

질의 기반 사용자 감정상태 예측

민혜진, 강인호
네이버

hyejin.min@navercorp.com, ihkang@navercorp.com

Query-based User Emotion Prediction

Hye-Jin Min, Inho Kang
NAVER Corp.

요 약

본 연구에서는 질의를 기반으로 사용자의 감정상태를 예측하는 방법을 제안한다. 제안방법은 자극-감정 규칙베이스 구축, 규칙확률 값 기반 질의 랭킹, 질의 랭킹 기반 사용자 감정예측의 단계로 구성된다. 방법의 적절성을 검증하기 위하여 힘들다와 심심하다에 대한 결과로 사용자평가를 실시하였다. 힘들다의 결과에서는 힘들다 정도에 대한 점수가 높은 질의들을 지속적으로 검색하는 사용자들을 힘들다라고 판단할 수 있다고 분석되었다. 심심하다의 결과에서는 방법 간 유의미한 차이를 보이지 않았으나, 특정 개별질의의 지속적인 패턴을 분석하는 것이 좀 더 높은 점수를 얻은 것으로 평가되었다.

주제어: 감정 파악, 평판 분석, 감정어휘 사전, 텍스트 마이닝

1. 서론

감정은 내용만으로 표현되지 않는 미묘한 정보를 전달하거나 어떤 의사를 결정하는데, 매우 중요한 역할을 한다고 알려져 있다. 이러한 특징으로 대화시스템, 웹 텍스트 마이닝 등의 연구 분야에서는 다양한 도메인의 텍스트로부터 감정 및 평판을 파악하는 연구가 꾸준히 이루어져 왔다[1]. 뉴스테이터로부터 감정을 파악하는 연구, 상품/영화리뷰로부터 다수 사용자들의 상품/영화에 대한 평판을 분석하는 연구 등이 대표적이라 할 수 있다[2,3,4]. 기존 연구에서는 감정이 태깅된 어휘사전 [5]을 활용하여 bag of words model로 감정값을 합산하거나[4] 문장의 통사적 구조를 분석하여 구조적 우선순위가 높은 성분의 감정값을 취하는 방식 등을 주로 활용하였다[6,7,8].

본 연구에서는 웹 포털 사이트에 접속한 사용자가 검색한 질의(query)를 기반으로 감정을 예측하는 방법을 제안한다. 즉, 질의가 입력 텍스트가 되는데, 다음과 같은 특징이 있다. 첫째, 텍스트가 대부분 명사구의 형태라 기존 방법에서 주로 활용되던 감탄사, 이모티콘 및 서술어 등이 거의 없다. 둘째, 텍스트의 주체를 파악할 수 없으므로 해당 텍스트 상의 감정의 주체가 사용자라는 보장이 없다. 이로 인하여 본 연구에서는 기존 접근 방식과는 다른 다음과 같은 방식을 제안한다. 첫째, 사전과 같은 고정된 감정과약 리소스를 사용하는 대신 사용자가 직접 작성할 가능성이 매우 높은 텍스트 데이터로부터 감정규칙을 마이닝하여 활용한다. 둘째, 주체를 알 수 없기 때문에 사용자가 검색한 여러 질의 중 사용자가 해당 질의를 통해 특정 정보를 획득한 경우의 질의만을 감정 예측에 활용한다. 셋째, 입력 텍스트 상의 주요성분(예: 이모티콘, 감탄사, 감정형용사, 감정동사) 대신 검색을 하는 문맥 내 질의 패턴을 분석에 활용한다. 본 연구에서는 이 같은 방식을 3단계-자극-감정 규

칙베이스 구축, 규칙확률 값 기반 질의 랭킹, 질의 랭킹 기반 사용자 감정예측-로 나뉘 구현하였다. 제안 방법의 적절성을 검증하기 위하여 2가지 감정 ‘힘들다’와 ‘심심하다’에 대한 결과를 바탕으로 사용자평가를 실시하였고, 평가 결과, 감정 타입에 따라 제안 방법의 적절성에 차이를 보였다.

