

스토리 창작 특성의 효과적 가시화를 위한 분류 좌표계 연구

김 명 준

이화여자대학교 융합콘텐츠학과

A Coordinate System of Classification for Effective Visualizations of Story Properties

Myoung-Jun Kim

Department of Content Convergence, Ewha Womans University, Seoul 03760, Korea

[요 약]

장르 및 행위는 스토리의 분류뿐만 아니라 그 특성 데이터의 분포를 가시적으로 나타내는 데에도 효과적으로 사용될 수 있다. 본 논문에서는 스토리 특성 데이터의 분포를 장르-행위의 2차원 평면에서 가시화함에 있어, 인접한 장르 및 인접한 행위가 서로 유사성을 가지는 즉 공간적 특성을 가지는 장르-행위 좌표계를 제안한다. 제안된 장르-행위 좌표계를 이용하여 스토리 특성 데이터의 분포를 가시화 해본 결과 유사도가 높은 항목들이 연이여 좌표계의 항목을 이루고 또한 관련성 있는 특성 데이터들이 군집을 이루어 나타나는 등 공간적 의미를 가지도록 스토리 특성 데이터의 가시화가 가능함을 확인하였다.

[Abstract]

Genres and actions of stories can be used to classify stories, and used effectively as well for visualizing story properties. This paper proposes a Genre-Action coordinate system for visualizing story property data in 2-dimension that has similarities between the genre and action items along the axes, i.e. a property of spatial continuum. With the proposed Genre-Action coordinate system we found that the genre and action items in the axes are arranged according to their similarities and we were able to achieve a spatially meaningful visualization of story properties where the related data form clusters.

색인어 : 스토리 특성 가시화, 외판원 문제, 장르-행위 좌표계, 창작 성향, 최소 해밀턴 경로

Key word : Creative Tendency, Genre-Action Coordinates, Shortest Hamiltonian Path, Story Property Visualization, Travelling Salesman Problem

<http://dx.doi.org/10.9728/dcs.2017.18.6.1119>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<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 26 September 2017; Revised 13 October 2017

Accepted 25 October 2017

*Corresponding Author; Myoung-Jun Kim

Tel: +82-02-3277-3347

E-mail: mjkim@ewha.ac.kr

1. 서론

장르 및 주제는 이야기를 분류하는 기본적인 방법이다. 이찬욱, 이재영은 장르, 주제, 인물, 배경으로 유형화한 스토리 데이터베이스가 필요하다고 보았다[1]. <스토리헬퍼>[2]-[5]는 1천여 건의 스토리 시놉시스를 장르, 주제(행위) 및 인물의 성격으로 분류하여 데이터베이스를 구축하였으며 사용자의 검색에 따라 유사도가 높은 스토리를 찾는 서비스를 제공한다. 사용자가 원하는 스토리는 찾는 데 있어 <표 1>의 질의어를 사용하는 데 여기서 장르와 행위가 가장 비중이 높음을 알 수 있다.

장르 및 행위는 분류뿐만 아니라 이야기 데이터의 분포를 가시적으로 나타내는 데에도 효과적으로 사용될 수 있다. 김명준은 <표 1>에 해당하는 스토리 특성 데이터를 좌표화하고 이를 장르 및 행위로 이루어진 2차원 평면에 점으로 표시함으로써 이야기 DB의 분포를 보면서 원하는 이야기를 찾아가는 검색 UI를 제안하였고[6], 같은 좌표계를 이용하여 스토리 검색 사용자들의 검색기록에 나타난 장르, 행위, 인물 창작성향을 장르-행위 평면에서 각각의 점으로 표현하여 전체적인 창작성향 분포를 한눈에 보여주는 <그림 1>과 같은 가시화 기법을 제안하였다[7]. 또한, 장르-행위 좌표계는 <그림 2>와 같이 인물 유형의 군집화 결과를 가시화하는 데에도 사용될 수 있다[8].

<그림 1>과 <그림 2>는 모든 조합의 장르-행위 데이터를 하나의 평면에서 잘 보여주는 가시화이지만 이를 장르-행위 좌표계로 간주하기에는 한 가지 문제가 있다. 장르 축에서 각 장르, 행위 축에서의 각 행위는 임의의 순서로 나열된 것일 뿐 인접한 장르, 인접한 행위 간의 유사성이 없다. 좌표계라고 함은 1, 2, 3이라는 눈금이 있다면 1과 2, 2와 3은 서로 인접한 것이고 1과 3 사이 보다 가깝다는 공간적인 특성이 있어야 한다. <그림 1>에서 장르, 행위의 순서는 그 공간적, 위치적 의미는 없으며 점들의 분포 또한 불규칙한 패턴을 보인다.

