 대화가 자주 발생하는 등장인물 간에는 높은 값이 할당 된다. 가중치는 2가지 기준으로 설정할 수 있다. 첫 번째는 식 5와 같이 인물간의 대화 횟수를 누적시켜 설정하는 횟수 가중치 방식이고, 두 번째는 식 6과 같이 등장인물 간의 대화 시간을 누적시켜 설정하는 시간 가중치 방식이다.

$$w_{ab} = NumTalk_{ab} / TotalNumTalk \quad (5)$$

$$w_{ab} = TalkTime_{ab} / TotalTalkTime \quad (6)$$

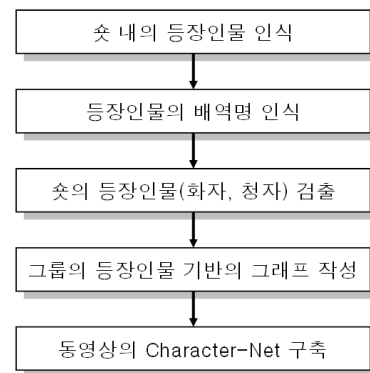
식 5, 6에서 w_{ab} 는 등장인물 a와 b 간의 가중치를 의미한다. 식 5에서 $NumTalk_{ab}$ 는 a, b 간의 대화 누적 횟수를 의미하며 $TotalNumTalk$ 는 동영상 전체의 대화 누적 횟수를 의미한다. 식 6에서 $TalkTime_{ab}$ 는 a, b 간의 대화 누적 시간을 의미하며 $TotalTalkTime$ 은 동영상 전체의 대화 누적 시간을 의미한다.

식 5와 같이 횟수 가중치 방식은 대화 1회가 1개의 정보를 전달한다는 관점에서 대화 횟수가 많을수록 많은 정보가 전달된다는 것을 의미한다

는 면에서 적절하지만 모든 대화의 정도가 동일하지 않기 때문에 단순 누적할 경우 적절한 대화의 양을 대변하지 못하는 단점이 있다. 식 6과 같이 시간 가중치를 사용할 경우 등장인물이 정보를 다른 등장인물과 공유한 시간이 많으면 주요 인물이라는 것을 파악하기엔 적절하지만 인물에 따라 말 빠르기가 다르고 얼마나 많은 정보를 제공하거나 받는지를 대변하지 못한다는 단점이 있다.

3.2.4 Character-Net의 구축 흐름

위에서 설명한 과정을 정리하면 그림 8과 같은 흐름이 된다.



(그림 8) Character-Net 구축 흐름도

우선 동영상의 쏫에 존재하는 등장인물을 인식한다. 그리고 인식된 등장인물의 배역명을 파악하기 위해 자막이나 대본을 이용한다[2,3,14]. 이후에 등장인물 중에 화자와 청자를 구별하기 위해 인식된 얼굴에서 입의 움직임을 파악하여 화자와 청자를 구별하게 된다[3,5,14]. 화자와 청자가 검출되면 의미의 기본 단위인 그룹의 등장인물 기반의 그래프를 작성한다. 등장인물은 노드로 표현되며 대화의 방향이 방향선(directed line)으로 표현된다. 등장인물 기반의 그래프로 표현된 그룹은 누적되어 장면이나 동영상의 Character-Net을 이루게 된다. 이때 대화의 양에 따라 라인은 가중치를

갖게 된다. 가중치는 대화 회수와 대화 시간을 누적하여 설정할 수 있다.

3.3 Character-Net에서 주요배역 추출

동영상에서 주요배역은 다수의 등장인물과 연계 된다. 즉, 주요배역은 다수의 인물과 만나고 많은 시간을 대화하게 되며 Character-Net에서 중심 노드로 표현된다. 따라서 Character-Net에서 연결 정도 중심성 분석을 하게 되면 주요배역을 추출할 수 있게 된다. 각 노드의 연결 정도 중심성(Degree Centrality, DC)은 식 7과 같이 표현된다.

$$DC_a = \sum_{i=1}^n w_{ia} + \sum_{j=1}^m w_{aj} - w_{aa} \quad (7)$$

여기서, n 은 노드의 인디그리의 출발 노드의 개수이고 m 은 노드의 아웃디그리의 도착 노드의 개수이다. DC값은 가중치의 총합인 1보다 작기 때문에 0과 1사이의 실수로 표현된다. 한 노드의 연결 정도 중심성은 다른 노드들과 연결하는 모든 방향선 가중치의 총합으로 식 7과 같이 정의하였다. 식 7의 우측의 첫 번째 항은 노드로 들어오는 방향선의 가중치들의 합인 인디그리의 값을 의미하고 두 번째 항은 노드에서 나가는 방향선의 가중치들의 합인 아웃디그리의 값을 의미한다. 세 번째 항은 자기 루프의 값을 의미한다. 세 번째 항의 자리루프를 빼주는 이유는 인디그리 항과 아웃디그리 항에 자리루프 성분이 중복적으로 적용되어 있기 때문이다. 이렇게 식 7의 3 개의 항에 의해 노드의 연결 정도 중심성이 결정된다. 이렇게 각 노드별로 연결 정도 중심성을 계산하여 일정한 값 이상인 노드를 주요배역으로 결정하게 된다.

4. 실험

앞 절의 등장인물 간의 Character-Net을 구축하기 위해 얼굴인식, 화자인식, 대화 상황정보 파악, Character-Net 표현의 기술 요소들이 필요하다.

