

기계학습 기반 블로그 신뢰도 시각화 웹 어플리케이션 개발

Development of a web visualization system with machine learning for weblog credibility

저자 (Authors)	조용걸, 전승곤, 한재호, 한경식 Yonggeol Jo, Seunggon Jeon, Jaeho Han, Kyungsik Han
출처 (Source)	한국HCI학회 학술대회 , 2019.2, 167-172(6 pages)
발행처 (Publisher)	한국HCI학회 The HCI Society of Korea
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08008098
APA Style	조용걸, 전승곤, 한재호, 한경식 (2019). 기계학습 기반 블로그 신뢰도 시각화 웹 어플리케이션 개발. 한국HCI학회 학술대회, 167-172
이용정보 (Accessed)	중앙대학교 175.124.46.*** 2019/10/08 00:11 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

기계학습 기반 블로그 신뢰도 시각화 웹 어플리케이션 개발

Development of a web visualization system with machine learning for weblog credibility

조용걸

Yonggeol Jo

아주대학교

Ajou University

yonggeol93@ajou.ac.kr

전승곤

Seunggon Jeon

아주대학교

Ajou University

vskylife5901@ajou.ac.kr

한재호

Jaeho Han

아주대학교

Ajou University

woghrnt2@ajou.ac.kr

한경식

Kyungsik Han

아주대학교

Ajou University

kyungsikhan@ajou.ac.kr

요약문

본 연구의 목적은 블로그의 신뢰성 판단을 지원하는 시스템을 디자인하는 것이다. 사용자 중심 접근 방법으로 시스템 디자인 요소를 도출하였으며, 이를 Pie-bar chart 와 Radar chart 를 통해 시각화하여 사용자의 신뢰성 판단을 지원하는 웹 확장 프로그램인 '사다리'를 개발하였다. 18 개 분야 (맛집, 국내여행, 뷰티 등)의 16,304 개 블로그를 이용하여 학습된 딥러닝 모델이 사용되었으며, 모델 결과에 대한 설명력을 높이기 위해 K-means 군집화 방법을 추가하였다. 우리 시스템은 잠재 사용자들에게 긍정적인 평가를 받아 실제 활용 가능성을 보였다.

주제어

신뢰성, 시스템 개발, 블로그, 기계학습, 딥러닝

1. 서론

SNS 에 대한 정보 습득의 의존도가 높아짐에 따라 이와 함께 정보의 신뢰성 문제가 대두되고 있다. 실제로 많은 가짜 정보들이 만들어지면서 개인적 및 사회적인 이슈를 만들어내어 왔다(e.g., 2016 년 미국 대통령 선거¹, 2017 년 앙겔라 메리켈 총리의 테러리스트와 셀카², 2017 년 프랑스 대통령 선거³). 이를 해결하기 위해 많은 연구자들이 온라인 정보의 신뢰도를 측정하고 예측할 방법에 대해서 제안하고 연구하고 있다 [1-9].

블로그는 검색 엔진의 결과로 우선적으로 보이기 때문에, 국내에서 블로그에 대한 정보 의존도는 상당히 높은 편이다. 또한 형식의 제약 없이 자신이 관심 있는 특정 주제에 대해 경험 또는 생각과 느낌을 자유롭게 표현할 수 있다는 점에서 일반 사용자들로 하여금 경험에 근거한 솔직하고 믿을만한 정보라고

인식을 주는 경우가 많다 [10]. 이는 다른 SNS (Instagram, Twitter, Facebook)들과는 다르게 블로그를 통해서 다양한 주제에 대해서 심도 있는 콘텐츠의 양과 질을 다양한 시각적 도구와 함께 제공하기 때문이다. 이러한 특징들 때문에 블로그를 통한 마케팅이 많이 이루어지고 있는데, 각종 마케팅 글들은 객관적 정보를 제공하기보다는 광고를 의뢰한 업체에 편향되게 글을 작성하여 사용자에게 왜곡된 정보를 제공하며 신뢰성을 훼손시키고 있다 [11].

