

오피니언 마이닝을 이용한 페이스북 팬 페이지 평가 시스템

판트랑녹*, 유 명 식^o

Facebook Fan Page Evaluation System Based on User Opinion Mining

Trong-Ngoc Phan*, Myungsik Yoo^o

요 약

본 논문에서는 페이스북에 게시된 글의 어휘 분석 및 호불호를 평가하여, 게시된 글에 대한 정확한 평가를 할 수 있는 시스템을 제안하였다. 기존 평가시스템과의 성능 비교를 통하여 제안 시스템의 평가 정확도가 높음을 확인하였다.

Key Words : mining, Facebook, fan page, opinion, lexicon-based analysis

ABSTRACT

In this paper, we propose the Facebook fan page evaluation system, which evaluates user opinions based on lexicon-based analysis and positive/negative response from users. By comparing the performance with existing evaluation systems, it is verified that the proposed system can evaluate the fan page in more accurate way.

I. 서 론

쇼셜 네트워크에서 사용자들이 작성하는 글에 대한 정확한 평가는 매우 중요하다. 게시글에 대한 평가 방법으로 게시글에 관심이 있는 사용자 수만을 고려할 수 있다. 그러나 이 방법은 간단한 응용프로그램으로 관심 사용자 수만을 증가 시키는 방법으로 쉽게 좋은

평가를 받을 수 있다는 단점이 있다. 또한 게시글에 대한 평가방법으로 일정기간동안 등록된 사용자 의견 수만 고려할 수 있다. 이 방법의 단점은 긍정적인 의견뿐만 아니라 부정적인 의견까지도 게시글에 대한 긍정평가에 반영된다는 점이다.

이에 본 논문에서는 게시글 및 게시글에 대한 댓글에 대해 어휘분석 및 호불호 점수를 반영한 정확한 평가시스템을 제안하고자 한다.

II. 페이스북 팬 페이지 평가 시스템

제안하는 평가시스템의 전체 흐름 처리를 그림 1에 도시하였다. 우선 크롤러를 이용하여 페이스북 팬 페이지 정보를 수집하고, 수집된 게시글에 대한 정보는 어휘 분석을 통하여 팬 페이지에 대한 평가를 수행하게 된다. 이때 사용자 게시글에 대한 평가는 게시글 자체에 대한 평가와 게시글에 대한 댓글의 평가를 종합하여 수행된다. 이후 평가 결과에 따라 게시글에 대한 선호 순위를 결정한다.

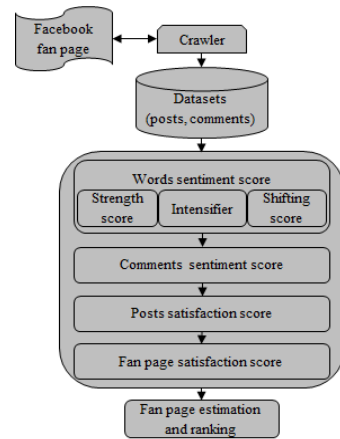


그림 1. 제안 평가시스템
 Fig. 1. Proposed evaluation system

2.1 댓글에 대한 어휘 분석 평균 점수 평가

게시글에 대한 댓글의 어휘 분석은 각 단어들에 대한 평가를 수식 (1)을 통하여 수행 한 이후, 전체 댓글에 대한 평균 점수를 수식 (2)를 통하여 계산한다. 이때 W_{sen} 은 각 단어에 대한 어휘 분석 점수를, S_{nor} 은 특

* 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음(IITP-2015-H8501-15-1008)

♦ First Author : Soongsil University, School of Electronic Engineering, phantr.ngoc@gmail.com, 학생회원

o Corresponding Author : Soongsil University, School of Electronic Engineering, myoo@ssu.ac.kr, 중신회원

논문번호 : KICS2015-12-385, Received December 7, 2015; Revised December 15, 2015; Accepted December 15, 2015

정 댓글에 대한 평균 점수를 나타낸다. 또한 l 은 특정 댓글의 단어 수를 의미한다.

$$W_{sen} = S_{str} \times (100\% + S_{inr}) + S_{shi} \quad (1)$$

$$S_{nor} = \frac{\sum_{i=1}^l W_{sen_i}}{L} \quad (2)$$

S_{str} 는 명사, 동사, 형용사 등에 대한 각 단어에 대한 Strength 점수로 AFINN^[2]에서 제공하는 단어 점수를 사용하였다. S_{inr} 은 강조단어에 대한 점수로 (Intensifier Score)^[3]에서 제공하는 점수를 사용하였다. 또한 S_{shi} 는 단어에 대한 긍정 및 부정을 표현하는 단어에 대한 점수로 긍정은 +4, 부정은 -4를 부여하였다.