2. 질의 기반 사용자 감정 파악 2.1 자극-감정 규칙 베이스 구축

본 연구에서는 일상생활에서 사람들이 주로 겪게 되는 사건이나 사람들의 행동과 관련된 감정상태를 파악하기 위하여 블로그 데이터로부터 자극-감정 규칙 베이스를 구축하였다. 감정동사 또는 감정형용사를 포함한 형태소 기반 패턴을 활용하여 특정 감정을 발생하게 하는 자극과 감정의 규칙쌍을 블로그 데이터로부터 추출하였다. 그림 1은 그 과정을 ‘심심하다’를 예로 들어 표현한 것이다. 각 문장은 형태소 분석과정을 거친 후 단어와 유효형태소쌍으로 변환된다. 변환된 단어-형태소 쌍에서 패턴 매칭을 통해 자극(명사-형용사/동사)-감정, 감정-결과 규칙을 추출한다. 규칙의 신뢰성을 높이기 위하여 본 연구에서는 인과관계를 나타내는 어미들 (‘어/아서’, ‘-니까’, ‘-는데’ 등 [9]) 가운데, 관계정도가 높은 어미 ‘-어/아서’만을 고려하였다. 또한 동사의 부정형태를 배제하기 위하여 패턴 매칭 시 부정어를 고려하였다.

패턴 매칭 시 고려한 감정어휘는 총 140여개로 세종사전의 의미부류 ‘내(외)재적 심리상태/행위’ [10], 한국어 정서어휘[11] 및 신체부위의 반응에 따른 감정표현 [12] 등을 참고하여 선정하였다. 1년치 블로그 데이터로부터 추출한 규칙은 자극-감정, 감정-결과의 2가지 형태로 각각 74,326개, 114,531개이다.

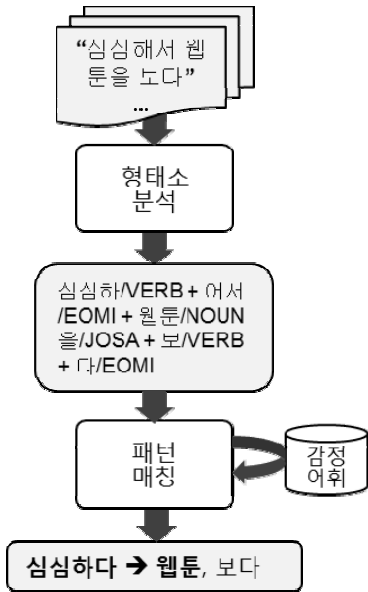


그림 1 규칙베이스 구축과정

가장 높은 비중을 차지하는 감정은 힘들다-좋아하다-아프다-귀엽다-놀라다-궁금하다 순이었으며 부정적인 감정이 긍정적인 감정에 비하여 더 많았다.

2.2 규칙확률 값 기반 질의 랭킹

구축된 규칙베이스를 활용하여 특정 질의와 가장 연관성이 높은 감정상태를 파악하기 위하여 (1)과 같이 특정 질의의 특정 감정에 대한 정도를 파악하였다.

$$P(e|q) = \sum_j P(e, v_j|q) \approx \lambda_1 * \sum_k \sum_j P(e, v_j|t_k) + \lambda_2 * \sum_j P(e, v_j|cat_q) + \lambda_3 * \sum_{w \in D_q} \sum_j P(e, v_j|w) \quad (1)$$

e: 감정, q: 질의, v_j: q가 논항인 형용사/동사, t_k: q의 색인어(term), cat_q: q가 속하는 주제군, D_q: q로 검색된 문서들 중 사용자가 클릭한 문서의 타이틀의 색인어 집합, λ: 가중치

위 식에서 q를 논항으로 취하는 모든 v_j에 대한 감정 확률 값을 더하는 이유는 질의는 대부분 명사형태로 되어 있고 이런 명사를 주어나 목적어로 취하는 동사를 검색을 하는 환경에서 파악하기가 매우 어렵기 때문이다. 또한, 자료 빈약성 (data sparseness)을 보완하기 위해 질의 색인어(t_k) 외에 질의가 속하는 카테고리(cat_q)와 질의로 획득되어 사용자가 클릭한 문서 타이틀(D_q)의 색인어도 같이 활용하였다. 이 때, 질의의 특성을 가장 잘 살리기 위하여 t_k에 가중치(λ₁)를 가장 높게 부여하였다. 질의 중, 실시간 검색어나 핫토픽에 해당되는 것들은 장기간에 걸쳐 축적된 감정(즉 규칙베이스로부터 파악되는 감정)보다 가장 최근에 일어난 이벤트로부터 도출되는 감정과 관련이 더욱 크기 때문에 본 연구의 목적에는 벗어나므로 제외하였고, 이는 추후 연구로 남겨둔다.