Character-Net 구현과 실험은 윈도우즈 XP 운영체제의 PC 기반에서 실행하였다. 구현과 실험을 위한 기술 요소들 중 얼굴인식을 위해서는 SDK 형태로 지원되는 상용 API(VeriLook Face Identification SDK)를 사용하였으며, 입의 움직임을 알아내어 화자인지를 인식하는 부분과 자막 정보로부터 대화시간을 파악하는 부분과 그룹의 대화 상황정보를 파악하는 부분 그리고 그룹에 대한 그래프 표현을 누적하여 Character-Net으로 표현하는 것은 VB 6.0과 Delphi 7.0을 사용하여 구현하였다. 구현된 어플리케이션에 대한 실험은 표1 대화행동의 종류의 각 항목을 대상으로 구현하여 소프트웨어 적용 가능성에 대하여 확인하였고, Character-net을 그리기 위해 실험에 이용된 데이터의 적용은 수작업으로 진행하였다.

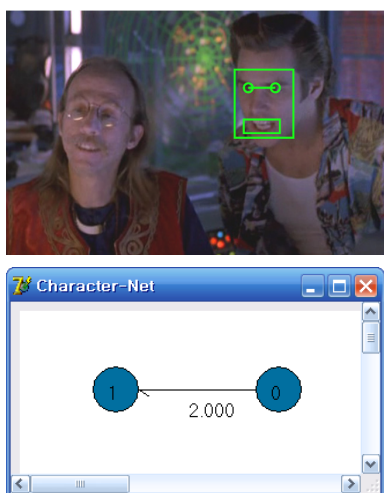
Character-Net의 구축을 위해 동영상 5개를 선택하여 실행하였다. 동영상은 영화중에서 ‘해리가 샬리를 만났을 때(When Harry Met Sally...),’와 ‘에이스 벤추라(Ace Ventura)’를 선택하였고 TV 시리즈로는 ‘히어로즈(Heroes)’ 시즌1의 에피소드 4와 23과 ‘프렌즈(Friends)’ 시즌3의 에피소드 1을 선택하였다.

4.1 등장인물기반의 그래프 추출

Character-Net의 구축을 위해 그룹이나 장면에서 발생하는 대화 상황을 3.1.3 절에 언급된 등장인물기반의 그래프로 표현하여야 한다. 그룹을 이루는 숫들을 정하기 위해 자막 시간을 활용하였다. 자막이 나타나는 시간대역을 화자가 청자에게 이야기하는 대화 상황의 시간대역으로 정의하였고 그 시간대역에 해당하는 숫들을 그룹으로 결정하였다. 일반적으로 1~3개의 숫들이 그룹으로 묶여졌다. 그룹의 등장인물기반의 그래프 추출 실험을 위해 영화 ‘에이스 벤추라’를 사용하였다. 그림 9와 그림 10과 그림 11은 ‘에이스 벤추라’의 일부 그룹들을 등장인물기반의 그래프로 표현한 실험 결과 화면이다.

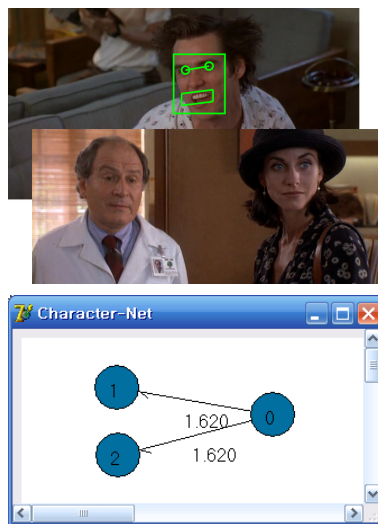
그림 9는 1개의 숫이 1개의 그룹을 이루는 경

우로 화자가 1인이고 청자가 1인인 경우다. 따라서 대화행동은 표 1의 ‘(화자) say to (청자)’가 되며 그림 5의 등장인물 기반의 그래프에서는 a(say to) 그래프로 표현된다. 그림 9에서 우측의 인물의 입 움직임이 검출되어 화자로 인식되었고 좌측의 인물은 청자로 인식되었다. 따라서 그림 9의 그룹에 대한 등장인물기반의 그래프는 0번 노드로 표현된 우측의 인물이 1번 노드로 표현된 좌측의 인물에게 방향선이 연결되는 그래프로 표현 되었다.



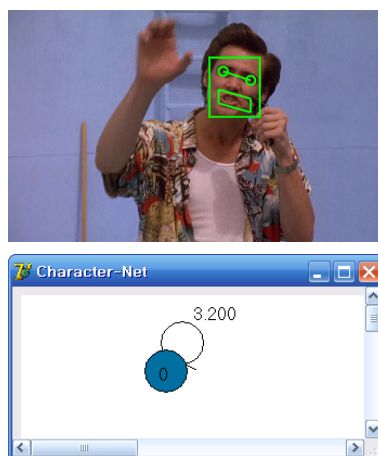
(그림 9) 등장인물기반 그래프의 ‘say to’ 예

그림 10은 2개의 숫이 1개의 그룹으로 이루어지는 경우로 화자는 처음 숫에서 검출된 1인이고 두 번째 숫의 2명은 청자인 경우다. 따라서 대화행동은 표 1의 ‘(화자) talk with (청자들)’이 되며 등장인물기반의 그래프는 그림 5의 b(talk with)로 표현된다. 그리고 그림 10의 처음 숫의 인식된 인물의 입의 움직임이 검출되어 화자로 인식되었으며 두 번째 숫의 2명은 입의 움직임이 없어 청자로 분류되었다. 따라서 그림 10의 밑의 그래프와 같이 0번 노드의 화자가 1번과 2번 노드의 청자 각각에게 방향선으로 연결되었다. 그림 9나 그림 10의 방향선은 값을 가지고 있는데 이 값은 대화가 진행된 자막 시간으로 지정 하였다.