<그림 1>에서 장르 및 행위는 그 순서를 바꾸어 그 분포가 보다 부드럽고 보기 좋게 바꿀 수 있다. 그 예로서 <그림 3>은 장르 및 행위의 각 좌표축에서 데이터의 빈도수가 높은 순으로 장르, 행위를 정렬한 것이다. 이는 <그림 1>에서 비해서는 상당히 부드러운 변화를 보이고 전반적으로 빈도수에 따라 정렬된 형태를 보이는 장점이 있다. 하지만 단지 분포 패턴의 변화가 상대적으로 부드러운 것으로 모든 인접한 장르, 인접한 행위 간의 유사성을 보이지는 않는다. 판타지, 드라마, SF를 행 단위로 비교해 보았을 때 제법 유사한 형태를 보인다. 이는 빈도순으로 정렬하면서 유사한 것끼리 인접하여 순서를 이룬 것으로 <그림 1>에 비해서는 적절한 정렬의 순서로 볼 수 있으며 결과적으로 분포 모양도 부드럽게 변하는 편이다. 하지만 이후 스릴러, 멜로로 이어지는 부분은 분포의 모양이 크게 달라 공간적 연속성을 보이지 않는다. 일반적인 인식으로도 판타지-드라마 사이의 유사성보다 SF-스릴러, 스릴러-멜로 간의 유사성이 적게 느껴지므로 SF-스릴러, 스릴러-멜로는 서로 인접하지 않는 것이 자연스러울 것이다.

표 1. 스토리 검색 질의/답변 및 가중치

Table 1. Query/answers for story search and weighting factors

Query	Answers	weight	
장르	서부극, SF, 갱스터, 전쟁, 멜로, 드라마, 로맨틱 코미디, 코미디, 재난영화, 에로, 액션, 사회, 역사, 뮤지컬, 댄스, 무협, 전기, 다큐, 실험, 어드벤처, 스릴러, 호러, 판타지	18	
타겟	Children, Adult, Family	3	
행위 (주제)	추구, 모험, 추적, 구출, 탈출, 복수, 수수께끼, 라이벌, 희생자, 유혹, 변신, 변모, 성숙, 사랑, 금지된사랑, 희생, 발견, 지독한행위, 상승, 몰락, 도전, 순애보	30	
Character	연령대	영유아, 청소년, 청년, 중장년, 노년	3
	성별	남, 여, 중성	3
	신체능력	장애, 병약, 보통, 건강, 우월	3
	종족	인간, 외계인, 영혼, 동물, 비인간	6
	유형	돈키호테형, 햄릿형	3
	성향	외향형인물, 내향적인물	3
	상황인지	감각, 직관	3
	판단방법	사실, 감정	3
	행동양식	판단, 인식	3
	관심상황	과거, 진행, 현재, 미래	6
	행동관심사	공감, 행동, 책취, 학습	6
	대응방식	계획, 연기, 구현, 고안	6
	내적문제 원인	기억, 충동적욕망, 전의식, 잠재의식, 인지, 의식	3
	계급	노예, 평민, 귀족, 왕족	6
	경제수준	저소득층, 서민, 중산층, 고소득층, 부유층	6
	교육수준	교육받은적없음, 낮은수준, 일반수준, 고등수준	6
	정치성향	보수, 중도, 진보, 혁명, 무정부주의자	6
인간관계	외톨이, 독립적, 사교적, 타인의존적	3	
인종	백인, 흑인, 황인, 혼혈	3	
직업	초, 중, 고등학생, 대학(원)생, 자영업, 회사원, 교사, 전문직, 연예인, 일용직, 경찰, 주부, 프리랜서, 공무원, 범죄자, 무직, 기타	15	
아버지성격	없음, 무관심, 가정적, 가부장적, 폭력적	6	
어머니성격	없음, 무관심, 유약, 가정적, 학대	3	
어린시절	불우, 유복, 총명, 아둔	3	
혼인관계	미혼, 기혼, 이혼, 재혼, 사별	6	
자녀유무	없음, 있음, 떨어져있거나사망	3	
성적취향	이성애자, 동성애자, 양성애자	3	

행위 축, 즉 열 단위로 보았을 때는 가장 빈도가 많은 ‘모험’, ‘사랑’이 인접해 있지만 그 모양이 상당히 다르고 실제로 주제의 유사성 또한 많지 않다고 볼 수 있다. 다른 예로 ‘탈출’과 ‘금지된사랑’도 인접해 있지만, 분포의 모양도 다르고 실제 인지되는 유사성도 적다고 볼 수 있다. <그림 3>과 같이 빈도수에 따라 장르, 행위를 정렬하는 것은 <그림 1>에 비해서는 전반적으로 분포의 모양이 좋기는 하지만, 유사성 있는 항목들이 인접하면서 분포의 공간적 의미를 가지기에는 여전히 부족하다고 할 수 있다.