정보의 신뢰성을 판단하는 것은 어려운 일이다. 사용자가 해당 정보에 대한 사실 여부를 매번 확인하기 어려울 뿐 아니라, 정보의 신뢰성 판단 과정은 사용자의 지식, 경험, 전문성과 같은 경험적인 요소에 좌우된다. 또한, 사람의 뇌에서 발생하는 정보인지 및 처리 과정의 한계가 합리적인 판단을 더욱 어렵게 만든다 [12,13]. 따라서 본 연구는 사람의 인지적 제한 요소를 보완하기 위해 블로그의 신뢰성 판단을 위한 시스템을 디자인하는 것을 목표로 한다. 이를 위해 우리는 두 가지 목적을 가진다.

- 사용자 스터디를 통해 블로그의 신뢰성을 평가할 때 어려운 요소들을 확인한다.
- 블로그 환경에서 사람의 인지 처리 과정의 한계를 보완하고 신뢰성 판단을 지원하기 위한 웹 어플리케이션을 디자인한다.

2. 이전 연구

본 연구는 이전 연구를 기반으로 확장되어 연구가 진행되었다 [14].

2.1 신뢰성

신뢰성과 관련한 많은 연구가 있었으나, 신뢰성의 개념과 평가하는 요소들은 연구자들마다 다르다. 그 중 Fogg 와 Tseng [15]는 신뢰도를 평가하는 많은 요소 중 진실성(trustworthiness)과 전문성(expertise) 이 두 가지 요소가 많은 연구자들이 신뢰성 연구를 진행할

¹ <https://www.npr.org/2018/04/11/601323233>

² <https://www.bbc.com/news/world-europe-38599385>

³ <http://www.bbc.com/news/world-europe-39495635>

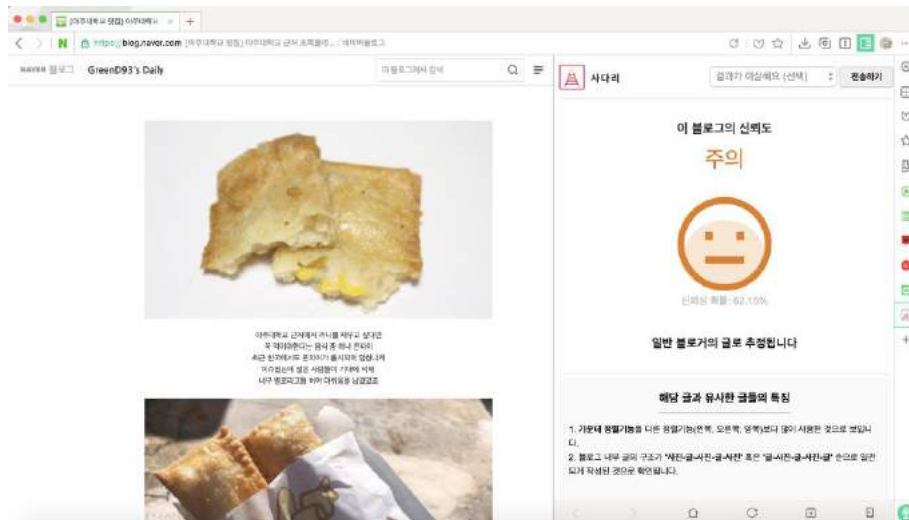


그림 1. 실제 블로그 화면의 신뢰도 결과와 속하는 군집을 보여주는 웹 확장 프로그램, 사다리

때, 고려하는 핵심 요소로 보았다. 진실성이란 제공된 정보가 정보 원천의 솔직한 의견이며 왜곡되지 않았다는 인식이다. 전문성이란 정보 원천이 관련 주제에 대해 타당한 진술을 할 수 있는 지식과 능력을 갖추었다는 믿음이다.