(그림 10) 등장인물기반의 그래프의 ‘talk with’ 예

그림 11은 1개의 숫이 1개의 그룹을 형성한 경우로 화자 혼자서 이야기하는 상황이다. 따라서 대화행동은 표 1의 ‘(화자) speak alone’이 되며 그림 5의 c(speak alone) 그래프로 표현된다. 그림 11의 아래 부분과 같이 0번 노드의 화자 혼자 3.2초 동안 자신에게 말하는 그래프로 표현되었다. 표 1의 나머지 대화행동 역시 화자와 청자를 인식하여 그래프로 표현하였다.



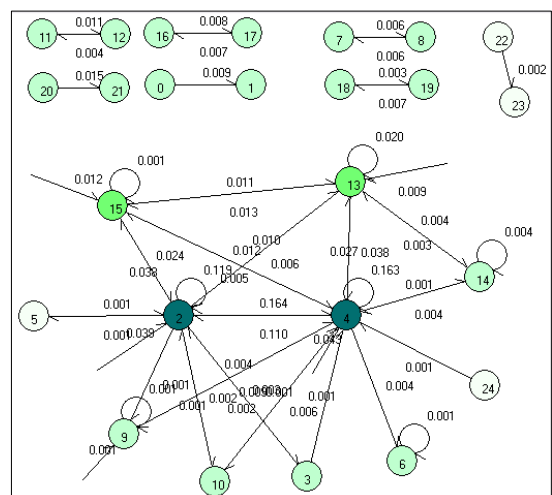
(그림 11) 등장인물기반 그래프의 ‘speak alone’ 예

4.2 Character-Net의 구축과 주요배역 추출

5개의 동영상 ‘해리가 샬리를 만났을 때’와 ‘에이스 벤츄라’와 ‘히어로즈’(2편), ‘프렌즈’에 대한 Character-Net을 각각 그림 12부터 그림 15까지와 같이 표현하였다. 각 그림 모두 등장인물 기반의 그래프를 누적하여 Character-Net을 구축하였으며 노드 간 방향선의 가중치는 식 6의 시간 가중치를 계산하여 사용하였다. 연결정도 중심성(DC)을 계산하여 0.10을 넘는 노드는 주요배역으로 추출하였으며, 0.05를 넘는 노드는 보조배역으로 추출하였다. 전체 대화 시간 중에 10% 이상을 차지하는 등장인물은 비중이 있는 주인공급의 배역이라 판단하였으며, 5 - 10% 정도의 연결정도 중심성을 갖는 노드는 이야기 전개에 도움을 주는 배역이라고 판단하였기 때문이다. 그리고 그 외의 나머지 노드는 비중이 없는 배역으로 고려하여 별도로 의미부여를 하지 않았다. 그리고 추출된 배역 분류 결과가 적절한 지를 검토하였다. 영화의 경우 주요배역이나 보조배역으로 추출된 등장인물을 영화배우, 텔레비전 시리즈 등에 관한 온라인 데이터베이스인 IMDb(The Internet Movie Database)*의 등장인물(Cast) 정보와 일치 여부를 비교하였다. IMDb에 있는 등장인물들의 정렬기준은 엔딩 크레딧(Ending Credit)에 나와 있는 순서이다. 영화의 경우 일반적으로 엔딩 크레딧은 배역의 중요도에 따라 순서가 정해진다. 따라서 제일 위에 있는 배역이 주인공이 되고 밑으로 내려갈수록 배역의 중요도가 낮아지게 된다. 하지만 IMDb의 등장인물의 순서가 TV 시리즈의 경우 배역의 중요도와 상관관계가 없이 정해져있어서 추출된 배역 분류 결과와 비교할 수가 없었다. 그래서 TV 시리즈의 경우 IMDb의 정보 대신에 직접적인 사용자 평가를 진행하였다. 사용자 평가는 10명의 사용자에게 배역별 중요도를 10점 만점으로 평가하게 하였으며, 10명의 평가 결과를 평균하여 배역별로 순위를 정하였다. 이렇게 사용자 평가된

결과와 본 논문에서 평가된 결과를 비교 검토하였다.

첫 번째 실험 동영상인 ‘해리가 샬리를 만났을 때’는 해리(Harry)와 샬리(Sally)라는 남녀 주인공들이 몇 가지 사건을 거치며 서로의 사랑을 확인해 가는 로맨틱 코미디 장르의 영화이다. 해리와 샬리 주변에는 친구가 각각 1명씩 존재하며 이들은 해리와 샬리의 고민을 들어주면서 해리와 샬리의 사랑을 키워가는데 도움을 주는 역할을 한다. 이들의 배역명은 마리(Marie)와 제스(Jess)이다. ‘해리가 샬리를 만났을 때’의 Character-Net을 그려보면 그림 12와 같이 나타나는데, 이때 4번 노드가 해리이고 2번 노드가 샬리로 연결정도 중심성을 계산하면 0.5 이상의 값으로 나타난다. 주인공 주변의 친구들은 15번과 13번 노드로 표현되었으며 0.1 이상의 값을 나타내었다. 이 동영상은 4명간의 주요배역들의 관계에 대한 장면묘사가 주를 이루었으며 중간에 노 부부들(0-1, 7-8, 11-12, 16-17, 18-19, 20-21, 22-23)이 나와서 인터뷰하는 형식의 장면들이 있으나 여러 장면에서 반복적으로 나오지 않아 의미있는 배역으로 판단되진 않았다. 그림 12에서 연결정도 중심성이 0.05를 넘는 노드들을 찾아서 정리하면 표 2와 같이 나타난다.



