

비정형자료로부터의 평화지수 분석을 통한 한반도 정세 파악 방법*

권오병** · 박다솔*** · 최지혜*** · 이재운***

Interpretation and Prediction of Situations on the Korean Peninsula by Peace Index Analysis from Unstructured Data*

Ohbyung Kwon** · Dasol Park*** · Jihye Choi*** · Jaeyoon Lee***

■ Abstract ■

Since acquiring intelligence about political situations around the Korea Peninsular in a direct manner is nearly impossible, it is inevitable for the individuals or companies to rely on open and indirect data such as newspapers. However, since the contents in the newspapers are substantially unstructured and very large, conventional content analysis is time-consuming and hence very costly. Hence, this paper aims to propose a sentimental analysis method which computes daily 'peace index' from unstructured data in the newspapers. From the content analysis, words and phrases which represent the sentiment of a nation are carefully identified. To show the feasibility of the idea proposed in this paper, a prototype system with vocabulary repository about political situations was developed for estimating peace index automatically.

Keyword : IT Solution, Sentiment Analysis, Peace Index, Unstructured Data, Content Analysis

논문투고일 : 2013년 10월 26일 논문수정완료일 : 2013년 12월 18일 논문게재확정일 : 2013년 12월 23일

* 본 논문은 경희대학교 교수학습지원센터의 2013년도 제4회 교수-학생협동학술연구 프로그램에 지원을 받아 연구되었음

** 경희대학교 경영대학 교수, 교신저자

*** 경희대학교 경영대학

1. 서론

한반도 정세는 한국 기업의 주요 거시환경 요인으로서 정확하고 적시적인 분석 및 예측은 한반도 평화 유지뿐 아니라 기업의 성공을 위해서도 매우 가치 있는 작업이다. 일반적으로 정세분석은 공개된 계량자료 등을 기반으로 하여 경제학적, 정치학적, 군사학적, 지리학적 등의 이론을 토대로 다방면으로 이루어진다[4]. 특히 한반도 정세는 다자간 이해관계의 불일치에 의하여 변화무쌍하게 이루어져 왔다. 일례로 2011년 11월 연평도 포격사건은 한국정부의 강경한 대북정책추진으로 남북한간 이해관계가 대립되었을 때 북한은 마찬가지로 이해관계가 부딪히는 미국의 위협으로부터 방어하기 위해 체제 생존적 전략의 일환으로 전략적 이해관계가 일치하는 중국을 파트너로 북중경협에 적극적으로 나서는 행보를 보인다[5]. 이러한 경향은 통일부에서 제공하는 남북교역액 현황에서도 나타나고 있다.

그런데 한반도 정세분석에 필요한 정보는 고도의 제한 사항이므로 개인이나 기업 경영 차원에서 확보하기란 불가능하거나 매우 비용이 드는 것이어서 간접적인 자료를 통해서 분석해야 한다. 특히 분석 기간을 좁혀서 주단위 혹은 일단위로 분석하기 위해서는 연구보고서나 책자의 형태는 불가능하고 언론 기사나 소셜미디어 정보 등 갱신 주기가 짧은 정보원천을 활용할 수밖에 없다. 또한 일반 언론 상에서 노출되는 각 지도자들의 발언들로부터도 각 국가별 이해관계의 양상과 이에 따르는 한반도 정세의 미묘한 변화를 감지할 수 있다. 대표적인 예로 2008년 이명박 대통령 중국 방문 시 중국 외교부 친강 대변인이 ‘한미군사행동은 냉전의 유물’이라고 한 발언이 기사화된 것을 통해 당시 중국과 한국 및 미국 간, 그리고 중국과 북한 간의 이해관계를 파악할 수 있다[5]. 이러한 이유로 실제로 남북관계의 변화나 남북의 현안, 주변 국가들의 반응에 따라 남한에 대한 북한 언론의 논조가 변할 것이라는 가정아래 언론에 소

개되는 보도 내용[1, 7, 10]이나 보도 프레임[9]을 통해서 질적 분석을 통해 정세 분석을 하는 연구들이 있어왔다. 이러한 연구 접근법은 북한 언론학자들이 예전부터 북한신문에 대해 남한 및 국제 정세자료를 반드시 포함해야 한다고 천명해온 것을 염두에 둘 때[2], 질적 연구 결과가 적지 않은 신빙성이 있을 것으로 보이는데 기인한다. 하지만 이들 연구들도 짧은 주기로 언론 내용을 인식하여 정세분석을 할 수 있는 틀을 제시하지는 못했다. 결국 간접자료로부터 유용한 정보를 추출함에 있어서 자료 자체가 가치가 있음에도 불구하고 자료의 비정형성으로 말미암아 분석 정확도와 효율성을 제고하기란 어려운 일이다.