선행연구에 기반하여 이전 연구에서는 ‘해당 분야의 지식을 갖춘 사용자의 솔직한 의견이 담겨있는 글’을 신뢰있는 글로 보았고, 반대로 ‘특정 분야의 경험과 전문적인 지식을 가지고 있지 않은 사용자의 글’을 신뢰성이 떨어지는 글로 정의하였다.

2.2 정보원

Naver⁴에서는 2008년부터 2014년까지 30개 이상의 Category(어학·외국어, 패션·뷰티, 국내여행 등) 별로 5개의 가이드라인(경험, 신뢰, 상업성, 저작권, 활동)과 함께 심사를 거쳐 1년에 한 번 우수한 정보를 제공하고 성실히 기여한 블로그를 해당 분야의 파워블로그로 선정하였고 2014년 이후부터는 이달의 블로그를 선정하여 진행하고 있다. 따라서 우리는 파워블로그에 선정되었거나 이달의 블로그로 선정된 블로그를 신뢰 있는 정보원으로 선정하였다.

Dbdbdeep⁵ 사이트에서는 업체에서 제공받은 사진과 가이드라인을 토대로 블로그 포스팅을 위한 아르바이트생을 구인하고 있으며, 작성 이후 업체에서 검색 여부와 가이드라인 준수를 확인하여 소정의 금액을 지급하고 있다. 따라서 Dbdbdeep에 게시된 업체에서 제공하는 가이드 라인과 사진을 토대로

작성된 포스트는 다른 마케팅 블로그(체험형 블로그)와 다르게 해당 분야에 대한 경험과 지식이 전혀 없는 사용자가 작성하는 글이기 때문에 이를 신뢰성이 떨어지는 정보원으로 정의하였다.

2.3 데이터 수집

데이터는 네이버 블로그를 대상으로 Python으로 작성된 크롤링 코드를 이용하여 수집되었다. 네이버의 경우, 네이버에서 선정된 블로그에 접속하여 각 분야에 해당하는 글을 가져왔으며, Dbdbdeep의 경우 업체가 제공한 가이드라인에 기재된 태그와 글귀로 검색된 블로그를 수집하였다. 데이터 수집 기간은 2018년 4월부터 7월까지 진행되었으며, 18개 분야(맛집, 국내여행, 뷰티 등)에서 Naver가 선정한 989명의 유저의 게시글 9,942개와 Dbdbdeep에서 제공된 정보를 가지고 작성한 유저 1955명의 게시글 6,362개로 총 16,304개의 블로그가 수집되었다.

2.4 모델 구축

모델의 feature는 선행 연구에서 신뢰성을 측정하기 위해 사용되었던 속성들을 참고하였고 [1-9], 추가로 사용자 스테디를 진행하여 글의 일관성(글의 구조, 스타일, 정성)이 신뢰성 여부를 구분하는데 중요한 요소임을 찾아내어 이를 포함시켜 모델을 구축하였다. 3개의 hidden-layer를 가진 Multilayer Perceptron (MLP)모델이 사용되었으며, 모델의 성능은 88.2%의 정확성과 90.2%의 F1-score을 나타낸다 [14].