(그림 12) ‘해리가 샬리를 만났을 때’의 Character-Net

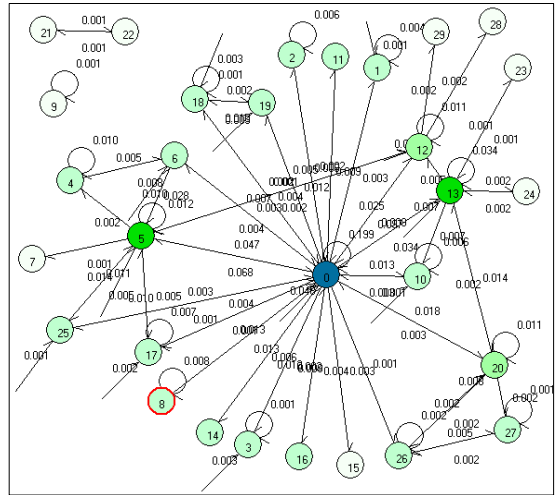
* <http://www.imdb.com/>

그림 12에서 추출된 주요배역들은 4번, 2번, 13번 15번 노드들이고 각각의 등장인물명을 표 2에 표시하였다. 표 2에는 나열된 배역들의 순서는 연결정도 중심성(DC)을 기준으로 하였으며 이 순서는 IMDb의 등장인물 정보의 정렬 순서와 동일한 것을 확인할 수 있다.

(표 2) '해리가 샬리를 만났을 때'의 연결정도 중심성

노드	등장인물	DC	분류결과	IMDb 순서
4	Harry Burns	0.594	주요배역	1
2	Sally Albright	0.519	주요배역	2
13	Marie	0.142	주요배역	3
15	Jess	0.116	주요배역	4

그림 13은 동물 탐정인 주인공이 여러 사람을 만나면서 범행을 일으킨 범인을 찾아내는 내용으로 0번 노드가 'Ace Ventura' 역의 짐캐리이다. 이야기 내용에 어울리게 0번 노드는 거의 모든 노드와 연결되도록 네트워크가 형성되었으며 연결정도 중심성 역시 0.687로 매우 높게 나타났다. 0번 노드 외에 5번 노드(0.216)와 13번 노드(0.204)가 주요배역으로 추출되었다. 5번 노드는 극의 주인공인 'Ace Ventura'의 여자친구 역으로 나오는 배역으로 초반부에서 후반부까지 지속적으로 등장한다. 13번 노드는 범행을 일으킨 범인으로 복수와 납치범 역할을 한다. 그림 13의 경우 0번과 5번과 13번 노드가 이야기 전개의 핵심적 인물로 등장하며 Character-Net에서 보여 지듯이 그들을 중심으로 한 서브네트워크가 존재한다. 그리고 12번과 20번 노드는 이야기 전개를 위한 보조적 인물로 등장 하지만 이야기 전개에 일정한 도움을 주는 인물들이다. 그림 13의 노드들 중에 연결정도 중심성이 0.05를 넘는 노드들을 순서대로 표시하면 표 3과 같다. IMDb의 등장인물 순서와 동일한 순서로 연결정도 중심성이 추출되는 것을 표 3을 통해 알 수 있다.



(그림 13) '에이스 벤츄라'의 Character-Net

그림 13에서 추출된 주요배역들은 0번, 5번, 13번 노드들이고 각각의 등장인물들의 이름을 표 3에 표시하였다. 표 3에 나열된 배역들의 순서는 연결정도 중심성(DC)을 기준으로 내림차순으로 정렬하였다. 0번 노드인 'Ace Ventura'가 여러 사람을 만나면서 사건을 풀어가는 내용이고 그림 13의 네트워크에서 보여지는 것처럼 거의 모든 노드들과 연결되어 있어서 0번 노드의 연결정도 중심성 값이 상당히 높게 나왔다. 영화의 경우 핵심적인 이야기가 1개이고 소수에 의해 이야기가 전개되기 때문에 핵심적인 인물들의 연결정도 중심성이 높게 관측된다. 그리고 연결정도 중심성에 의해 추출된 배역의 중요도 순서가 IMDb의 순서와 동일한 것을 확인하였다.