본 논문에서는 장르-행위 좌표계를 이용하여 스토리 데이터의 분포를 가시화하는 데 있어, 인접한 장르 및 인접한 행위가 서로 유사성을 가지는 즉 공간적 특성을 가지는 장르-행위 좌표계를 제안하고자 한다. 먼저 스토리 검색 서비스에 나타난 사용자 기록 데이터를 이용하여 각 장르 사이의 유사도 및 행위 사이의 유사도를 정의한다. 이를 이용하여 가능한 한 인접한 장르 및 행위의 유사도가 높도록 정렬을 하는 방법으로 장르-행

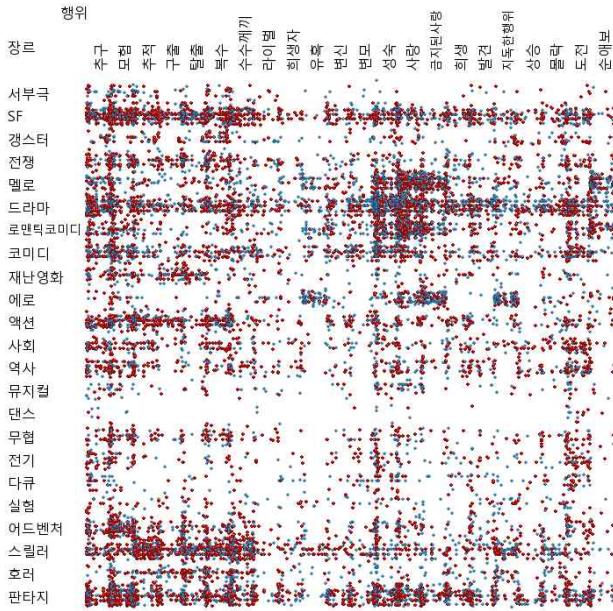


그림 1. 스토리 검색 서비스에 나타난 인물성향(판단방법)의 장르-행위 분포 가시화
 Fig. 1. Visualizing genre-action distribution of a character tendency appeared in a story search service

위 좌표계를 구성하였고, 보다 공간적 의미를 가지도록 스토리 특성 데이터의 가시화가 가능함을 확인하였다.

II. 장르-행위 좌표계

2-1 장르, 행위간 유사도 거리

유사한 장르(또는 행위)가 서로 인접하도록 나열하기 위해 장르 간의 유사도 수치가 필요하다. 스토리의 장르 간 및 행위 간의 유사도는 사람이 얼마나 이를 가깝게 인식하는 가이다. 본 논문에서는 스토리 검색 서비스인 <스토리헬퍼>의 사용자 검색 기록 데이터를 이용하여 사용자들이 선택한 장르-행위 데이터로 부터 장르 간 유사도, 행위 간 유사도를 계산하여 사람들의 유사하게 인식하는 정도를 수치화한다. 이를 위하여 2만 여 건의 검색에서 선택된 장르, 행위의 건수를 계산하였다. <그림 4>는 장르-행위 별로 그 검색 건수를 나타낸다. 본 논문에서는 두개의 장르가 서로 유사도가 높다면 그 장르에서 다루어지는 주제(행위)의 분포 또한 유사하다고 가정하였다. 즉, <그림 4>에서 각 장르 별로 행위에 따른 건수들이 '행' 데이터로 나타나는 데, 유사한 장르는 이 '행' 데이터가 서로 비슷할 것이라고 가정한다. 따라서 두개의 장르 사이의 '유사도 거리'를 해당 '행' 데이터 간 차이의 절대값의 총합으로 정의한다. 마찬가지로, 두개의 행위가 서로 비슷한 의미를 가진다면 그 행위(주제)를 다루는 장르의 건수 분포인 '열' 데이터가 서로 유사할 것이

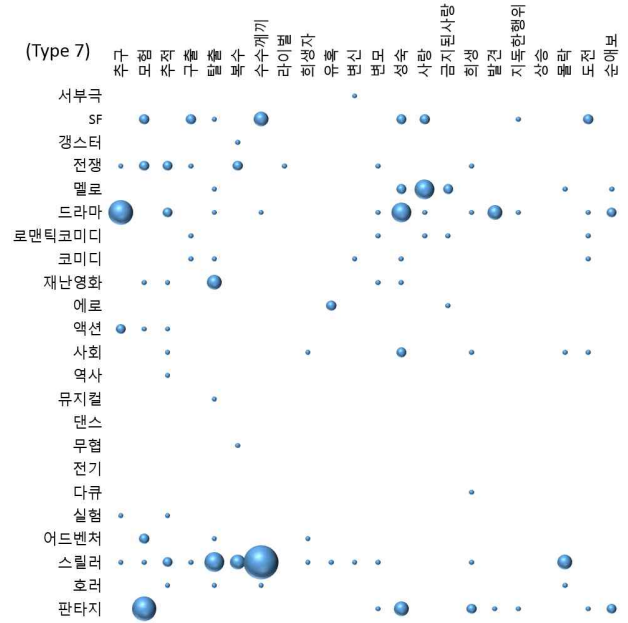


그림 2. 인물 유형의 장르-행위 분포 가시화
 Fig. 2. Visualizing genre-action distribution of a character type

라고 가정하고, 두 개의 '열' 데이터 간 차이의 절대값의 총합으로 행위 사이의 '유사도 거리'를 정의한다. <그림 5>는 이와 같이 정의된 장르 간 유사도 거리 및 행위 간 유사도 거리를 나타낸다. 각각의 유사도 거리는 그 거리가 가장 큰 것을 100으로 환산하여 표시하였다.