⁴ <https://www.naver.com/>

⁵ <http://dbdbdeep.co.kr/>

군집	신뢰 유무	신뢰성 있는 포스트 비율	신뢰성 없는 포스트 비율	샘플 수	특징
1	O-1	0.966	0.034	1388	- 경험적 글이 많음 (1 인칭, 2 인칭 대명사를 제일 많이 사용함). - 블로그 내부 글들의 구조가 자유롭고 다양함 (규칙적인 방식으로 글을 쓰지 않음).
2	O-2	0.741	0.259	769	- 경험 기능을 제일 많이 사용함. - 주관성이 낮고 냉정함 (감정표현을 제일 쓰지 않음). - 이미지가 가장 많이 사용됨.
3	O-3	0.667	0.333	1043	- 긍정 부정의 단어의 표현이 가장 많이 사용됨. - 블로그 내부 글들의 구조가 자유롭고 다양함 (규칙적인 방식으로 글을 쓰지 않음). - 이미지가 다른 집단과 비교하여 많이 사용 됨.
4	△-1	0.457	0.543	702	- 감정 단어 점수 값이 높으며 스티커를 많이 사용함. - 해당 분야의 평균 글 길이와 이미지의 개수보다 많음.
5	X-1	0.408	0.592	1728	- 물음표가 가장 많이 사용되었으며 긍정 단어 비율이 높음. - 글의 길이가 길고 태그가 가장 많이 사용됨. - 띄어쓰기 오류가 많이 발견됨.
6	X-2	0.371	0.629	6366	- 중앙경향이 비교적 많음. - 글의 구조가 "사진-글-사진-글-사진" 혹은 "글-사진-글-사진-글" 순으로 일관되어있음.

표 2. K-means 모델 결과 및 군집 특징

3. 본론

3.1 사용자 스터디

본 연구에서는 사용자 중심 디자인(User-centered design)접근을 통해 사용자가 블로그의 신뢰성을 판단할 때, 고려하는 요소를 찾고 이를 시스템에 반영하기 위한 설문조사를 진행하였다. 설문 대상자는 숙소(펜션), 음식(맛집), 화장품(틴트/파운데이션) 중 하나를 선택해서 해당 분야에서 신뢰 있는 블로그 한 개를 선정하고, 아래의 질문에 답하도록 요청 받았다. 설문은 약 15 분의 시간이 소요되었으며, 대면면접조사를 통해 진행되었다. 모든 문항은 주관식으로 구성되어 있으며, 다음과 같다.

1. 이 블로그를 선택한 이유는 무엇인가요.
2. 신뢰도를 판단할 때, 어려웠던 점은 무엇 인가요.
3. 어떤 기능이 추가된다면, 블로그의 신뢰성 판단이 용이해질 것으로 생각되나요.

총 10 명의 대상자가 설문에 참여하였으며, 설문에 응해준 응답자 전원에게 2,000 원 상당의 기프티콘을 지급하였다. 설문 결과는 표 1 과 같다.

블로그 사용자가 신뢰도를 평가할 때, “광고성 사진 혹은 개인적 사진 여부와 글에 사용된 단어 및 문구에 대한 비판과 경험”을 중요하게 생각하였다. 이는 선행 연구 결과와 비슷하다 [1]. 다음으로, “블로그 정보에 대해서 직접 체험 여부 혹은 비판적인 포스팅을 나타낼 수 있는 기능이 시스템에 있으면 좋을 것 같다”라고 응답한 대상자가 많았다. 이를 통해서 우리는 사용자가 블로그의 신뢰성을 평가할 때, 비판(부정)적인 단어의 사용을 중요시 고려하는 것을 확인하여 본 어플리케이션에 비판적인 단어의 비율을 확인할 수 있는 시스템 요소를 추가하였다.

4. 디자인

4.1 네이버 웨일 확장 프로그램

웨일⁶은 네이버가 2016 년에 자체 웹 브라우저 기술로 구글 Chromium 오픈소스를 활용해 만든 브라우저다. 웨일 브라우저는 확장 프로그램 기능을 제공하는데, 이는 추가적인 페이지 이동 없이 브라우저 내에서 손쉽게 사용이 가능하다. 구글의 크롬과 다르게 네이버 웨일의 경우 사이드 바 기능을 추가로 제공하기 때문에, 사용자가 손쉽게 브라우저 내의 확장 프로그램 조정과 관리를 가능하게 한다. 따라서 본 논문은 사용자의 사용성과 접근성을 고려하여 네이버 웨일 확장 프로그램으로 시스템을 디자인하였다.