2.2 게시물 및 전체 게시물에 대한 만족도 점수
댓글에 대한 호불호의 개수는 게시물에 대한 만족도 평가에 매우 중요하다. 이를 반영하기 위하여 다음의 수식 (3)을 적용한다. 이때 N_{likes} 는 댓글에 대한 좋아요 수를, S_{pow} 는 해당 댓글에 대한 Power Score를 나타낸다.

$$S_{pow} = N_{likes} + 1 \quad (3)$$

이후 수식 (4)를 적용하여 해당 댓글에 대한 최종 어휘 분석 점수 (S_{com})를 평가한다.

$$S_{com} = S_{nor} \times S_{pow} \quad (4)$$

하나의 게시물에 대한 만족도 점수 평가는 게시물 자체에 대한 좋아요 수 (g)와 게시물에 대한 다수의 댓글에 대한 평균 어휘 분석 점수를 반영하여 수식 (5)를 통하여 계산된다. 이때 m 은 총 댓글의 수이다.

$$P_{sat} = \frac{\sum_{i=1}^m S_{com_i}}{m} \times g \quad (5)$$

특정한 하나의 팬 페이지에는 보통 다수의 게시물이 존재한다. 팬 페이지 전체에 대한 만족도 점수는 다음의 수식 (6)을 적용하여 계산한다. 이때 n 은 팬 페이지에 게시된 총 게시물 수를 의미한다.

$$S_{sat} = \frac{\sum_{i=1}^n P_{sat_i}}{n} \quad (6)$$

III. 실험 및 분석

본 논문에서 제시한 팬 페이지 평가 시스템에 대한 성능 분석을 위하여 소셜 팩킷 크롤러 (Social Packet Crawler)^[4]를 사용하여 페이스북에 게시된 7인 (Rihanna, Eminem, Bod Marley, Taylor Swift, David Guetta, Selena Gomez, and Usher)의 유명인에 대한 게시글을 수집하였다. 제안 시스템의 성능 비교 분석을 위하여 팬 수만을 고려한 평가 시스템과 게시글 수(People Taking About : PTA)만을 고려한 평가 시스템의 성능평가도 수행하였다.

그림 2는 팬 수만을 고려한 평가시스템의 평가 결과 그래프이다. 팬 수만을 고려할 경우 인위적인 조작이 쉽고, 또한 팬 중에는 게시글에 대한 긍정적인 팬뿐만 아니라 부정적인 팬이 포함될 수 있다. 실제로 그림 2에서는 Eminem의 팬 수가 Usher의 팬 수보다 많지만, 그림 3의 평가 결과를 보면 Usher에 대한 댓글이 Emimem 보다 많을 것을 볼 수 있다. 따라서 팬 수만을 고려한 평가시스템은 팬 페이지 평가에 대한 신뢰도가 가장 낮다는 것을 알 수 있다.

그림 3과 그림 4는 PTA만을 고려한 평가시스템과 제안 평가시스템의 평가 결과에 대한 그래프이다.

두 결과 그림을 통하여 두 방식이 유사한 결과를 보임을 알 수 있다. 그러나 일부 팬 페이지 평가에서 상이한 결과가 도출 되었다. 특히, David Guetta와 Selena Gomez의 팬 페이지를 비교하면 PTA 평가시스템에서는 David Guetta의 팬 페이지 평가가 높게 나온 반면, 제안 평가시스템에서는 Selena Gomez의 팬 페이지가 높게 평가되었다. 이 경우에 대한 심도있는 비교를 위하여 표1의 데이터를 참조하면, 전체 댓글 대비 부정적 댓글의 비가 David Guetta는 16.2%, Selena Gomez는 11.2%로 David Guetta에 대한 부정적인 댓글이 높음을 알 수 있다. 또한 제안 평가시스