(표 3) '에이스 벤츄라'의 연결정도 중심성

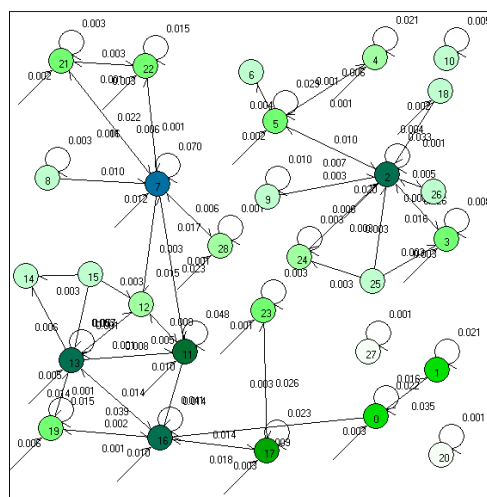
노드	등장인물	DC	분류결과	IMDb 순서
0	Ace Ventura	0.687	주요배역	1
5	Melissa Robinson	0.216	주요배역	2
13	Lt. Lois Einhorn	0.204	주요배역	3
12	Emilio	0.065	보조배역	4
20	Dan Marino	0.059	보조배역	5

그림 14는 TV 시리즈인 ‘히어로즈’ 시즌1의 4번째 에피소드와 23번째 에피소드를 각각 표현한 Character-Net이다. ‘히어로즈’는 초능력을 가진 다양한 인물이 등장하여 주변에서 벌어지는 문제들을 풀어가면서 이야기가 전개되어 가는 내용이다. ‘히어로즈’는 주요배역들이 자신의 주변 인물들과 관계를 맺어가면서 조금씩 성장해 가면서 자신에게 닥친 문제를 하나씩 해결해 나가면서 이야기 전체에 관련된 문제를 해결해 나가는 이야기 전개를 가진다. 따라서 시리즈의 초반에는 주연들 주변의 작은 사건들이 이야기 전개의 동기가 된다. 따라서 시리즈 초반에 해당하는 에피소드 4의 동영상은 주연들이 서로 만나지 않으면서 자신의 주변 인물들과 서브네트워크를 형성한다. 그림 14의 대상이 된 동영상은 시즌1의 4편으로 시리즈의 초반에 해당하는 동영상이다. 따라서 주연들 주변으로 작은 서브네트워크를 형성하는 네트워크 형태를 띤다. 또한 2번 노드 주변의 서브네트워크는 다른 서브네트워크와 연결되지 않은 형태를 띠기도 한다. 아직 이야기 전개상 2번 노드는 다른 주연들로부터 영향을 받지 않는 상태에 있는 것이다. 시리즈 결말인 에피소드 23은 에피소드 4와 달리 모든 서브네트워크들이 연결되어 있다. 서브네트워크들이 연결되어 있다는 것은 모든 주요배역들이 서로를 만난다는 것을 의미하고 이야기가 종결되어질 때 나타나는 Character-Net 형태이다.

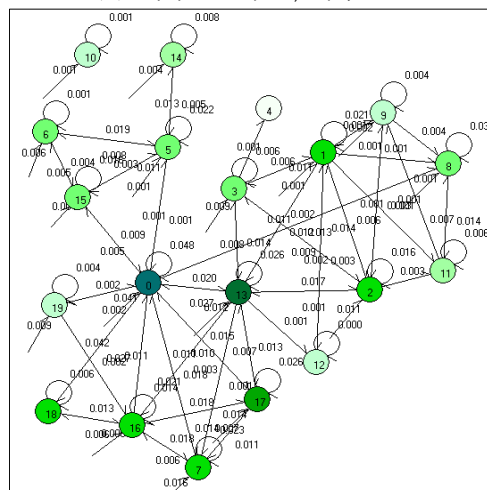
그림 14의 Character-Net에서 연결정도 중심성을 노드별로 계산하여 분류된 결과는 표 4와 같다. 에피소드 4에서는 7번, 13번, 16번, 2번, 11번 노드들이 10%가 넘는 연결정도 중심성을 나타내었고 에피소드 23은 0번, 13번, 17번, 7번, 16번 노드들이 10% 넘는 값을 나타내어 각각의 에피소드의 주요배역으로 추출되었다. 표 4의 나머지 노드들은 5%가 넘는 연결정도 중심성을 나타내어 보조배역으로 추출되었다.

IMDb의 등장인물 순서는 TV 시리즈의 경우 배역의 중요도와 전혀 무관하게 정렬되어 있어서

실험결과와 비교를 할 수가 없다. 따라서 비교를 위한 자료로 ‘히어로즈’를 시청한 시청자들의 평가 결과를 사용하였다. 시청자 평가 결과는 연결정도 중심성 결과와 순위에 있어 차이를 나타내었다. 에피소드 4의 경우 전체 평균 1.6 순위의 차이를 나타내었고 에피소드 23의 경우 평균 2.4 순위 차이를 나타내었다.