조사 항목	답변	응답 수
신뢰성 판단 요소	광고성 사진 혹은 개인적 사진 여부	5
	비판적 문구 여부	5
	일상적인 단어 사용 여부	3
	광고 글로 느껴짐 혹은 카테고리가 이상함	3
	객관적인 비교 여부	2
	사용자의 경험이 느껴짐	1
	블로그의 전문성이 느껴짐	1
	사진의 수	1
신뢰성 지원을 위한 필요 기능	공감, 댓글 수, 이웃 수	1
	직접 체험 여부	4
	비판적인 포스팅 여부	3
	글의 구조, 길이 요약	2
	블로그들의 공통적인 평가	2
	댓글의 상호의사소통 정도	1
신뢰성 지원 이외의 추가 요청 기능	사진의 보정 정도	1
	블로그 포스트의 만족도 별점 기능	2

표 1. 블로그 신뢰성 요소 디자인을 위한 사용자 스터디 결과

⁶ <https://whale.naver.com/ko/>

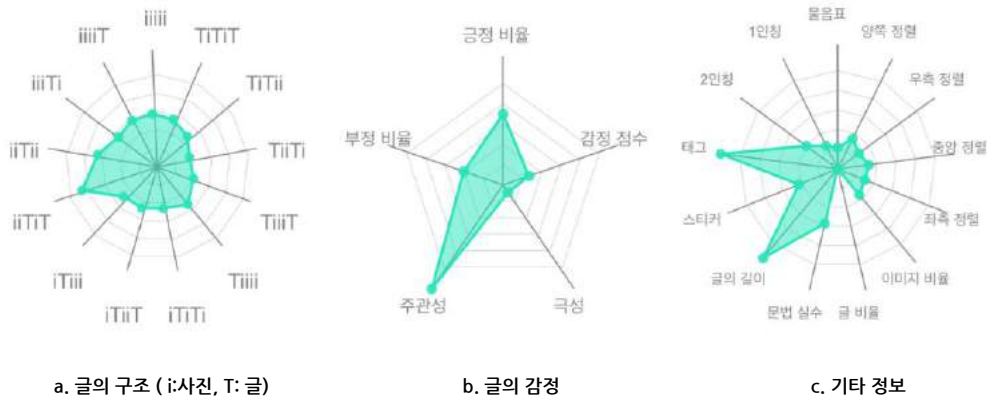


그림 3. Radar chart

4.2 모델

본 연구에서는 블로그의 신뢰성을 평가하기 위해 이전 연구에서 만들어진 딥러닝 모델과 K-means 군집화 모델 2 가지를 사용하였다. 먼저, 이전 연구에서 구축된 딥러닝 모델의 결과는 시스템 상단에서 예측 블로그의 신뢰성 결과를 3 개의 단계와 그에 해당하는 확률로 나타낸다. 1 : 좋음 (신뢰있음으로 분류될 확률 70%이상), 2 : 주의 (신뢰있음으로 분류될 확률 30%이상 70%미만), 3 : 경계 (신뢰있음으로 분류될 확률 30%미만). 신뢰 있음과 없음으로 분류하는 것이 아니라 해당 블로그에 대한 모델의 확률을 함께 제시함으로써 최종적인 판단은 사용자가 하도록 하였다.

딥러닝 모델의 경우 신뢰성 여부에 대한 예측은 가능하지만, 이 블로그 포스트가 신뢰성이 있는지에 대한 이유와 해석을 제시하는 것은 어렵다. 이러한 한계점을 해결하기 위해 블로그를 여러 개의 군집으로 나누어 신뢰성 있는 블로그와 신뢰성이 떨어지는 블로그의 특징을 세분화하여 K-means 군집화 모델을 통해 딥러닝 모델의 설명을 보완하였다. 실험을 통하여 총 6 개의 집단이 수집한 블로그 데이터에 적합하다는 것을 찾아내었고, 표 2 에 집단 별 특징을 정의하였다. O 는 신뢰 있는 블로그의 비율이 높은 집단을 의미하고, 비율이 비슷한 집단을 △로 표시하였으며 마지막으로, X 는 신뢰성이 떨어지는 블로그의 비율이 높은 집단으로 정의하였다. 비록 모든 집단이 이상적으로 분포 되어있지 않아 딥러닝 모델의 결과와 군집화 모델의 결과가 다소 차이가 있을 수 있지만, 딥러닝 모델의 해석을 군집의 속성으로 나타냄으로써 결과에 대한 설명력을 높였다. 해당 결과는 시스템에서 제공하는 그래프와 블로그 분석에 나타난다.