표1은 감정 ‘힘들다’에 대하여 (1)의 식으로 특정 질의의 질의에 대한 랭킹 일부를 나타낸다. 상위권에 위치하

는 질의의 경우 주로 ‘병원’, ‘질병’, ‘대출’, ‘시험’ 등과 같은 주제에 속하는 질의이다.

순위	질의	순위	질의
1	유디치과	7	2013년 6월 모의고사
2	사랑니	8	6월 모의고사 범위
3	2014년 3월 모의고사	9	치과
4	모의고사	10	공황장애
5	2014년 6월 모의고사	11	치아보험
6	6월 모의고사	12	폐렴

표 1 ‘힘들다’에 대한 질의 랭킹 일부

2.3 자극-감정 규칙 베이스 구축

2.2절에서 제안한 수식에 의해 질의 A의 특정감정 E에 대한 스코어가 매우 높더라도 질의 A를 입력한 사용자의 감정이 E일 것이라고 단정짓기는 어렵다. 예를 들어, ‘자살’을 1회 검색해본 사용자가 있다고 가정해보자. 이 사용자는 ‘자살’에 관한 뉴스가 궁금해서 검색하거나 인기 검색어에 ‘자살’이 포함되어 있어 호기심에 검색해 보았을 수 있다. 본 연구에서는 이 같은 경우를 배제하고 특정 시간대에 빈번하게 그리고 여러 시간에 걸쳐 지속적으로 검색한 질의 또는 질의 주제군으로부터 파악된 감정이 곧 질의한 사용자의 감정을 가리킨다는 가정으로 (2)와 같은 식에 의해 특정 감정 E에 대한 감정 정도를 측정하였다.

$$UScore(e, Q_u) = \lambda_1 * \sum_{Q_u} P(e|q) + \lambda_2 * PeakScore(e, Q_u) + \lambda_3 * PatternScore(e, Q_u) \quad (2)$$

Q_u: 사용자 u의 질의집합 Q_u = {q₁, q₂, ..., q_n}, λ: 가중치

PeakScore는 특정시간대 주제별 또는 개별 유효질의 (특정 시간대의 특정 사용자의 P(e|q)값이 존재하는 질의 집합)의 비중을 계산한 값이다. PatternScore는 주제별 또는 개별 유효질의의 지속패턴을 계산한 값이다(논문 마지막 장 수식 3,4 참조). 그림 2(논문 마지막장)는 PeakScore와 PatternScore가 각각 높은 예에 해당하는 사용자의 질의 사례이다.

3. 실험 및 평가

3.1 실험 방법

제안한 방법의 적절성(feasibility)을 평가하기 위하여 5일치 질의로그 데이터를 대상으로 다음의 2가지 형태의 평가실험을 진행하였다: 1) 제안하는 수식(2)에서 중요한 비중을 차지하는 PeakScore와 PatternScore에 의해 특정감정으로 판단되는 결과의 적절성에 차이가 있는가?, 2) 제안하는 수식(2)의 값과 피실험자의 평가점수가 연관성을 가지는가? 1)번 실험에서는 2가지 감정 (힘들다, 심심하다)에 대하여 Score 타입 (PeakScore, PatternScore) 및 유효질의의 카테고리 (개별 질의, 주

제별 질의)에 따른 조합결과 데이터를 무작위로 섞어서 피실험자들에게 제시하고 해당감정이 느껴지는 정도에 따라 1~3점의 점수를 부여하라고 요청하였다. 2)번 실험에서도 역시 2가지 감정에 대하여 수식에 의한 값의 범위를 다양하게 분포시키고 무작위로 섞어서 제시한 후, 역시 해당감정이 느껴지는 정도에 따라 1~3점의 점수를 부여하라고 요청하였다. 총 10명의 피실험자가 평가하였고, 연령은 20대~30대로 남:여 비율은 8:2였다.

