

Predicate Logic Form을 이용한 자연어 텍스트로부터의 감정인식

설용수[○], 김동주^{**}, 김한우^{*}, 박정기^{***}

[○]한양대학교 컴퓨터공학과

^{**}한양대학교 컴퓨터공학과

^{***}한양대학교 전자및통신공학과

e-mail: {iamtbn, kimhw}@hanyang.ac.kr, djkim@anyang.ac.kr, parkjk@hanyang.ac.kr

Emotion Recognition from Natural Language Text Using Predicate Logic Form

Yong-Soo Seol[○], Dong-Joo Kim^{**}, Han-Woo Kim^{*}, Jung-Ki Park^{***}

[○]Dept. of Computer Science, Hanyang University

^{**}Dept. of Computer Science, Anyang University

^{***}Dept. of Electronics and Communication Engineering, Hanyang University

● 요 약 ●

전통적으로 자연어 텍스트로부터의 감정인식 연구는 감정 키워드에 기반한다. 그러나 감정 키워드만을 이용하면 자연어 문장이 원래 갖고 있는 통사정보나 의미정보는 잃어버리게 된다. 이를 극복하기 위해 본 논문에서는 자연어 텍스트를 Predicate Logic 형태로 변환하여 감정 정보처리의 기반데이터로 사용한다. Predicate Logic 형태로 변환하기 위해서 의존 문법 구문분석기를 사용하였다. 이렇게 생성된 Predicate 데이터중 감정 정보를 갖고 있는 Predicate만을 찾아내는데 이를 위해 Emotional Predicate Dictionary를 구축하였고 이 사전에는 하나의 Predicate마다 미리 정의된 개념 클래스로 사상 시킬수 있는 정보를 갖고 있다. 개념 클래스는 감정정보를 갖고 있는지, 어떤 감정인지, 어떤 상황에서 발생하는 감정인지에 대한 정보를 나타낸다. 자연어 텍스트가 Predicate으로 변환되고 다시 개념 클래스로 사상되고 나면 KBANN으로 구현된 Lazarus의 감정 생성 규칙에 적용시켜 최종적으로 인식된 감정을 판단한다. 실험을 통해 구현된 시스템이 인간이 인식한 감정과 약 70%이상 유사한 인식 결과를 나타냄을 보인다.

키워드: 감정인식(emotion recognition), 감정(emotion), 자연어처리(natural language processing)

1. 서론

감정 컴퓨팅의 한 분야로 최근 활발히 연구되고 있는 감정인식은 모달리티에 따라 음성, 비전, 생체신호, 텍스트 등으로 분류가 가능하다. 다른 모달리티에 비해 텍스트 기반 감정인식은 최근 발전된 연구 결과를 보여주지 못하고 있으나 텍스트기반 인터페이스를 갖는 웹이나 전자문서, 책 또는 Speech to Text를 통한 응용 등에서 여전히 중요한 비중을 차지하고 있다. 텍스트로부터 감정을 인식하는 능력은 기계가 인간의 감정을 인식하는데 있어 음성이나 비전을 통한 감정인식처럼 단순히 현재 입력 데이터에 대하여 감정 분류를 해주는 수준을 넘어 과거 기억이나 감정 주체, 감정주체의 성향이나 성격등에 따라 지능적으로 감정을 인식하는 보다 인간다운 감정인식을 가능하게 하는데 필요하다. 그동안 텍스트로부터의 감정인식은 감정 키워드를 추출하고 그 감정키워드를

근거로 감정을 판단하는 접근이 대다수를 이루었다. 그러나 자연어 문장속에 들어있는 다양한 구문 정보나 의미정보를 잃어버리기 때문에 감정과 같은 복잡한 인간의 생각을 인식하는데에는 한계가 있었다. 본 논문에서는 자연어를 Predicate Logic 형태로 변환하여 이를 감정 처리의 기반 데이터로 삼는다. 각각의 Predicate은 주어에 대한 정보를 가지므로 텍스트 속에 등장하는 다양한 감정 주체들에 대하여 각각의 감정을 인식해 낸다. Predicate Logic 형태로 만들어진 감정 지식들은 미리 정의된 Predicate Class에 의해 추상화 되고 이 추상화된 개념을 가지고 Lazarus[1]의 감정생성 규칙에 적용을 하여 8가지의 다양한 감정 생성을 시도 한다. 실험에서는 어린이의 일기를 인간이 먼저 감정 인식을 시도 하고 본 논문에서 제안하는 감정인식 시스템으로 감정인식을 시행한다. 실험 결과 인간의 감정인식 결과와 약 70% 유사한 결과를 보였다.