장르에서 그 유사도 거리가 가깝게 나타나는 예를 살펴보면 다큐/전기 3, 뮤지컬/댄스 4, 갱스터/무협 6 등이며 실제로도 유사하다고 인지되는 장르들의 유사도 거리가 작게 나타났다. 유사도 거리가 먼 장르들의 예로는 판타지/멜로 100, 판타지/댄스 99, 댄스/드라마 89 등으로 실제로 유사성이 적게 인지되는 경우들로 나타났다. 판타지의 경우에는 대부분의 다른 장르들과의 유사도 거리가 큰 값으로 나타났는데, 이는 판타지 자체가 빈도수가 많은 장르로서 값들이 크게 나타나기 때문이다. 따라서 판타지 장르와의 유사도 거리 값을 상대적으로 해석하여 판타지 장르와의 유사성을 판단할 수 있다. 판타지와 가장 유사한 장르는 SF로 거리가 44로 그 값이 크지만 다른 장르에 비해 월등히 작은 값이며, 실제로도 장르 중에서 SF가 판타지와 가장 유사하게 인지된다.

행위에서는 라이벌/상승 3, 희생/라이벌 4, 발견/변신/변모 4, 몰락/변모 5 등이 유사도 거리가 가장 가깝게 나타나며 실제로 인지되는 유사도에 부합하는 수치임을 보이고 있다. 유사도 거리가 서로 먼 행위로는 모험/사랑 100, 모험/금지된사랑 76, 추적/사랑 66 등이 나타났다. 장르에서와 마찬가지로 거리 값은 상대적인 것으로 빈도수가 많은 '사랑'과 다른 행위 사이의 유사도 거리 값은 대부분 큰 값으로 나타나 최소값이 성숙 48이다. 이는 수치로는 큰 값이지만 상대적으로 작은 값이며 실제

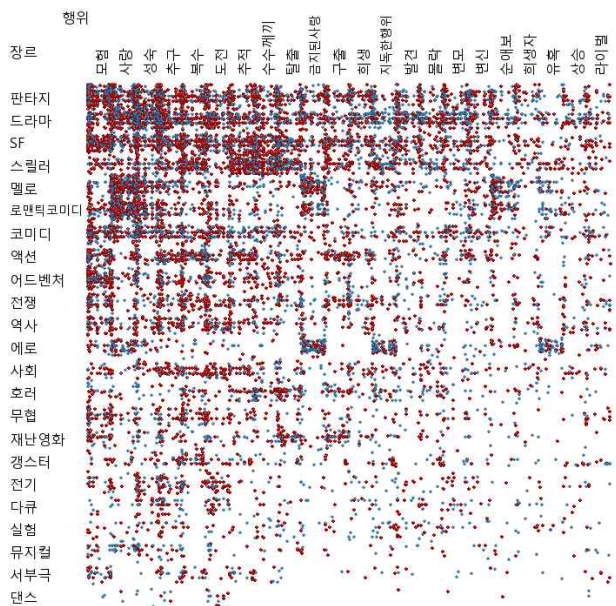


그림 3. 빈도수로 높은 순으로 장르, 행위 각 좌표축을 정렬한 장르-행위 분포 가시화
 Fig. 3. Visualization of genre-action distribution, where each genres, actions are sorted by their counts in the axes

로 느껴지는 유사도로도 성숙과 사랑은 관련성이 많다.

2-2 외판원 문제 해를 이용한 장르-행위 좌표계

서로 유사한 장르(또는 행위)들이 서로 인접하여 공간적 의미를 가지는 장르-행위 좌표계를 구성하기 위해서는 장르를 정렬하여 가능한 한 서로 유사한 장르는 인접하거나 가까이 위치하도록 정렬하는 것이 필요하다. 이는 장르를 정렬하는 순서 중에서 그 인접하는 장르 사이의 거리들의 합이 가장 작은 순서를 의미한다. 장르의 개수가 23개 이므로 23!의 순서 중 인접한 장르의 거리 합이 최소인 순서를 찾는 문제이다. 이는 일종의 외판원 문제(Traveling Salesman Problem)로서 최적해(optimal solution)를 구하는 데 아주 많은 시간이 드는 문제이다[9]. 외판원 문제에서는 외판원이 정해진 도시들을 한번씩 방문하고 처음의 위치로 돌아오는 데 있어 최소한의 이동거리를 최소화 하는 문제이다. 장르를 도시로 장르 간 유사도 거리를 도시간의 거리로 간주하면 외판원 문제가 된다. 단, 외판원 문제는 처음으로 되돌아오는 순환 경로를 찾는 것이므로 중간에서 이를 끊어 장르 축을 위한 순서로 사용한다. 외판원 문제는 그 최적해를 구하는 데 아주 많은 시간이 드는 어려운 문제로서 불필요한 검색을 피하는 전통적인 분기한정(branch-and-bound) 방법을 사용하는 경우 시간이 많이 걸린다. 현재 외판원 문제의 최적해를 구하는 가장 빠른 방법은 concode[9][10]로 알려져 있다. 논문에서는 이를 사용하여 수초 이내에 최적해를 구하였다. 인

