

정보이론에 근거한 지식분배 관점에서의 내러티브 복잡도 평가

An Evaluation of Narrative Complexity Based on Knowledge Distribution Model and Information Entropy

권호창, 권혁태, 윤완철
한국과학기술원

Hochang Kwon, Hyuk Tae Kwon, Wan Chul Yoon
KAIST

요약

내러티브의 복잡도는 수용자의 이해와 흥미에 직접적으로 영향을 미치기 때문에 창작과정에서 체계적으로 관리되어야 한다. 본 논문에서는 내러티브 구성에 있어서의 지식 분배 작업에 초점을 맞추어 정보 엔트로피 개념을 활용한 복잡도 평가 방법을 개발하였다. 수용자의 지식상태 변화 과정에서 발생하는 정보량을 엔트로피로 계산하여 복잡도 척도로 활용하였다. 실제 영화 내러티브를 대상으로 사례 연구를 수행하였고, 본 방법이 내러티브의 구조적 특성과 전개과정에서의 변화를 효과적으로 반영함을 확인하였다.

I. 서론

영화, 드라마, 소설 등에서 내러티브 복잡도는 텍스트의 내용과 구조에서 기인하며, 수용자의 이해와 정서적 반응에 중요한 영향을 미친다. 소프트웨어나 인터페이스 설계 분야에서의 달리 내러티브의 복잡도는 억제되어야 하는 것이 아니라 창작과정에서 정교하고 체계적으로 관리되어야 한다. 내러티브 수용과 창작과정에서 복잡도가 갖는 중요성 때문에 내러티브의 구조적/구문론적 특성과 수용자의 경험 및 반응 등에 관하여 여러 논의들이 있어 왔다. 하지만 많은 경우 주관적인 인상 비평에 기대거나 너무 포괄적이고 표면적으로 복잡도를 다루어 내러티브 분석과 평가에 직접 활용하기가 어려웠다. 이러한 한계들로 인해 창작과정에서 복잡도를 분석·관리하기는 더욱 어려운 실정이다.

본 논문에서는 이러한 문제를 해결하고자 내러티브의 지식 분배 측면에 초점을 맞추어 복잡도를 정량적으로 평가하고 다면적으로 시각화하는 방법을 제안한다. 먼저, 복잡도 평가를 위한 프레임워크로서 지식 분배 모델을 설명하고, 정보 이론에서의 엔트로피를 복잡도의 정량적 척도로 활용할 수 있는 방법을 제안한다. 그리고 실제 영화 내러티브에 대한 비교 분석을 통해 본 방법의 유효성을 검증한다.

II. 내러티브 지식 분배 모델

내러티브 창작과정에서 작가가 수용자와 캐릭터들의 지식을 정교하게 조절하는 작업은 매우 중요하다. 내러티브가 전개됨에 따라 수용자와 캐릭터는 각자의 시간축과 관점을 따라 사건을 경험한다. 이 경험을 통해 사건에 포함된 정보를 축적하고 추론을 통해 스토리 세계에 대한 '지식 상태'를 갱신해나간다. 이 변화 궤적을 수용자와 캐릭터의 '지식 흐름'(KF: Knowledge Flow)이라 하고, 이를 조절하는 작가의 업무를 '지식 분배'라고 한다. 이전

연구에서 우리는 지식 분배 업무를 효과적으로 지원할 수 있는 '지식 분배 모델'을 제안하였다[1].

이 모델에서 내러티브는 사건, 정보, 지식, 그리고 메타 지식의 연결로 구성된 '지식 구조'로 표상되며, 각 캐릭터와 수용자의 지식 상태는 주어진 시점까지 경험한 사건들에서 인지한 정보들을 근거로 확률 추론 모델을 통해 -1에서 1 사이 값의 확신도로 표현된다. 여기서 -1은 '확실히 아니다', 1은 '확실히 그렇다', 중간 값은 의심스러운 상태, 0은 판단할 수 없거나 정보가 없는 상태를 의미한다. 지식 흐름 $KF(\cdot)$ 은 지식 상태의 변화로서, 사건 순서를 가로축, 지식 상태를 세로축으로 하는 그래프로 시각화된다(그림 1).