(a) ‘히어로즈’ 시즌1, 에피소드 4



(b) ‘히어로즈’ 시즌1, 에피소드 23

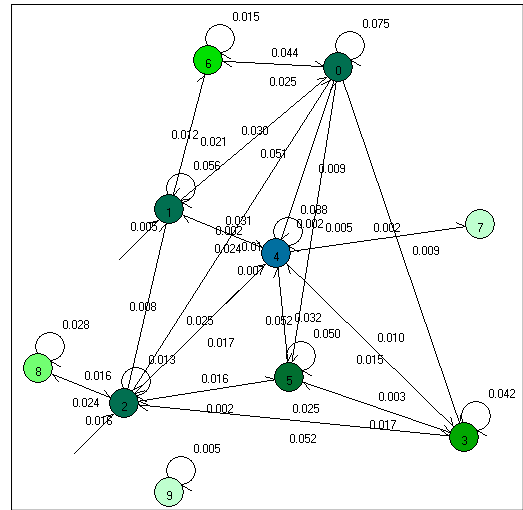
(그림 14) ‘히어로즈’의 Character-Net

(표 4) '히어로즈'의 연결정도 중심성

에피소드 4	등장인물	노드	DC	분류결과	시청자 평가	
					순위	차이
	Niki Sanders	7	0.202	주요배역	5	(-4)
	Mohinder Suresh	13	0.161	주요배역	3	(-1)
	Peter Petrelli	16	0.161	주요배역	1	(+2)
	Claire Bennet	2	0.149	주요배역	2	(+2)
	Nathan Petrelli	11	0.132	주요배역	6	(-1)
	Simone Deveaux	17	0.087	보조배역	8	(-2)
	Hiro Nakamura	0	0.085	보조배역	4	(+3)
	Ando Masahashi	1	0.079	보조배역	7	(+1)
	Brody Mitchum,	3	0.056	보조배역	10	(-1)
	Noah Bennet	5	0.054	보조배역	9	(+1)
	Tina	21	0.050	보조배역	11	(0)
	시청자 평가 순위와 DC 순위의 차이 평균					1.6

에피소드 23	등장인물	노드	DC	분류결과	시청자 평가	
					순위	차이
	Peter Petrelli	0	0.302	주요배역	1	(0)
	Noah Bennet	13	0.186	주요배역	4	(-2)
	Claire Bennet	17	0.135	주요배역	2	(+1)
	Angella Petreli	16	0.123	주요배역	7	(-3)
	Nathan Petrelli	7	0.123	주요배역	2	(+3)
	Molly Walker	2	0.095	보조배역	8	(-2)
	Charles Deveaux	18	0.094	보조배역	11	(-4)
	Mohinder Suresh	1	0.093	보조배역	5	(-3)
	Niki Sanders	8	0.062	보조배역	8	(+1)
	Hiro Nakamura	5	0.084	보조배역	6	(-4)
	Matt Parkman	3	0.068	보조배역	8	(-3)
	시청자 평가 순위와 DC 순위의 차이 평균					2.4

그림 15는 TV 시리즈인 '프렌즈'의 시즌 3의 에피소드 1에 대한 Character-Net을 표현한 것으로 6명의 친구들이 만들어내는 상황을 코믹하게 그린 시트콤이다. 시트콤은 등장인물이 7명 내외로 고정된 등장인물과 동일한 배경을 바탕으로 진행된다. '프렌즈' 역시 이러한 시트콤의 정의를 비슷하게 따르고 있으며 이야기 진행이 6명의 친구들에 의해 이루어진다. 6명 이외에 1~2명이 추가되어 내용의 약간의 변화를 이끌지만 주요한 이야기 전개는 6명의 주연들에 의해 이루어진다. 그림 14의 Character-Net은 6개의 노드들이 골고루 연결되어 네트워크를 형성하고 있다.



(그림 15) '프렌즈'의 시즌 3의 에피소드1의 Character-Net

그림 15의 Character-Net에서 연결정도 중심성을 노드별로 계산하면 표 5와 같이 나타난다. 4번, 0번, 1번 5번, 3번, 6번 노드가 주요배역으로 추출되었으며 8번 노드는 보조배역으로 추출되었다. 시청자 평가 결과는 연결정도 중심성 결과와 순위에 있어 평균 1.6 순위 차이를 나타내었다.

(표 5) '프렌즈'의 연결정도 중심성

노드	등장인물	DC	분류결과	시청자 평가
4	Dr. Ross Geller	0.315	주요배역	3 (-2)
0	Joey Tribbiani	0.268	주요배역	1 (+1)
1	Chandler Bing	0.238	주요배역	4 (-1)
2	Monica Geller	0.233	주요배역	2 (-2)
5	Rachel Green	0.181	주요배역	5 (0)
3	Phoebe Buffay	0.175	주요배역	6 (0)
6	Janice Litman	0.117	주요배역	7 (0)
8	Jack Geller	0.068	보조배역	8 (0)
시청자 평가 순위와 DC 순위의 차이 평균				0.75

4.3 고찰

Character-Net으로 표현된 동영상에서 주요배역

들로 추출되는 노드들이 갖는 특징은 연결정도 중심성(DC)이 높게 나타난다는 것이다. 연결정도 중심성이 갖는 의미는 식 7의 각 항이 의미하는 것처럼 자신이 다른 등장인물에게 대화한 정도와 다른 등장인물로부터 말을 들은 정도와 자기 혼자 말한 정도의 합이다. 즉 다른 등장인물에게 정보를 많이 제공했거나 다른 등장인물로부터 정보를 많이 제공받거나 자기가 가진 정보를 스스로 많이 확인한 등장인물이 연결정도 중심성이 높게 된다. 영화나 TV 시리즈에서 주요배역은 주변인물과 많은 의사소통을 하거나 자기 자신의 심리 상태나 기분을 표현하는 행동이나 말을 많이 하게 된다. 따라서 주요배역은 주변인물과 높은 가중치의 관계를 형성하게 되며 결과적으로 연결정도 중심성이 높게 나타나게 된다. ‘에이스 벤추라’의 경우 주인공은 사건의 해결을 위해 다수의 배역들과 만나게 되며 이들과의 대화를 통해 다양한 정보를 파악하게 된다. 따라서 주인공의 연결정도 중심성은 상당히 높게 측정되었다. ‘해리가 샬리를 만났을 때’의 경우 두 남녀의 사랑이야기를 다루었기 때문에 두명의 남녀가 주인공이 되며 2명의 연결정도 중심성이 상당히 높게 측정되었다. TV 시리즈의 경우 마찬가지로 주인공들의 연결정도 중심성이 높게 측정되었으나 다수의 등장인물이 출연하고 여러 이야기가 동시에 진행되므로 인해 주요배역들 간에 고른 연결정도 중심성이 측정되었다. 하지만 주요배역들은 0.1 이상의 연결정도 중심성을 나타내었으며 시리즈 전체에선 주요배역이나 개별적인 에피소드에서 이야기 전개가 미미할 경우 ‘히어로즈’의 에피소드 4의 ‘Hiro Nakamura’나 ‘Noah Bennet’처럼 낮은 연결정도 중심성 값을 나타내기도 하였다.