3.2 평가결과

- 1)번 평가결과: ‘힘들다’의 경우, Score 타입 (PeakScore, PatternScore) 및 유효질의의 카테고리 (개별 질의, 주제별 질의)에 의한 결과의 적절성에 유의미한 차이 (ANOVA, $F(3,996) = 223.56$; $p = .000$)를 보였다. Post Hoc Test 결과를 분석해보니, PeakScore&개별질의, PeakScore&주제별질의 각각이 PatternScore&개별질의, PatternScore&주제별질의와 차이를 보였다. 결론적으로 ‘힘들다’의 경우, 특정 시간의 질의 비중보다는 지속패턴을 측정하는 것이 더 적절하다고 볼 수 있다. 반면, ‘심심하다’의 경우, 유의미한 차이가 없었으나, PatternScore&개별질의의 평가점수의 평균값이 가장 높아져서, ‘심심하다’ 역시 개별질의의 지속패턴이 적절하다고 보인다.
- 2)번 평가결과: ‘힘들다’의 경우, 제안한 수식값과 피실험자들의 평가점수 간의 양의 상관관계(Pearson Correlation $r = .386$, $p = .000$)를 보였고, ‘심심하다’의 경우, 상관관계가 보이지 않았다. 평가 후, 피실험자들의 의견을 종합해보면, ‘심심하다’의 경우, ‘힘들다’에 비해 개인의 취향, 주관에 따라 변동이 커서 애매한 경우가 많아서가 그 원인이라 판단된다. 또한, 실험데이터에 속하는 기간 중 주말이 포함되어 있어서 ‘심심하다’ 여부에 관계없이 방송명을 질의하는 경우도 한 원인이라 분석되었다. 정교한 분석을 위해서는 실험 데이터를 보다 다양하게 구성해야 할 것으로 보이는데, 추후연구로 남겨둔다.

4. 토의 및 결론

본 연구에서는 질의를 기반으로 사용자의 감정상태를 예측하는 방법을 제안하였다. 제안 방법은 크게 자극-감정 규칙베이스 구축, 규칙확률 값 기반 질의 랭킹, 질의 랭킹 기반 사용자 감정예측의 단계로 구성된다. 방법의 적절성을 검증하기 위하여 2가지 감정 ‘힘들다’와 ‘심심하다’에 대한 결과의 일부를 실험 셋으로 생성하여 평가를 실시하였다. ‘힘들다’의 결과에서는 힘들다 정도에 대한 점수가 높은 질의들을 지속적으로 검색하는 사용자들의 경우, ‘힘들다’라고 판단할 수 있다고 분석되었다. ‘심심하다’의 결과에서는 유의미한 차이는 없었으나 특정 개별질을 지속적으로 검색하는 사용자들의 경우, ‘심심하다’고 판단할 수 있다고 분석되었다. 그러나, ‘심심하다’는 사용자의 주관적 취향에 큰 영향을 받으므로 보다 정교한 분석이 필요하다.

향후에는 보다 다양한 감정에 대한 결과를 추가하고, 실시간 검색어처럼 최근 발생한 이벤트로부터 도출되는 감정에 대한 처리방법을 고안하려고 한다. 또한 정확도

향상을 위해 질의 입력의 문맥정보를 파악할 수 있는 방법에 대한 연구를 진행하려고 한다.

참고문헌

- [1] Pang, B., and Lee, L. (2008). “Opinion mining and sentiment analysis.” *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135.
- [2] Wilson, T., Wiebe, J., and Hoffman, P. (2005). “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis.” In *Proceedings of HLT/EMNLP*, pp. 347-354.
- [3] Pang, B., Lee, L., and Vaithyanathan, S. (2002). “Thumbs up? Sentiment Classification Using Machine Learning Techniques.” In *Proceedings of EMNLP’02*, pp. 79-86.
- [4] Hu, M. and Liu, B. (2004). “Mining and summarizing customer reviews.” In *Proceedings of SIGKDD*, pp.168-177.
- [5] Baccianella S., Esuli,A., and Sebastiani, F. (2010). SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, LREC’10.
- [6] Moilanen, K., and Pulman, S. (2007). “Sentiment Composition.” In *Proceedings of RANLP*, pp. 378-382.
- [7] Choi, Y., and Cardie, C. (2008). “Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis.” In *Proceedings of HLT/EMNLP*, pp. 793-801.
- [8] Min, H.-J., and Park, J.C. (2011). “Detecting and Blocking False Sentiment Propagation.” In *Proceedings of 5th IJCNLP*, pp. 354-362.
- [9] 남기십, 고영근. (1998). 표준국어문법론, 박이정.
- [10] 세종전자사전, 21세기 세종계획, 국립국어원.
- [11] 안신호, 이승혜, 권오식. (1993). “정서의 구조: 한국어 정서단어분석,” *한국심리학회지* 7(1), pp. 107-123.
- [12] 임지룡. (1999). “감정의 생리적 반응에 대한 언어화 양상”, *담화와 인지*, 6(2), pp. 89-117.