따라서 본 연구의 목적은 개인 혹은 집단의 감성을 분석하는 방법(sentiment analysis)을 기본으로 언론기사 자료와 같이 간접적이지만 정세이해에 유용하고 기업이나 개인이 수집 가능한 자료를 바탕으로 동북아 이해관계집단의 현재 상태를 파악하고 미래 일정 시점의 상태를 예측할 수 있는 방법을 제안하는 것이다. 이를 위해 감성분석을 응용한 평화지수(Peace Index) 계산법을 개발하였다. 여기서 평화지수란 동북아 정세가 전쟁이나 국지도발이 억제되는 정도를 의미한다. 그리고 감성분석이란 텍스트에 표현된 의견, 감성 또는 감정을 계산하는 방법으로서[14], 정치 분야 뿐 아니라 교육, 마케팅, 온라인 쇼핑 등 방대한 분야에서 유용한 분석 방법으로 알려져 있다[12, 15]. 즉, 평화지수 및 그 정도에 해당하는 어휘를 내용분석을 통해 파악한 후에 어휘 저장소를 구축하고, 이를 통해 매일 생성되는 기사자료를 기반으로 한국과 북한, 미국, 중국 등 관련 국가의 대 북한 의견을 습득, 평화지수를 계산한 후에 각 국가 간 연관도 분석 등을 수행하였다. 또한 미래의 북한의 대외 평화지수 예측의 토대가 되는 인과관계 모형을 자동으로 수립하는 방법을 제안하였다. 본 방법의 성능을 측정하기 위해 실제로 2013년 3월부터 5월까지의 북한관련 기사를 가지고 실험을 수행하였다.

기사에 대한 내용분석 결과를 통해 정세분석 도

메인의 어휘 저장소를 구축했으며, 이 어휘 저장소는 데이터베이스 형태로 저장되어 본 연구에서 개발한 프로토타입에서 자동방식으로 평화지수를 계산하는데 사용했다.

2. 한반도정세 기사내용 분석

평화지수 계산에 의한 정세분석을 위해 본 연구는 한국 내 발행 신문에 게재된 북한관련 기사에 대해 내용분석을 수행하여 특정 국가의 북한 관련 언급이 기사화된 후에 북한의 반응이 어떻게 변화하는지를 파악함으로써, 각 국가 간 동향이 북한의 행동에 어떤 영향을 주는지를 분석하기 위해 먼저 내용분석을 실시하였다. 내용분석이란 분석 대상이 되는 내용을 객관적/수량적으로 분류하고 이 기준에 입각하여 체계적으로 분석하는 조사방법론이다[6]. 김영주[1]의 연구에서도 북한의 기사 분석을 위해 기사의 논조를 매우 긍정에서 매우 부정까지 5단계로 분류한 바 있다. 그러나 그의 연구에서는 기사의 논조를 판단하는 근거를 제시하지 않고 있어 객관성 확보가 어렵다. 이 경우 내용분석을 실시하여 질적 분석의 약점이 될 수 있는 객관성 확보 문제를 해결해야 한다. 더구나 본 연구에서의 내용분석의 또 다른 목적은 평화지수 계산법에 등장할 어휘(vocabulary)나 어구(phrase)를 확정하고 정세분석 도메인의 어휘 저장소를 구축하기 위함이다.

그 동안 북한을 비롯한 한반도 정세 분석을 위해 사용되어 온 원천자료는 각 당국의 공식 언명(official statements)이나 군사제안(military proposals) 등을 활용하는 문헌 중심의 내용분석 방법이었다[8]. 그러나 본 논문에서는 다음과 같은 이유로 북한의 한국에 대한, 그리고 한국 및 주변국가의 북한에 대한 언급이 포함된 기사들을 분석하는 방법을 선택했다. 그 이유는 첫째, 신문기사는 모든 사람들이 자유로이 접근할 수 있는 공개된 자료이며, 둘째, 기사의 내용이 제목의 형태로 잘 요약되어 있기 때문에 내용 파악이 쉽고, 셋째, 언론사에서 관

심 있어 할 만한 언급 혹은 반응을 보이는 군사, 외교 관계자나 기관들이 자신의 행동이나 발언수위를 결정할 때 서로 상대방의 언론 기사를 참조할 개연성이 높기 때문이다.