추출된 주요배역을 IMDb의 등장인물 순서와 비교하면 영화에 해당하는 그림 12와 그림 13의 경우 동일한 순서로 나타났으나 TV 시리즈인 그림 14와 그림 15의 경우 서로 다른 순서로 나타났다. 이는 TV 시리즈의 경우 엔딩크레딧이 에피소드에 나타나는 배역의 비중의 순서대로 되어있

지 않기 때문이다. TV 시리즈의 경우 배역의 비중보다는 별도의 기준에 의해 엔딩크레딧의 순서가 정해지는 것으로 판단된다. 그래서 TV 시리즈의 경우 시청자의 평가를 통해 등장인물의 중요도를 측정하여 실험하여 얻어진 연결정도 중심성과 비교하였다. 비교 결과 ‘히어로즈’의 에피소드 4는 평균 1.6순위의 차이가 발생하였으며, 에피소드 23은 평균 2.4 순위의 차이가 발생하였고, ‘프렌즈’의 경우 0.75 순위 차이를 발생하였다. 결과는 ‘히어로즈’의 에피소드 23의 경우 평균 2.4 순위 차이로 차이가 크게 나타났지만 전체적으로는 연결정도 중심성에 의한 순위와 시청자 평가에 의한 순위가 비슷하게 나타났다. 시청자 평가와 실험결과의 차이는 시청자들의 등장인물 중요도를 평가하는 기준이 본 연구에서 제안하는 방식과 차이를 갖기 때문이라고 생각된다. 설문에 응한 시청자들에게 평가의 기준을 추가로 질문한 결과 등장인물의 중요도를 판단하기 위해 등장인물의 대화 빈도 외에도 대화의 내용과 이야기 진행에 영향을 미치는 비중을 함께 고려하여 등장인물의 중요도를 판단한다고 하였다. 즉 시청자는 영상에 나타나는 등장인물의 정보뿐만 아니라 음성과 대화의 내용까지 고려하기 때문에 평가결과의 차이를 나타내었다고 판단된다.

5. 결 론

본 논문에서 동영상의 등장인물과 등장인물간의 대화 상황을 토대로 하여 Character-Net을 구축하는 방법을 제안하고 그 효용성을 주요배역 추출방법을 통하여 입증하였다. Character-Net은 네트워크의 형성과 주요배역 추출을 위하여 사회연결망의 개념과 방식을 채용하였다. 이렇게 형성된 Character-Net은 동영상의 중요한 기본 요소인 등장인물과 등장인물 간의 상황이 기초가 된 의미 기반의 표현 방식으로 등장인물과 배역명을 인식하기 위해 얼굴 인식 기술[3,14]과 화자 인식 방법[5] 그리고 대본으로부터 배역명 추출방법[3,14]

을 이용하였다. 그리고 등장인물 간의 대화 상황을 표현하기 위해 그림 5에서 등장인물기반의 그래프 표현 방식을 제안하고 다양한 대화 상황을 정의하였다. 또한 등장인물 기반의 그래프를 누적하여 Character-Net을 그릴 때 방향선에 할당되는 가중치의 계산방식과 주요배역 추출을 위한 연결정도 중심성을 계산하는 식을 함께 제안하였다.

제안한 Character-Net 구축 방법론과 주요배역 추출 방법론을 실험을 통해 증명하였다. 동영상 5개를 Character-Net으로 표현하였고 표현된 Character-Net에서 연결정도 중심성 분석을 통해 주요배역과 보조배역을 추출하였다. 추출된 주요배역은 영화의 경우 각 동영상의 엔딩크레딧상의 주연과 동일한 결과를 얻었다. 하지만 TV 시리즈의 경우에는 엔딩크레딧의 배역순서가 극중의 비중 순서와 상이 하여서 추출된 주요배역 순서와 다른 결과를 나타내었다. 따라서 TV 시리즈의 경우 시청자의 평가와 비교하였으며 실험결과에 의한 주요배역 순서와 유사한 평가 결과를 얻을 수 있었다.

본 논문에서 제안한 Character-Net은 다양한 분야에서 활용될 수 있다. 등장인물이나 등장인물 간의 관계를 가지고 동영상 내용기반의 검색을 하는데 사용할 수 있으며 Character-Net의 구축이 그룹의 등장 인물정보의 누적으로부터 시작되므로 등장인물 위주의 계층적인 브라우징(Browsing)에도 활용될 수 있다. 본 논문의 연구결과를 통하여 관련연구에서 언급한, 콘텐츠의 내용 또는 의미와 관계된 특징을 추출하여 표준화된 디스크립션으로 어노테이션하는 MPEG-7 등의 방법과 접목한다면 표준화된 표현이 가능할 것으로 판단되며, MPEG-7의 표준을 통해 기술된 어노테이션 내용을 분석하여 이야기 구조를 갖는 영화나 드라마와 같은 동영상에서 인물들 간의 대화 이벤트를 기반으로 한 캐릭터 네트워크 구성을 통하여 주요배역을 추출하고 배역들 간의 관계를 모델링하는 연구를 진행할 수 있을 것이다. 또한 현재는 등장인물간의 상황을 대화의 방향만으로 국한하였으

나 대화의 의미를 알아내어 추가한다든지 총소리와 같은 음향효과로 인한 특수한 상황을 추가적으로 표현[18]한다면 더욱 다양한 형태의 의미도 표현할 수 있을 것이다.