5. 시각화

그림 1 은 본 연구에서 개발된 웹 확장 프로그램이다. Radar chart 와 Bar chart 를 이용하여 현재 블로그와 해당 블로그가 속하는 군집을 시각화하여 비교할 수 있도록 디자인되었으며, 블로그의 31 가지(글의 배치, 긍정 단어 비율, 부정 단어 비율 등)의 특징들을 크게 글의 감정, 글의 구조, 기타 정보 3 가지 그룹으로 나누어 시각화 하였다. 각 그룹에 속해 있는 특징들은 부록에 작성되어 있다. 글의 구조의 iTiIT 와 같은 항목들은 이미지와 글의 배치를 말하며, 이전 연구의 5-gram 모델을 이용하였다. 글의 감정은 서울대학교 감정 사전 (KOSAC)을 이용하였다.⁷

5.1 Pie-bar chart

Pie-bar chart 는 블로그의 모든 특징들을 보여주기 위하여 Pie chart 와 Bar chart 를 결합한 시각화이다. 그림 2 의 가운데 있는 Pie 의 그룹 크기는 각 그룹 안에 속해 있는 특징의 개수를 전체 특징 개수로 나누는 것이며, Pie 위에 있는 Bar 그래프는 각 그룹 안에 속해 있는 특징들의 값이다. Pie-bar Chart 모양을 통해 블로그 전체 속성의 분포를 확인할 수 있고, 어떤 군집과 비슷한 분포를 가지고 있는지 확인할 수 있다.



그림 2. Pie-bar chart

⁷ <http://word.snu.ac.kr/kosac/corpus.php>

5.2 Radar chart

그림 3은 블로그의 정보를 글의 구조, 글의 감정, 기타 정보로 세분화하여 Radar chart를 통해 현재 블로그와 블로그가 속하는 집단과 특징을 각각 비교 가능하게 하였다.

5.3 감정 분석

비판(부정)적 단어 뿐만 아니라 긍정과 중립 단어의 비율과 함께 각 단어의 빈도를 확인할 수 있게 그림 4와 같이 Pie-chart와 Bar-chart를 이용하여 시스템 하단에 나타내었다.



그림 4. 감정 단어 분포 및 단어 시각화

6. 결론

본 논문은 블로그 환경에서 사용자의 신뢰성 판단 지원 시스템을 지원하기 위해 이전 연구와 사용자 스터디를 통해 도출된 항목들을, 두 개의 머신러닝 모델(Neural Network, K-means 군집화)에 기반하여 이를 시각화(Pie-bar chart, Radar-chart)한 웹 확장 프로그램인 '사다리'를 개발하였다. 사용자는 이 확장 프로그램을 이용하여 해당 블로그에 대해 고려해야 할 특징과 전반적인 블로그의 속성을 파악하여 신뢰성 판단에 도움을 받을 수 있을 것으로 기대한다.

본 연구는 다음과 같은 한계점을 가진다. 첫째, 개발된 시스템의 디자인 요소와 평가는 소수의 사용자 스터디로 평가되었다. 향후에는 다수의 사용자 스터디를 통해 해당 시스템에 대한 유용성과 용이성의 검증과 시스템 디자인 요소의 평가가 고려되어야 한다. 둘째, 사다리는 잠재적 사용자들로부터 긍정적인 평가를 받기는 하였지만, 장기적 사용을 통한 시스템의 타당성이 검증되지 않았다. 실제로 개발된 프로그램을 사용했을 때, 사용자의 블로그 신뢰성 판단 결과가 향상 되었는지를 확인하여 시스템에 대한 타당성 검증이 필요하다. 향후 연구에는, 이전 연구의 결과와 비교하여 시스템의 타당성을 검증할 예정이다.