장르	판타지	드라마	SF	스릴러	멜로	로맨틱코미디	액션	코미디	역사	어드벤처	전쟁	역사	에로	사회	무협	전기	다큐	실험	뮤지컬	서부극	댄스	
장르	30	44	17	2	3	23	7	6	0	2	2	3	3	1	0	2	4	3	1	0	4	2
행위	204	471	117	101	157	136	123	17	31	6	63	41	134	76	24	97	62	40	22	42	117	16

그림 4. 스토리 검색 기록에 나타난 장르-행위 창작 성향
 Fig. 4. Creative tendency of Genre-Action appeared in a story search service

장르	판타지	드라마	SF	스릴러	멜로	로맨틱코미디	액션	코미디	역사	어드벤처	전쟁	역사	에로	사회	무협	전기	다큐	실험	뮤지컬	서부극	댄스		
장르	59	6	16	47	86	41	27	8	17	25	11	14	5	5	8	6	6	6	23	49	12	96	
행위	55	43	77	62	70	34	53	63	36	49	45	59	62	52	55	57	58	36	50	50	44		

그림 5. 장르 간의 유사도 거리 (상), 행위 간의 유사도 거리(하)
 Fig. 5. Similarity distance matrix between Genres (upper), between Actions (lower)

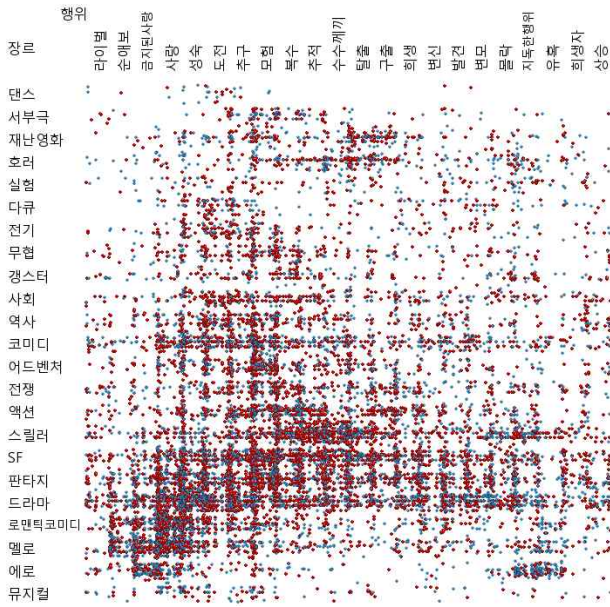


그림 6. 외판원 문제의 해로 각 축을 정렬한 장르-행위 좌표계
 Fig. 6. Genre-action coordinates found by solving Traveling Salesman Problem.

접한 장르 및 행위의 인접 유사도 거리가 최소화 되는 순서는 각각 다음과 같이 계산되었다.

- (장르) 댄스, 서부극, 재난영화, 호러, 실험, 다큐, 전기, 무협, 갱스터, 사회, 역사, 코미디, 어드벤처, 전쟁, 액션, 스릴러, SF, 판타지, 드라마, 로맨틱코미디, 멜로, 에로, 뮤지컬
- (행위) 라이벌, 순애보, 금지된사랑, 사랑, 성숙, 도전, 추구, 모험, 복수, 추적, 수수께끼, 탈출, 구출, 희생, 변신, 발견, 변모, 몰락, 지독한행위, 유혹, 희생자, 상승

장르에서 ‘댄스’와 ‘뮤지컬’은 서로 이어지는 순환 순서이다. 역시 행위에서는 ‘라이벌’과 ‘상승’이 서로 이어지는 순서이다. 이들 순서를 좌표축의 순서로 사용하기 위해서 적절한 위치에서 끊어야 하는 데 빈도수가 적은 장르 및 행위에서 끊어 빈도수가 많은 데이터들이 가운데에 모여서 나타나도록 하는 것이 하나의 좋은 방법이다. 장르에서 가장 빈도수가 낮은 것은 ‘댄스’이고, 행위에서 가장 빈도수가 낮은 것은 ‘상승’이었다. 위의 순서는 가장 빈도수가 낮은 것의 위치에서 순환 순서를 끊은 것이다.

<그림 6>은 이렇게 결정된 순서로 장르-행위 좌표계를 구성하고 <그림 1>을 가시화를 재구성한 것이다. 스토리 특성 분포의 모양이 전반적으로 점진적 변화를 보이며 <그림 1>의 불규칙한 패턴과는 많은 차이를 보이며, 빈도수로 정렬한 <그림 3>과의 비교에서도 보다 정돈되고 부드럽게 변하는 분포 패턴을 보인다. 빈도수가 많은 주요 장르를 볼 때 ‘스릴러-SF-판타지-