창작과정에서 작가는 수용자(KFR), 캐릭터(KFC), 수용자가 판단하는 캐릭터(KFCIR)의 지식 흐름을 파악할 수 있는데, 대부분의 내러티브에서 수용자와 각 캐릭터가 경험하는 사건이 다르고, 사건을 경험하는 순서가 다르며, 같은 사건을 경험하더라도 다른 정도로 정보를 인지하기 때문에 각각의 KF·는 다르게 나타난다.

III. 정보 엔트로피와 내러티브 복잡도

정보이론에서 엔트로피(새넨 엔트로피)는 한 메시지에 포함된 평균 정보량이고, 쿨백-라이블러 발산(Kullabck-Leibler divergence, KLD)[2]은 동일 사건 공간에서 정의된 두 확률 분포의 차이를 나타내며 다음과 같이 엔트로피 차이로 계산된다.

$$D_{KL}(P||Q) = \sum_i p_i (\log_2 p_i - \log_2 q_i) \quad (1)$$

정보 처리 관점에서 내러티브 이해는 작가가 사건을 통해 분배한 정보를 수용자가 모두 정합적(coherent)으로 처리하는 과정이라고 할 수 있다[3]. 이 절에서는 지식 분배 결과인 지식 흐름에서 발생하는 정보량을 내러티브 복잡도 척도로서 제안하고, 지식 흐름 변화와 KLD에 기초한 엔트로피 계산방식을 제시한다.

1. 처리 정보량

지식 분배 모델에서 지식 상태는 확신/불신의 확률 분포이다. 내러티브가 담화의 형식으로 전달될 때 수용자는 새로운 정보를 처리하여 자신의 지식 상태를 갱신하는데, 이때의 정보 이득은 전후 지식 상태의 KLD로 나타난다. 따라서 내러티브에서 수용자가 t 시점까지 획득하여 처리하는 정보량은 다음과 같다.

$$H_R(t) = \sum_{i=1}^t D_{KL}(KS_R(i) \| KS_R(i-1))$$

$$KS_R(i) = [p, 1-p] (\because p = KF_R(i)/2 + 0.5) \quad (2)$$

수용자는 또한 내러티브 이해의 핵심 요건 중 하나로 서 등장인물들의 지식 상태를 내러티브 전개에 따라 유추하고 추적한다(KF · IR). 수용자는 자신이 몰입한 등장인물과 자신, 또는 지식 흐름이 유사한 등장인물들을 묶어서(chunking) 처리하여 정보량을 줄일 수 있다. 수용자가 자신의 지식 상태와의 차이로 등장인물(A)의 지식 상태를 유지한다고 가정하면 KFAIR을 위한 소요 정보량은 다음과 같다.

$$H_C(A,t) = \sum_{i=1}^t D_{KL}(KS_R(i) \| KS_{A|IR}(i)) \quad (3)$$

내러티브를 통해 수용자가 처리하는 총 정보량은 지식 흐름의 전송 관점에서 모든 지식(K)에 대한 식 (2), (3)의 결과를 누적하여 다음과 같이 계산된다.

$$H_P = \sum_{k \in K} \left\{ H_R^{(k)}(n) + \sum_{d \in D} H_C^{(k)}(d,n) \right\} \quad (4)$$

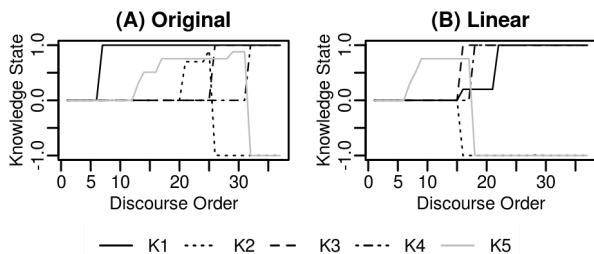
2. 요구 정보량

수용자가 내러티브 내 지식을 모두 올바르게 이해해야 메타 지식을 획득할 수 있다는 가정 하에, 시점 t에서의 지식 상태 KFR(t)와 목표 지식 상태(참인 지식: 1, 거짓인 지식: -1) 간 차이를 요구 정보량으로 정의한다.