2.1 자료 수집

자료조사 대상으로 한국 일간지 기사를 선정하여 최대한 대표성과 신뢰성을 확보하였다. 기사 검색 기간은 한반도 긴장과 완화의 변화 정도가 심했던 2013년 3월 초부터 5월 말로 한정하였다. 그리고 기사 분석이 되는 주체는 한반도 정세에 대한 관심도가 높은 한국과 북한, 미국, 중국, 기타(EU 및 UN)로 한정하였다. 수집 대상이 되는 신문으로 발행부수가 가장 많은 조선일보, 중앙일보, 동아일보 중 북한에 관한 기사를 따로 관리하고 있는 '중앙일보 북한네트(<http://nk.joins.com/>)'를 택했다. 더 많은 신문을 참조하지 않은 것은 각 국가의 성명 혹은 행동은 사실이지 각 신문사의 편집방침이나 논설이 아니므로 신문사 간 차이가 거의 없기 때문이다. 자료는 먼저 중앙일보 북한 네트워크를 통해 북한과 관련된 기사를 모두 모은 뒤, 북한과 대북한의 정세에 영향을 주는 기사를 다시 추출하였다. 추출된 기사는 날짜 별로 각 기사의 지면, 대제목, 소제목, 주제로 범주를 나누어 평화지수(peace index) 및 그 위협 용어의 강도(intensity)에 대한 점수를 매겼다. 평화지수와 강도 계산에 있어 대제목과 소제목만을 참고한 것은 대제목과 소제목들만으로 충분히 기사의 성격을 반영할 수 있었기 때문이다. 기사 본문은 대부분의 경우 기사의 제목과 같은 취지의 평화지수를 보일 것으로 보여 중복이므로 분석에서 제외하였다.

2.2 어휘 선정

평화지수에 영향을 주는 어휘들을 선정하는 원칙은 한양대 아태지역연구소에서 한반도 평화지수 산정 시 채택되고 있는 가이드라인(http://www.hanyang.ac.kr/code_html/H5EAKB/indexC2.html)

을 준수하여 문헌으로부터 국제정세평화, 개선, 긴장완화 또는 전쟁, 위협, 악화, 긴장조성이라는 용어와 함께 등장하는 동사 및 형용사 어휘들(예 : 응징, 대응, 대화, 무대응, 감시 등)을 찾았다. 한편 북한의 경우에는 용어가 특이하기 때문에 북한 언론매체에서 즐겨 사용하는 ‘괴뢰’, ‘반역’ 등 북한식 악평어(bad label)도 참조하였다[1]. 그리고 내용분석 담당자들이 수차례에 걸친 조정회의를 통해 각 용어에 대해서 -3에서 3점까지 점수를 결정했다. 추가로 신문 기사 제목 분석 시에 새로운 용어가 등장할 때는 기준 어휘를 근간으로 하여 상대적으로 강약을 평가하여 그 용어의 평가지수를 정했다. 한편 명사형도 평화지수에 포함했는데 특별히 전쟁과 평화를 상징하는 물건(예 : 핵무기, 미사일, 친서), 장소(예 : 진지, 부대 등), 사람(예 : 장군, 대사 등)을 고려했다.

2.3 기사별 평화지수 계산

다음으로 대제목과 소제목을 참조하여 평화지수를 계산하였다. 위협 또는 안심은 감성(sentiment)의 일종이므로 평화지수 분석은 감성분석의 특수형으로 간주할 수 있다. 따라서 본 연구에서는 먼저 내용분석에 의하여 수집된 위협을 표현하는 언어와 정도부사를 통해 어휘 저장소를 구축한다.

어떤 특정 기사 d 는 시간(t), 발언 국가(??), 대상 국가(???), 그리고 i 번째 기사의 j 번째 어구에 대한 위협어 및 정도부사 쌍의 집합으로 이루어진다.

$$d_i = \langle t, C_A, C_B, \{(w_{ij}, a_{ij})\} \rangle \quad (1)$$

$\{(w_{ij}, a_{ij})\}$ 로 하는 이유는 한 문장에도 복수개의 의견 혹은 감성이 포함될 수 있기 때문이다. 그리고 발언 국가의 평화지수 ν_A 는 $\{(w_{ij}, a_{ij})\}$ 의 함수로 이루어진다.