사사의 글

참고 문헌

- [1] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. S. (2013, July). What yelp fake review filter might be doing?. In *ICWSM* (pp. 409-418).
- [2] Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)* (Vol. 6, No. 2010, p. 12).
- [3] Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th international conference on World Wide Web* (pp. 675-684). ACM.
- [4] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96-104.
- [5] Han, K. (2018). How do you perceive this author? Understanding and modeling authors' communication quality in social media. *PloS one*, 13(2), e0192061.
- [6] Lee, K., Caverlee, J., & Webb, S. (2010). Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 435-442). ACM
- [7] Li, F., Huang, M., Yang, Y., & Zhu, X. (2011). Learning to identify review spam. In *Proceedings-International Joint Conference on Artificial Intelligence* (Vol. 22, No. 3, p. 2488).
- [8] Lu, Y., Zhang, L., Xiao, Y., & Li, Y. (2013). Simultaneously detecting fake reviews and review spammers using factor graph model. In *Proceedings of the 5th annual ACM web science conference* (pp. 225-233). ACM
- [9] Stringhini, G., Kruegel, C., & Vigna, G. (2010). Detecting spammers on social networks. In *Proceedings of the 26th annual computer security applications conference* (pp. 1-9). ACM.
- [10] 소현진. (2013). 신뢰받는 블로그 콘텐츠의 특성 탐구. *한국콘텐츠학회논문지*, 13(1), 73-82.
- [11] 장재영, 이병준, 조세진, 한다혜, & 이규홍. (2016). 기계학습기법을 이용한 광고 외식 블로그의 자동분류. *한국인터넷방송통신학회 논문지*, 16(2), 55-62.
- [12] Dörner, D., & Schaub, H. (1994). Errors in planning and decision-making and the nature of human information processing. *Applied psychology*, 43(4), 433-453.
- [13] Simon, H. (1957). *Abehavioral model of rational choice. Models of man, social and rational: Mathematical essays on rational human behavior in a social setting*.
- [14] Jo, Y., Kim, M., Han, K. (2019). How Do Humans Assess the Credibility of Weblogs: Qualifying and Verifying Human Factors with Machine Learning. *SIGCHI* (to appear).
- [15] Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 80-87). ACM.

부록

그룹	특징	설명
글의 구조	글의 배치	이미지나 글의 배치 순서 (e.g., 글·사진·글·사진·글)
글의 감성	긍정 비율	긍정 단어 사용 비율
	부정 비율	부정 단어 사용 비율
	주관성	(부정 단어 + 긍정 단어) / 총 단어 개수
	극성	(부정 단어 - 긍정 단어) / (부정 단어 + 긍정 단어)
	감성 점수	(긍정 단어 - 부정 단어) / 총 단어 개수
기타 정보	물음표 사용	물음표 사용 횟수
	1 인칭 대명사	1 인칭 대명사 사용 횟수
	2 인칭 대명사	2 인칭 대명사 사용 횟수
	태그 사용	태그 사용 횟수
	스티커 사용	스티커 사용 횟수
	글의 길이	포스트 총 길이
	띄어쓰기 실수	띄어쓰기 실수 횟수
	분야 대비 글 비율	포스트에서 사용한 글 길이 / 분야에서 사용한 글 평균 길이
	분야 대비 이미지 비율	포스트에서 사용한 이미지 수 / 분야에서 사용한 이미지 평균 수
	글의 정렬	이미지나 글 정렬한 횟수 (좌측, 우측, 중앙, 양쪽)