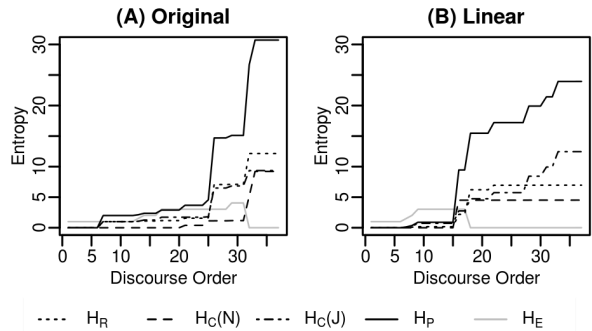
$$H_E(t) = \sum_{k \in K} D_{KL}(KS_D^{(k)} \| KS_R^k(t)) \quad (5)$$

IV. 사례 연구

이상의 방법을 실제 영화 <그을린 사랑>(드니 빌뇌브, 2010) 분석에 적용하여, 원래 내러티브(A)와 사건 순서를 스토리 시간 순서로 배열한 선형 구성(B)을 비교하여 지식 분배 측면에서 내러티브 복잡도가 어떻게 변하는지 비교 분석하였다. 그림 1, 2는 각각 수용자의 지식 흐름과 엔트로피 분석 결과를 나타낸다.



▶▶ 그림 1. <그을린 사랑> 수용자의 지식 흐름



▶▶ 그림 2. 영화 <그을린 사랑> 복잡도 분석 결과

그림 2(A)에서는 크게 두 가지 양상이 나타나는데, 먼저 수용자의 요구 정보량(HE)이 계속 증가하다가 결말 부분에서 급격히 감소한다. 그리고 수용자의 처리 정보량(HP)이 내러티브의 특정 국면에서 급격히 증가한다. 그림 1(A)의 지식흐름에서도 나타나는 이러한 양상은 반전 결말이라는 영화의 구조적 특성과 사건 전개과정에서의 중요 변화와 그에 따른 수용자의 반응을 반영한다. 그림 2(B)에서 먼저 주목할 점은 달라진 변화 패턴이다. 요구 정보량과 처리 정보량 모두 내러티브 초반부에 급격한 변화가 나타나고 이후는 완만하게 진행되는데, 이는 원래 내러티브가 갖고 있었던 시점과 시간의 교차에 의한 지식 분배의 리듬과 역동성이 줄어들었음을 의미한다. 이러한 변화에 따라 총 처리 정보량(HP) 역시 다소 감소하였는데(30.72→23.94) 이는 초기에 주요 정보를 획득한 수용자의 편향된 확신 처리를 반영하는 것으로 보인다. 선형 구조에서도 일정 정도 유지되는 엔트로피는 담화 형식과 상관없는, 내러티브의 지식 구조 자체에 내재한 복잡도라고 볼 수 있다.

V. 결론

본 논문은 정보이론에 근거하여 수용자의 인지적 수준에서 처리되는 내러티브 복잡도를 체계적으로 평가했다는 점에서 의의를 갖는다. 본 논문에서 제안한 방법은 다면적으로 복잡도를 평가하기 때문에 보다 정교한 내러티브 분석과 평가에 적용할 수 있다. 또한 구성 중인 내러티브의 사건 전개에 따른 변화를 시각화할 수 있으므로 창작과정에 효과적으로 활용할 수 있다. 후속 연구에서는 엔트로피 변화 패턴에 따른 내러티브 유형화, 내러티브의 다른 측면에서의 복잡도와와의 관계, 수용자의 정서적 반응과의 관계 등이 규명되어야 한다.

■ 참고 문헌 ■

[1] Kwon, H., et al. (2014). A Knowledge Distribution Model to Support an Author in Narrative Creation, HIMI 2014, Part II, LNCS 8522, 511-522.
 [2] Kullback, S. and Leibler, R. A., "On information and sufficiency", Annals of Mathematical Statistics, 22, pp. 76-86, 1951.
 [3] Branigan, E. (2013). Narrative comprehension and film. Routledge. 63-83.