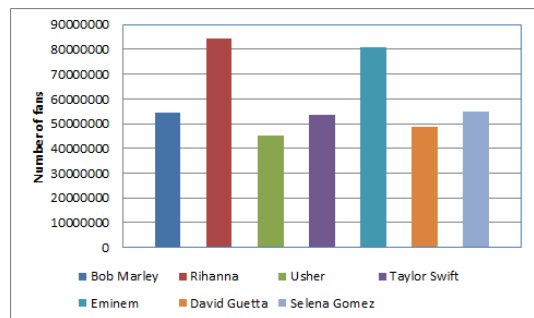


그림 2. 팬 수만을 고려한 평가시스템 평가결과
Fig. 2. Page estimation by the number of fans

템에서 부정적 댓글을 반영한 어휘 분석을 통한 전체 게시글에 대한 만족도 점수는 표1에서와 같이 Selena Gomez의 팬 페이지가 높이 평가되었다.

따라서 PTA와 같이 팬 페이지에 게시된 글 수만을 고려할 경우 부정적인 댓글에 대한 평가가 반영이 되지 않아 정확한 평가가 어려운 반면, 제안 평가시스템은 II장에서 설명한 바와 같이 긍정적 댓글 및 부정적 댓글에 대한 어휘분석을 통한 정확한 평가를 수행함을 알 수 있다.

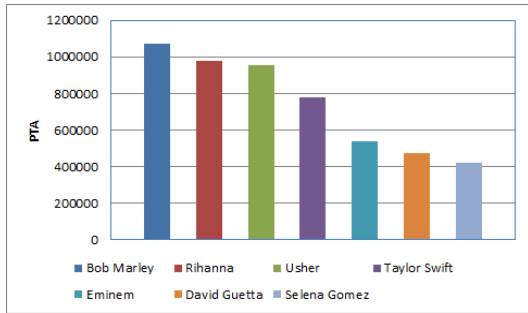


그림 3. 게시글 수만 고려한 평가시스템 평가결과
Fig. 3. Page estimation by PTA

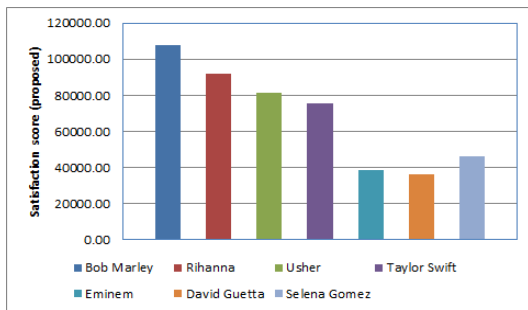


그림 4. 제안 평가 시스템 평가결과
Fig. 4. Page estimation by proposed system

표 1. PTA와 제안방식 비교
Table 1. Comparison between PTA and the proposed

Page	PTA	Proposed	Comments Polarity	
			Neg.	Pos.
David Guetta	475,197	36,032.91	28,334	145,943
Selena Gomez	461,466	46,079.12	203,164	1,602,947

IV. 결론

최근 소셜 네트워크를 통하여 게시되는 글들이 폭발적으로 증가하고 있고, 이러한 게시글에 대한 신뢰성 있는 평가시스템이 절실히 요구되고 있다. 본 논문에서는 기존의 관심있는 팬 수만을 고려하거나 게시글 수만을 고려한 평가시스템이 가지는 단점을 실험을 통하여 입증하였고, 이를 보완하기 위하여 게시글에 대한 어휘 분석 및 호불호를 게시글 평가에 반영하여 정확한 평가를 수행할 수 있는 평가시스템을 제안하였다. 실험 결과를 통해 제안 시스템의 기존의 평가시스템에 비하여 보다 정확한 평가를 수행함을 입증하였다.

References

- [1] S. Moghaddam and F. Popowich, *Opinion Polarity Identification through Adjectives (2010)*, CoRR, Retrieved Oct. 30, 2014, from <http://arxiv.org/abs/1011.4623>
- [2] F. Å. Nielsen, *AFINN word database an affective lexicon (2011)*, Retrieved Sept. 15, 2014, from http://www2.imm.dtu.dk/pubdb/view/s/edoc_download.php/6010/zip/imm6010.zip
- [3] J. Brooke, "A semantic approach to automated text sentiment analysis," M.S. Thesis, Simon Fraser University, Burnaby, Canada, 2009.
- [4] S. F. Wu and F. Erlandsson, *Social Packets crawler(2013)*, Retrieved Aug, 15, 2013, from <https://github.com/dslfaithdev/SocialCrawler>