II. 본론

텍스트로부터 감정인식을 하기위해 우리는 자연언어 문장을 Predicate Logic으로 변환 한다. 이때 말하는 Predicate은 문장 내의 본동사를 말하며 Dependency Parser를 사용하여 대상 본동사가 문장을 만들어내는데 필요한 파라미터(주어, 목적어 등)를 찾아낸다. Dependency parser는 Stanford Parser를 사용한다. 모든 문장은 본동사를 필수적으로 갖게 되는데 이 본동사를 Predicate의 이름으로 정하고 그 동사가 갖는 파라미터를 괄호 안에 주어, 목적어, 보어 순으로 적어준다. 예를 들면 Like(I, you), Give(I, you, an apple) 와 같다.

위와 같이 자연어를 Predicate형태로 변환하면 이론적으로 동사의 개수만큼의 Predicate이 만들어질 수 있다. 그러나 모든 동사가 감정을 갖고 있는 것은 아니다. 우리는 감정을 가진 동사만을 걸러내기 위하여 Emotional Predicate Dictionary(EPD)를 구축하였다. 이는 감정 정보를 파라미터로 갖는 동사들만을 모아놓은 사전이다. 이 사전에 해당되는 Predicate만을 유효한 감정 지식으로 판단하고 나머지는 모두 버린다. 이 사전에는 Predicate과 추상화된 개념 Class를 갖고 있다. 따라서 자연어 문장으로부터 생성된 Predicate이 어떤 개념 Class에 사상되는지를 EPD로 찾아내어 감정 생성 규칙 적용 단계의 입력데이터로 이용한다.

감정생성 규칙은 Lazarus의 감정 생성 규칙[1]을 토대로 간략화 하여 만들었다. 감정의 복잡성과 Predicate변환 및 개념 Class 변환과정에서의 추상화로 인한 정보손실로 인해 순수하게 규칙만으로 적용은 무리가 있기 때문에 우리는 KBANN을 이용하여 감정 생성 규칙을 신경망 형태로 구현하고 이를 학습하여 만들어진 분류기로 감정 분류를 시도한다.

우리는 감정 표현이 많이 담겨져 있으며 복잡한 표현이 많지 않은 미국 어린이들의 일기를 학습 및 실험 데이터로 선택하였고 총 5300 문장이다. 각각의 문장은 수작업을 통하여 각각의 문장별로 8가지 감정(anger, fear, hope, sadness, happiness, love, thank, neutral) 중 어떤감정인지를 태깅하였다. 우리는 10-fold validation을 통하여 KBANN을 학습 및 실험하였고, 테스트 결과는 아래 그래프와 같다.

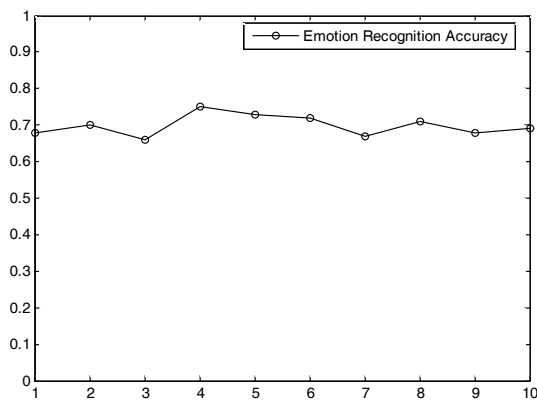


그림 1. 감정인식의 정확도

실험 결과는 위 그림과 같이 평균 70% 정도의 정확도를 보였다.

III. 결론

우리는 텍스트로부터 감정인식을 하기 위하여 기존에 시도되지 않은 Predicate으로의 변환을 통한 감정인식을 시도하였다. Predicate 변환을 위하여 의존 구문분석기를 사용하였으며 EPD를 만들어 개념을 추상화 하였고 이를 KBANN을 통하여 Lazarus의 감정 생성 규칙에까지 적용하였다. 시스템의 감정인식 성능은 약 70%가 나왔다. Predicate으로의 변환 부분, EPD를 통한 개념 추상화, KBANN적용 등 각 부분별로 감정 정보를 잃어버리는 부분을 더 보완하면 좀더 나은 감정인식 성능을 얻을 수 있을 것이라 생각한다.

참고문헌

- [1] R. Lazarus, Passion and reason: Making Sense of our emotions. New York: Oxford University Press, 1994.
- [2] Picard, R.W. Affective Computing, MIT Press. Cambridge, MA, 1997.
- [3] P. Lang. "The emotion probe: Studies of motivation and attention", American Psychologist, vol.50, no.5, pp. 372-385, May 1995.