

## YouTube 데이터를 활용한 미디어 트러스트 요소 연구

문승현, \*정성관, \*\*박해민

KAIST IT 융합연구소

{myth624, \*skjung, \*\*hmin01}@kaist.ac.kr

### A Study of Media Trust Factors using YouTube Data

Seunghyeon Moon \*Sungkwan Jung \*\* Heamin Park

KAIST Institute for IT Convergence

#### 요 약

정보·통신 기술이 발전함에 따라, 적절한 검증을 거치지 않은 미디어 콘텐츠가 공급되고 소비되는 미디어 소비 환경이 형성되고 있다. 이러한 환경에서는 적절치 않은 미디어 콘텐츠의 유통을 제어 하기 위해 기존의 보안 및 인증 기술을 적용하는데 한계가 있기에 새로운 접근 방법의 고민이 필요하다. 본 연구에서는, 건전한 미디어 콘텐츠 소비환경 조성을 위한 미디어 트러스트(Trust)의 개념을 정립하고, 이를 위한 미디어 트러스트에 영향을 주는 4 가지 트러스트 요소들을 제안한다. 그리고, 제안된 요소들의 유효성을 검증하기 위한 YouTube 데이터의 활용 방안에 대해 기술한다. 이러한 접근 방법을 통해 미디어 소비환경에서의 트러스트 개념에 대한 연구 및 기술개발을 활성화 시키고자 한다.

#### 1. 서론

정보통신 기술의 발전으로 인해 손쉽게 인터넷에 접속하여 다양한 미디어 콘텐츠를 소비할 수 있는 환경이 형성 되었다. 이러한 환경 하에서, YouTube 등 UCC 서비스들이 점점 더 활성화 되고 있으며 광고 등을 통한 수익 모델이 대중화 되면서 사용자 관심을 끌기 위한 자극적인 콘텐츠 제공의 유혹이 증가되고 있다. 이에 따라, 서비스 제공자들은 이러한 사용자 콘텐츠를 적절히 관리할 기술에 대한 요구가 증대되고 있다. 이러한 미디어 콘텐츠 소비환경에 대해 기존의 단순한 인증 방식의 관리로는 수많은 미디어 콘텐츠의 위해성 검증 등에 한계가 있으며, 이를 해결하기 위해 트러스트(Trust)라는 개념의 도입 필요성이 점차 부각되고 있다. 본 연구에서는, 건전한 미디어 콘텐츠 소비환경 조성을 위한 미디어 트러스트(Media Trust)의 개념을 정립하고, 이를 위한 미디어 트러스트에 영향을 주는 4 가지 트러스트 요소들을 제안한다. 그리고 제안된 요소들의 유효성을 검증하기 위한 YouTube 데이터 및 메타데이터의 활용 방안에 대하여 기술한다.

#### 2. 미디어 트러스트의 개념

트러스트는 영향을 미치는 다양한 요소가 복합적으로 고려되는 개념으로써 다양한 정의가 존재하며, 트러스트 개념을 사용하는 분야와 상황에 따라 사용되는 정의에 조금씩 차이가 있다. ICT 정보 인프라 영역에서는 특정 임무에 대해 두 개체 A 와 B 가 존재하고 정해진 구체적 기간 동안 존재하는 측정 가능한 믿음을 트러스트라 정의한다[1]. Mayer et al. 은 신뢰를 받는 대상 (Trustee)이 신뢰를 주는 대상(Trustor)에

대해 수행할 특정 행동(Action)에 대하여 신뢰를 주는 대상(Trustor)이 그 행동으로 인해 야기될 위험(Risk)을 감수할 의지라고 트러스트를 정의 하였다[2]. ITU-T 에서는 이와 같이 추상적이고 복잡한 '트러스트'의 개념 정립과 관련 기술에 대한 표준화가 논의되고 있으나 이는 일반적인 IT 서비스 전반에 대한 개념적 논의로 개별 서비스 영역에 필요한 구체화는 아직 진행되고 있지 않다[3]. 본 연구에서는 미디어 서비스 영역에서 트러스트 개념을 사용자와 미디어 콘텐츠의 관점에서 접근하여 사용하고자 한다. 이러한 관점에서 미디어 트러스트(Media Trust)는 사용자와 미디어 콘텐츠 간의 상호작용 관계에서 사용자가 미디어 콘텐츠에 기대하는 것과 다른 결과 혹은 위험을 미디어 콘텐츠가 초래할 가능성이라고 볼 수 있다. 다음 장에서는 이러한 미디어 트러스트에 영향을 미칠 것이라 예상되는 요소들을 크게 4 가지로 나누어 제안하며, 각 요소들을 정량화 할 수 있다면, 미디어 트러스트 또한 이러한 요소의 조합으로 측정할 수 있는 하나의 지표로 건전한 미디어 소비환경의 구성에 사용할 수 있을 것이다.

#### 3. 미디어 트러스트 요소

멀티미디어 서비스 환경에서 일반적인 트러스트 요소를 개체들 간 트러스트 관계에 기반하여 정리하는 연구가 진행되었다[4]. 본 연구에서는 미디어 서비스 영역에서 미디어 트러스트에 영향을 미칠 것으로 고려되는 미디어 트러스트 요소를 다음과 같이 제안한다.