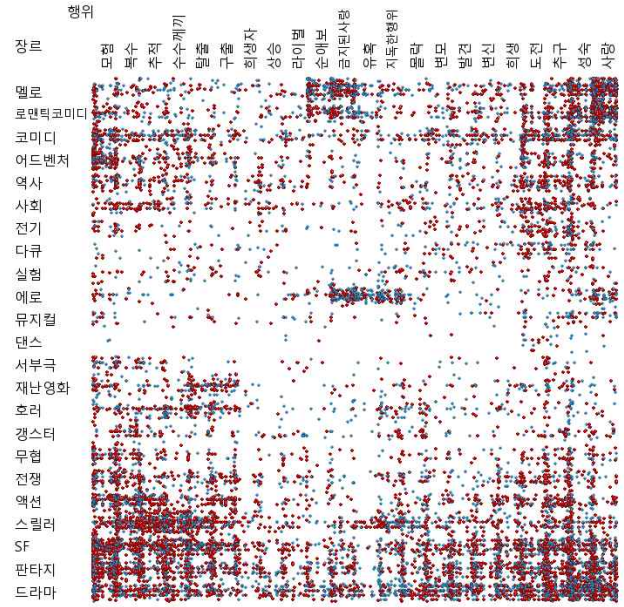


그림 7. 최소 해밀턴 경로 문제의 해로 각 축을 정렬한 장르-행위 좌표계
 Fig. 7. Genre-action coordinates formed by finding the shortest Hamiltonian path

드라마-로맨틱코미디-멜로’로 이어지는 순서가 자연스러워 보이고, 행위에서도 ‘금지된사랑-사랑-성숙-도전-추구-모험-복수’로 순서가 부자연스럽지 않다. 이는 인접 장르 및 행위 간의 유사도 거리를 최소화하는 과정의 결과이다. 또한, 주요 장르 및 행위의 위치에 주로 많은 데이터들이 나타나는 데 <그림 6>에서와 같이 그 분포가 삼각형 형태의 특별한 모양을 이루어 전반적인 분포 해석이 용이하며 이는 장르-행위 좌표계가 공간적 특성을 가짐을 의미한다.

빈도가 낮은 장르나 행위도 유사한 것들이 서로 인접하는 모습을 보인다. 장르에서 재난영화-호러, 다큐-전기, 무협-갱스터가 인접해 있는 것이 의미 있어 보인다. 한편, 빈도가 낮은 곳에서 순환 순서를 끊어서 만든 정렬이기에 상승-라이벌은 매우 유사하고 실제로 계산된 유사도 거리도 짧지만 서로 양끝에 배치되어 있다. 또한, 장르에서 댄스-뮤지컬 또한 매우 유사하지만 양끝으로 떨어져 있는 문제가 있다.

2.3 최소 해밀턴 경로를 이용한 장르-행위 좌표계

서로 유사도가 높지만 빈도가 낮아 좌표계에서 서로 멀리 배치되는 문제를 해결하기 위해서 순환 순서상에서 인접한 항목의 유사도 거리를 최소화하지 않고, 일렬로 늘어난 순서에 대해 인접한 항목 사이의 유사도 거리를 최소화 하는 것으로 좌표축의 항목들을 정렬할 수 있다. 이는 항목들을 한번씩 비순환 방문하면서 움직이는 총 유사도 거리를 최소화하는 ‘최소 해밀턴 경로[11] 문제’이다. 이렇게 최종적인 좌표계에서 일렬로 정렬

된 항목간의 인접 유사도 거리만을 최소화하면 인접 항목 간의 유사도가 보다 높은 장르-행위 좌표계를 구성할 수 있다. 반면, 빈도가 높은 항목을 중심부에 위치시키는 것과 같은 조정은 할 수 없다.

최소 해밀턴 경로 문제는 외관원 문제로 변환하여 그 해를 구할 수 있다. 두 개의 항목은 양끝으로 떨어질 것이기 때문에 최소해에서는 그 거리가 계산되지 않는다. 따라서 모든 두개의 항목 조합에 대해 임시로 거리를 0으로 만들어 각각에 대해 외관원 문제를 풀고 그 중에서 최소 거리를 보이는 경로를 찾는다. 또, 다른 방법으로는 가상의 항목을 추가하고 이 가상항목에서 다른 항목사이의 거리를 모두 0으로 하고 외관원 문제를 푼 후 가상의 항목을 잘라내면 최소 해밀턴 경로가 된다. 본 논문에서는 이와 같은 방법으로 최소 해밀턴 경로 문제를 외관원 문제로 변환하고 *concorde*[10]를 이용하여 그 해를 구하였다.

장르 및 행위의 항목 사이의 인접 유사도 거리가 최소화 되는 순서는 각각 다음과 같이 계산되었다. 앞서와는 달리 비순환 순서로서 실제로 양끝의 항목인 멜로-드라마, 모험-사랑 사이의 유사도 거리는 멀게 나타난다.