$$PeakScore(e, Q_u) = \max \frac{average_peak_score(SubQ_{u, type}, e)}{expected_peak_score} \tag{3}$$

$$average_peak_score(SubQ_{u, type}, e) = \frac{\sum_h \frac{n(valid_queries(e, SubQ_{u, type}, h))}{n(all_queries(Q_u, h))} * \log(n(valid_queries(e, SubQ_{u, type}, h)) + 1)}{n(h)}$$

$SubQ_{u, type}$: 사용자 u의 type별 검색어 부분집합, type: 개별질의(q) 또는 주제별(cat) 질의
 h: 시간
 valid_queries: 시간 h의 사용자 u의 $\sum_{Q_u} P(e|q)$ 값이 있는 q의 수
 all_queries: 시간 h의 사용자 u의 모든 q의 수

$$Pattern Score (e, Q_u) = \max \left(\frac{observed_qpat_score(e, SubQ_{u, type})}{expected_qpat_score(e, SubQ_{u, type})} * W \right) \tag{4}$$

$$expected_qpat_score(e, SubQ_{u, type}) = avg_q * n(valid_queries(e, SubQ_{u, type}, h))$$

$$observed_qpat_score(e, SubQ_{u, type}) = \sum_h \frac{n(valid_queries(e, SubQ_{u, type}, h))}{\Delta time}$$

$$w = n(valid_queries(e, SubQ_{u, type}, h)) * \frac{1}{\sum h}$$

avg_q : 사용자의 시간당 평균 질의 수, w: 가중치

시	모든 질의	유효 질의	
16시	엑소 크리스(2) 엑소(1) 노무현(1) 일보드(1) 해커스토익(1) 브금저장소(1) 슈퍼맨이 돌아왔다(1) 엑소 기자회견(1) 아빠어디가(1) 안정환(1) 디시인사이드 갤러리(2)	아빠어디가(1) 엑소(1) 디시인사이드 갤러리(2) 엑소 기자회견(1)	Peak Score ↑
19시	무한도전 시청률(1) 박건형(1) 구혜선(1) 황금레시피 닭볶음탕(1) 박세영(2) 기분좋은날(2) 박봄(4)	기분좋은날(2) 무한도전 시청률(1)	
20시	정덕희(1) 윤소정(1) 튀니지전(1) 구원파 신도(2) 보육교사(2) 기분좋은날(1) 에듀스충남(1)	기분좋은날(1)	
...			
22시	밴드(2) 편성표(1) 왔다 장보리(1) 이상우(2) 무한도전(2) 노홍철(2) 무한도전 치호진(2) 이마트(2) 기분좋은날(1) 케이بل 편성표(1)	밴드(2) 편성표(1) 왔다 장보리(1) 이상우(2) 무한도전(2) 노홍철(2) 무한도전 치호진(2) 이마트(2) 기분좋은날(1) 케이بل 편성표(1)	

시	모든 질의	유효 질의		
7시	넷마블(1) 한게임(1) 다음게임(1)	넷마블(1) 한게임(1) 다음게임(1)	Pattern Score ↑	
8시	한게임(1) 다음게임(1)	한게임(1) 다음게임(1)		
9시	다음게임(1)	다음게임(1)		
10시	넷마블(1) 한게임(1) 다음게임(1)	넷마블(1) 한게임(1) 다음게임(1)		
...				
16시	한게임(1)	한게임(1)		
17시	한게임(1) 다음게임(1)	한게임(1) 다음게임(1)		
18시	다음게임(1)	다음게임(1)		
...				

그림 2 PeakScore와 PatternScore가 높은 질의 사례