- **Media Integrity:** 미디어 콘텐츠의 메타데이터들과 실제 미디어 콘텐츠가 관련 있는 정도

- **Media Reputation:** 미디어 콘텐츠에 대한 평판 정도
- **Media Harmfulness:** 미디어 콘텐츠의 유해 정도
- **Media Legality:** 미디어 콘텐츠의 합법성 정도

Media Integrity 는 미디어 콘텐츠의 제목, Tags, Description, Recording detail 등의 메타데이터가 실제 콘텐츠의 내용과 얼마나 관련 있는지에 대한 요소로 미디어 트러스트에 긍정적인 영향을 미칠 것으로 예상되는 요소이다. Media Reputation 은 미디어 콘텐츠에 대한 좋아요, 싫어요, Comment 등 콘텐츠의 평판 정보를 고려하는 요소로 미디어 트러스트에 긍정적인 영향을 미칠 것으로 예상되는 요소이다. Media Harmfulness 는 미디어 콘텐츠가 선정적이거나 음란한지, 범죄 충동을 일으키는지, 폭력행위와 약물의 남용을 자극하는지, 도박 및 사행심을 조장하는지 등 미디어 콘텐츠의 유해한 정도를 나타내는 요소로 미디어 트러스트에 부정적인 영향을 미칠 것으로 예상되는 요소이다. Media Legality 는 미디어 콘텐츠가 합법적인 콘텐츠를 보는 요소로 저작권 정보의 유무, 저작권 주장의 정당한 정도로부터 추정할 수 있으며 합법성이 높을수록 미디어 트러스트에 긍정적인 영향을 미칠 것으로 예상되는 요소이다.

#### 4. 미디어 트러스트 요소 유효성 검증을 위한 YouTube 데이터 활용 방안

미디어 트러스트는 앞서 제안한 미디어 트러스트 요소들이 고려되어 Index 의 형태로 정량화 될 수 있다. 미디어 트러스트 Index 는 각 트러스트 요소들을 정량화 하고 이들의 Weighted Sum 의 형태로 모델링하여, 각 미디어 트러스트 요소의 Weight 을 최적화하는 접근방법을 사용할 수 있으며, 정량화된 각 트러스트 요소들을 Input features 로 사용하여 기계학습 알고리즘들을 이용하는 접근방법을 사용할 수도 있다. 이러한 접근 방법을 사용하기 위하여 현재 미디어 트러스트 요소를 도출하여 정량화하는 방법을 연구 중에 있다. 미디어 트러스트 요소를 도출하는 체계화된 미디어 콘텐츠와 풍부한 메타데이터가 필요하다. 대표적인 미디어 서비스 플랫폼인 YouTube 는 미디어 콘텐츠와 더불어 다양한 사용자 콘텐츠 간 상호작용 데이터를 YouTube Data API(v3)를 통해 제공하고 있다[5,6]. 표 1 은 각 미디어 트러스트 요소를 정량화하는데 이용 가능할 것으로 예상되는 YouTube 메타데이터를 보여준다.

| 미디어 트러스트 요소       | YouTube 메타데이터 |  |
|-------------------|---------------|--|
| Media Integrity   | Video         | Title, Description, Tags, Recording details, Content details |
| Media Reputation  | Video         | View count, Like count, Dislike count, Comment count         |
|                   | Comment       | Text, Viewer rating, Author id                               |
| Media Harmfulness | Video         | Title, Description, Tags, Topic details                      |
|                   | Comment       | Text, Author id  |
| Media Legality    | Video         | Licensed content, License                                    |

표 1 미디어 트러스트 요소와 수집 필요한 YouTube 데이터

미디어 트러스트 요소들이 미디어 트러스트 Index 의 정량화에 유의미한지 알기 위해서는 미디어 콘텐츠와 메타데이터를 수집하고 각 트러스트 요소와 미디어 트러스트에 대한 Ground Truth 데이터를 생성하여, 정량화된 미디어 트러스트 Index 가 Ground Truth 데이터와 얼마나 일치하는지 정도를 파악할 필요가 있다. 이를 위해 현재 YouTube 로부터 News 비디오 콘텐츠 150 개, UCC 비디오 콘텐츠 150 개에 대한 비디오와 메타데이터 수집을 진행 하였으며, Ground Truth 데이터 생성을 위해 수집된 데이터에 대해 각 콘텐츠 별로 4 가지 미디어 트러스트 요소와 미디어 트러스트에 대하여 1~ 5 점으로 Annotating 하는 작업을 진행하고 있다. Annotating 작업이 완료되면, 생성된 Ground Truth 데이터의 자체의 유효성을 Annotator 들 간의 일치 정도를 이용하여 검증할 예정이다. Ground Truth 데이터의 생성 및 검증 작업이 완료되면 추후 검토를 통해 공개하여 미디어 트러스트 연구 활성화에 기여하는 것을 고려하고 있다. 미디어 트러스트 요소들에 대한 정량화, 이들의 조합을 이용한 미디어 트러스트 Index 도출 생성한 Ground Truth 데이터를 이용한 검증 단계를 거친 미디어 트러스트 Index 정량화 기법의 개발이 완료되면, 건전한 미디어 소비환경 조성에 기여할 수 있을 것이라 기대한다.

#### ACKNOWLEDGMENT

본 연구는 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2017-0-00318, 건전한 미디어 소비환경 제공을 위한 소셜 IoM 기반 트러스트 미디어 생성 제어 프레임워크 기술 개발)

#### 참 고 문 헌

- [1] 조철희, et al. "ICT 트러스트 정보 인프라 기술 연구동향," 2017.
- [2] Mayer, Roger C., James H. Davis, and F. David Schoorman. "An integrative model of organizational trust." *Academy of management review* 20.3, p.709-734, 1995.
- [3] "ITU-T Technical Report: Trust Provisioning for future ICT infrastructures and services," 2016.
- [4] Moon, S., Jung, S., & Jung, S. "A study on trust factors in a multimedia service environment." *Information Networking (ICOIN), 2018 International Conference on*. IEEE, p.67-69, 2018.
- [5] YouTube, <http://youtube.com>
- [6] YouTube Data API(v3), <https://developers.google.com/youtube/v3/>