- (장르) 멜로, 로맨틱코미디, 코미디, 어드벤처, 역사, 사회, 전기, 다큐, 실험, 에로, 뮤지컬, 댄스, 서부극, 재난영화, 호러, 갱스터, 무협, 전쟁, 액션, 스틸러, SF, 판타지, 드라마
- (행위) 모험, 복수, 추적, 수수께끼, 탈출, 구출, 희생자, 상승, 라이벌, 순애보, 금지된사랑, 유혹, 지독한행위, 몰락, 변모, 발견, 변신, 희생, 도전, 추구, 성숙, 사랑

<그림 7>은 최소 해밀턴 경로로 항목을 정렬하여 장르-행위 좌표계를 구성하고 <그림 1>을 가시화를 재구성한 것이다. <그림 6>의 외관원 문제의 해로 구성된 장르-행위 좌표계와는 상반된 형태로써 가운데 부분이 빈도수가 적고 네 귀퉁이 부분이 빈도가 높게 분포되었다. 이는 인접 항목간의 유사도 거리를 최소화하는 과정에서 빈도수가 많으면서 서로 유사하지 않은 항목들이 서로 멀리 배치되었기 때문이다. 즉, 네 귀퉁이의 항목들은 빈도가 크면서도 서로 상반된 특성을 보이는 것이다. 장르에서는 한쪽 끝이 멜로이고 다른 반대 끝은 드라마이다. 빈도가 크면서 구별되는 장르로 해석할 수 있다. 행위에서는 한쪽 끝이 모험이고 다른 반대 끝은 사랑이다. 역시 비중이 크지만 서로 구별되는 항목이 양끝에 배치된 것을 알 수 있다. 네 귀퉁이는 각각 행위-장르로 표현하면 '모험 멜로', '사랑 멜로', '모험 드라마', '사랑 드라마'이다. 이 중 '모험 멜로' 상대적으로 흔하지 않은 조합이어서 빈도수가 적지만 나머지 세 개는 대표적인 주제-장르이다.

최소 해밀턴 경로로 정렬된 장르-행위 좌표계 상에서 표현된 스토리 특성 분포인 <그림 7>은 주요 장르-행위 들이 네 귀퉁이에 분산되어 있는데, 네 귀퉁이에서 서로 유사성이 높은 것들이 따로 모여 있는 모습을 보인다. 이는 빈도가 높은 것들이 함께 모여 있는 <그림 6>과 달리 분산이 되어 있는 반면 인접 항

목 간의 유사도는 더 좋게 나타날 수 있다.

장르의 항목 순서를 살펴보면 빈도수가 많은 것에서 '멜로-로맨틱코미디-코미디'가 이어진다. <그림 6>에서는 '로맨틱코미디-코미디'가 이어지지 않았다. 반대편의 장르에서는 '드라마-판타지-SF-스틸러-액션-전쟁'로 이어지는 순서는 <그림 6>과 같으나 연이어 '무협-갱스터-호러-재난영화'로 빈도가 적은 항목까지 자연스럽게 이어지며 인접 항목간의 유사도가 더 좋게 나타나는 것을 보인다. 그리고 '댄스-뮤지컬'이 이어져 있다.

행위의 순서에서는 '사랑-성숙', '모험-복수-추적-수수께끼-탈출-구출'이 <그림 6> 및 <그림 7>에서 공통적으로 나타난 순서이다. <그림 7>에서 나타난 '금지된사랑-유혹', '상승-라이벌', '구출-희생자'의 인접관계는 <그림 6>에 비해 인접 유사도가 개선된 것으로 볼 수 있다.

실제 인접 항목 간의 유사도 거리의 총합을 비교해보았을 때도 <그림 6>보다 <그림 7>의 장르-행위 좌표에서 유사도 거리 총합이 작게 나타난다. 장르의 경우 유사도 거리의 총합이 15.4% 감소하였고, 행위의 경우 21.3% 감소하였다.

III. 장르-행위 좌표계 및 창작 성향 가시화 고찰

본 논문에서는 장르-행위로 분류하는 2차원 좌표계를 구성함에 있어 좌표계의 공간적 의미를 부여하기 위해서 장르, 행위 각각의 인접 항목간의 유사도 거리가 최소화되도록 항목을 정렬하여 장르-행위 좌표계를 구성하고 스토리 특성을 가시화하는 방법을 제시하였다. 다음과 같이 두 가지 장르-행위 좌표계를 제안하였다.

외관원 문제의 해로 정렬한 장르-행위 좌표계 (<그림 6>)

빈도가 많은 항목들이 중심에 위치하는 것이 보기 좋은 배치로 생각되는 경우에 사용할 수 있는 장르-행위 좌표계이다. 주된 것들이 내부에 모이고 빈도가 낮은 것들이 주변에 배치되면서도 가능한 한 인접 항목 간의 유사도 거리를 최소화한 장르-행위 좌표계이다. 주된 분포들이 특별한 모양을 이루면서 나타나 이를 한 번에 살펴보기 좋다.

해밀턴 경로의 해로 정렬한 장르-행위 좌표계 (<그림 7>)

유사도가 서로 낮은 구별되는 것들이 네 귀퉁이로 분산되면서 각각 유사한 항목끼리 뭉치를 이루는 특성을 보이는 장르-행위 좌표계이다. 장르에서 '멜로, 로맨틱코미디 ↔ 판타지, 드라마', 행위에서 '모험 ↔ 사랑'처럼 서로 구별되는 대비 관계를 잘 보여 주는 좌표계로서 대비가 필요한 가시화에 적합하다. 외관원 문제의 해로 정렬한 좌표계에 비해 인접 항목 간의 유사도 거리가 보다 최소화되어 인접한 항목 간의 유사도가 좀 더 높다.