참 고 문 헌

- [1] J.R. Cózar, N. Guil, J.M. González-Linares, E.L. Zapata, E. Izquierdo, "Logotype detection to support semantic-based video annotation," *Signal Processing: Image Communication*, Volume 22, Issues 7-8, pp. 669-679, Aug.-Sep. 2007.
- [2] J. Yang, R. Yan, A.G. Hauptmann, "Multiple instance learning for labeling faces in broadcasting news video," in: *Proceedings of the ACM International Conference on Multimedia*, pp. 31 - 40, 2005.
- [3] M. Everingham, J. Sivic, A. Zisserman, "Hello! My name is... Buffy" - automatic naming of characters in TV video," *Proceedings of the 17th British Machine Vision Conference*, Edinburgh, UK, pp. 889 - 908, 2006.
- [4] L. Liang, G. Haifeng, L. Li, and W. Liang. "Semantic event representation and recognition using syntactic attribute graph grammar," *Pattern Recognition Letters*, Vol. 30, Issue 2, pp. 180-186, 15 Jan. 2009.
- [5] 박승보, 김유원, 조근식, "얼굴인식을 이용한 동영상 상황정보 어노테이션," *한국지능정보시스템학회*, 2008 추계 학술대회 논문집, pp.319-324, 2008.11.28.
- [6] V. Roth, "Content-based retrieval from digital video," *Image and Vision Computing*, Vol. 17, no. 7, pp. 531-540, 1999.
- [7] D. Cristinacce and T.F. Cootes, "Feature Detection and Tracking with Constrained Local Models," *Proc. 17th British Machine Vision Conf.*, pp. 929-938, 2006.

- [8] B. Leibe, E. Seemann, B. Schiele, "Pedestrian detection in crowded scenes," *Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 878 - 885, 20-25 June 2005.
- [9] S. H. Khatoonabadi, M. Rahmati, "Automatic soccer players tracking in goal scenes by camera motion elimination," *Image and Vision Computing*, Vol. 27, Issue 4, pp. 469-479, 3 Mar. 2009.
- [10] P. Buehler, M. Everingham, D.P. Huttenlocher, A. Zisserman, "Long Term Arm and Hand Tracking for Continuous Sign Language TV Broadcasts," *Proceedings of the British Machine Vision Conference*, 2008.
- [11] 손동원, "사회 네트워크 분석," pp.25-38, 161, 242-244, 경문사, 서울, 2002.
- [12] Y. Rui, T.S. Huang, S. Mehrotra, "Constructing Table-of-Content for Videos," to appear in *ACM Multimedia Systems Journal*, Special Issue *Multimedia Systems on Video Libraries*, Sep. 1999.
- [13] F. Nack, A. Parkes, "The Application of Video Semantics and Theme Representation in Automated Video Editing," *Multimedia Tools and Applications*, Vol. 4, No. 1, pp. 57-83, Jan. 1997.
- [14] M. Everingham, J. Sivic, A. Zisserman, "Taking the bite out of automated naming of characters in TV video," *Image and Vision Computing*, In Press, Corrected Proof, Available online, 4 May 2008.
- [15] ISO, "MPEG-7 Overview", Oct. 2004.
- [16] Mehmet Emin Donderler, Ediz Saykol, Umut Arslan, Ozgur Ulusoy and Ugur Gudukbay, "BilVideo: Design and Implementation of a Video Database Management System", *Kluwer Academic Pub*, pp.20-21, 2003
- [17] Menzo Windhouwer, "Feature Grammar Systems", *Incremental Maintenance of Indexes to Digital Media Warehouses*, pp. 32-33, Nov. 2003
- [18] R. Lienhart, S. Pfeiffer, W. Effelsberg, "Video abstracting," *Communications of the ACM*, Vol. 40, pp. 54-62, Dec. 1997.

● 저 자 소 개 ●



박 승 보

1995년 인하대학교 전기공학과 졸업(학사)
1997년 인하대학교 대학원 전기공학과 졸업(석사)
2003년 9월~현재 인하대학교 정보공학과 박사과정
관심분야 : 영상정보 표현, 얼굴인식, USN, etc.
E-mail : molaal@eslab.inha.ac.kr



김 유 원

1987년 경희대학교 기계공학과 졸업(학사)
2003년 인하대학교 대학원 컴퓨터정보공학과 졸업(석사)
2003년 9월~현재 인하대학교 정보공학과 박사과정
관심분야 : 디지털 방송, 멀티미디어 프로세스, 컴퓨터 비전, etc.
E-mail : yoowon@eslab.inha.ac.kr



조 근 식

1982년 인하대학교 전자계산학과 졸업(학사)
1985년 Queens College/CUNY M.A., Computer Science 졸업(석사)
1991년 City University of New York Computer Science 졸업(박사)
1991년 3월~현재 인하대학교 컴퓨터정보공학과 교수
관심분야 : 인공지능, Semantic Web, 지능형 에이전트 시스템, etc.
E-mail : gsjo@inha.ac.kr