본 논문에서 제안하는 장르-행위 좌표계를 사용하여 스토리

특성 데이터를 가시화함으로써 군집화 하여 나타나는 분포를 잘 보여주는 결과를 얻을 수 있었다. <그림 7>에서 살펴보면 {멜로, 로맨틱코미디}-{성숙, 사랑}이 오른쪽 위에서 군집을 이룬다. 가운데 위에서 {멜로, 로맨틱코미디}-{순애보, 금지된 사랑, 유혹}이 아주 많은 빈도는 아니지만 확연한 군집을 보인다. 또한 가운데에서 {에로}-{금지된사랑, 유혹, 지독한행위}가 독특하고 강한 군집을 보이고 있다. 또 다른 의미 있는 군집으로서 {재난영화, 호러}-{탈출, 구출}이 주변에 비해 드러나는 군집을 보임으로서 서로 관련 있는 장르/행위 관계임을 보여 주고 있다. 이처럼 관련성 있는 항목들이 군집 분포를 이루면서 주변과 구별되는 것은 본 논문에서 제시하는 장르-행위 좌표계에서 유사한 항목들이 인접하면서 전반적으로 공간적 연속성을 형성하였음을 의미한다.

IV. 결 론

본 논문에서는 스토리 특성 데이터의 분포를 장르-행위의 2차원 평면에서 가시화함에 있어, 그 좌표계의 인접항목들의 유사도를 최소화하여 좌표계에서 거리가 멀수록 장르나 행위의 유사도가 낮아지는 공간적 특성을 가지는 장르-행위 분류 좌표계를 제안하였다. 주요 항목이 중심부에 나타나는 좌표계 및 구별되는 항목이 서로 멀리 분산되는 두 가지 좌표계를 제안하였으며, 이들 좌표계의 항목 순서를 외관된 문제 및 최소 해밀턴 경로 문제로 구할 수 있음을 보였다.

제안된 장르-행위 좌표계를 이용하여 스토리 특성 데이터의 분포를 가시화 해본 결과 유사도가 높은 항목들이 연이여 좌표계의 항목을 이루고 또한 관련성 있는 특성 데이터들이 군집을 이루어 나타나는 등 제안된 좌표계가 2차원 공간적 의미를 가지는 것을 확인할 수 있었다. 장르, 행위 외에도 인물 같은 다른 스토리 특성에 대해서도 텍스트 데이터에 공간적 의미를 부여하는 후속 연구가 필요할 것으로 보인다.

감사의 글

본 연구는 과학기술정보통신부와 한국연구재단의 여성신산업융합인재양성사업(2015HIC3A1064579)의 지원을 받아 수행된 연구임.

참고문헌

[1] C.-W. Lee and C.-Y. Lee, "A Study of the Movie & Animation Storytelling Planning and Creative Support System," *Humanities Contents*, No. 19, pp. 533-557, 2010.
 [2] Ewha Womans University, "Development of Authoring tool

for supporting Movie and Animation," Project Report (1st year), KOCCA, 2011.
 [3] C.-G. Lyou and Y. Jeong, "A Study about <The Principle of Causality> of Digital Narrative Tool," *Humanities Contents*, No. 22, pp. 183-207, 2011.
 [4] C.-G. Lyou and H.-Y. Yun, "A Comparative Study on the CBR Model of Story Creation Program : focusing on the <Minstrel> and the <Storyhelper>," *Journal of Digital Contents Society*, Vol. 13, No. 2, pp. 213-224, 2012.
 [5] Story Helper. <http://www.stotyhelper.co.kr>.
 [6] M.-J. Kim, "An Analysis of Search Log from a Story Database Service and a New Story Search Method based on Story Map," *Journal of Digital Contents Society*, Vol. 16, No. 5, pp. 797-805, 2015
 [7] M.-J. Kim, "An Analysis and Visualization of Creative Tendency appeared in Query Log of a Story Database Service," *Journal of Korea Multimedia Society*, Vol. 19, No. 8, pp. 1609-1618, 2016
 [8] M.-J. Kim, "Clustering Character Tendencies found in the User Log of a Story Database Service and Analysis of Character Types," *Journal of Digital Contents Society*, Vol. 17, No. 5, pp. 383-390, 2016
 [9] David L. Applegate, Robert E. Bixby, Vasek Chvátal, William J. Cook, *The Traveling Salesman Problem: A Computational Study*, Princeton University Press, 2007
 [10] Concorde TSP Solver [Internet]. Available: <http://www.math.uwaterloo.ca/tsp/concorde.html>
 [11] https://en.wikipedia.org/wiki/Hamiltonian_path



김명준(Myoung-Jun Kim)

1991년 : 한국과학기술원 (전산학 석사)
 1996년 : 한국과학기술원 (전산학 박사)

1996년~1997년: University of Washington Visiting scholar
 1997년~2000년: 한국전자통신연구원 선임연구원
 2001년~현재: 이화여자대학교 융합콘텐츠학과 교수
 ※ 관심분야 : 컴퓨터그래픽스, 영상처리, 클라우드 컴퓨팅, 어플리케이션 개발 등