식 (1)에 따라 먼저 각 동북아 정세 관련 기사의 제목이 어떤 성향(강경, 완화)을 지니는지에 따라

7점 척도(-3에서 3까지)를 사용하여 점수를 매겼다. 즉, 3점에 가까울수록 완화를 의미하며 -3점에 가까울수록 강경의 입장이라 정의하였다. 또한 행동적 성격을 지닐 경우 1.0, 계획의 성격을 지닐 경우 0.8, 그리고 태도의 성격을 가지는 경우에 대해서는 0.6정도의 강도를 곱했다. 예를 들어 개성공단 재개가 2점이라면, 개성공단 재개 의사를 밝히는 것은 계획의 성격으로 0.8을 곱한 1.6점, 개성공단 재개에 대해 환영한다면 태도의 성격으로 0.6을 곱한 1.8점이 된다. 이렇게 되면 실제로 3월 6일 한 일간지의 종합 면의 1면에 게재된 ‘안보리 제재 앞두고 북, 정진협정 백지화 국지도발 가능성도’이라는 대제목과 ‘천안함 공격한 김영철 발표 미그기도 최전방 진전배치’라는 소제목의 기사에서는 강경의 입장을 취하고 있으며 그 입장이 행동적 성격을 지니고 있기에 5점 척도 중 -2점을 줄 수 있다. 반면, 4월 17일 국방/외교면의 10면에 게재된 ‘위협 수위 낮춘 북 ... 긴장국면 출구 찾나’라는 대제목과 ‘주한미군 고위 관계자 분석/ 최근 들어 원색 비난 줄고 “대화 원하면 ...” 조건문 사용’이라는 소제목의 기사에서는 완화의 입장을 취하고 있으며 그 입장이 발언적 성격을 지니고 있기에 5점 척도 중 2점에 계획의 성격으로 0.8을 곱한 1.6점이 매겨진다.

또한 평화지수 관련 어휘의 강도에 영향을 주는 보조 어휘에 대해 결정하였다. 보조 어휘는 일종의 승수(multiplier)로서 정도부사의 유의어 연구에서 정도부사의 하위분류와 등급설정 부분을 참조했다[11]. 특히 등급 설정을 정도의 크기에 따라 분류한 김영희[3]의 분류에 초점을 두었다. 정도의 크기는 0.5부터 2까지 차등 분류했으며 종류별로는 정도부사의 형태(예 : 매우, 완전, 약간 등), 가능성(예 : 반드시, 실제로, ~듯, ~일지도 모르는, 시사), 부정 혹은 취소(예 : ~하지 않는, ~는 아닌), 의문형(예 : ~?), 평화지수 관련 명사를 수식하는 형용사(예 : 단호한, 강력한, 전면적인, 대폭적인 등), 시간의 긴박성을 알리는 어휘(예 : 전격적, 곧) 등으로 나뉘었다. 새롭게 등장하는 어휘들은 내용분석

수행자들의 합의에 의하여 기준 점수를 보면서 조정했다.

또한 하루에 다양한 동북아 정세에 대해서 복수개의 기사가 나가는 경우도 있었는데 이때에는 한 기사의 보조 기사인 경우에는 같은 강도를 보일 것으로 보고 생략하고도 여러 기사들이 있는 경우에 대해서만 각각 제목 상의 평화지수 관련 어휘들에 대해서 평화지수를 결정한 후에 두 기사의 지수 값을 합산하였다. 가중합산은 지나치게 내용 분석가들의 주관에 허용하게 되므로 고려하지 않았다.

한편 한 기사에 여러 국가의 발언이 등장하는 경우가 있다. 예를 들어 3월 6일자의 “(미국) 북돋줄 포함 제재 강화. 중국도 동의”에는 미국과 중국의 북한에 대한 태도가 동시에 담겨있다. 이 경우는 각 국가의 입장 별로 분리를 한 다음 각각 평화지수를 계산하게 된다.

2.4 일관성 검증

분석 작업은 세 사람이 각자 독립적으로 작업을 수행하고, 완성 후 회합하여 그 각자의 평화지수에 대한 일치도를 파악하여 일정 오차 이상의 기사에 대해서는 상호 협의를 통해 일치시켜나갔다. 이러한 과정을 통해 모든 기사에 대해 오차 없이 일치된 평화지수와 강도를 정했다.

2.5 일별 평화지수 계산

하루에 복수개의 기사가 등장할 수 있다. 따라서 기사별 평화지수를 계산한 후에는 일별로 평화지수를 계산해야 한다. 계산한 결과의 예는 <표 1>과 같다. 여기서 평화지수는 해당일의 각 기사들의 승수와 점수를 곱한 수치의 평균이다. 단, 기사가 없었던 3월 10일의 평화지수가 -2.3으로 기록된 것은, 특정일에 기사가 없을 때 그 전날의 기조가 그대로 유지되는 것으로 보았기 때문이다. 다만 추후에 결측치(missing value) 처리 방법에

따라 직전 평화지수와 직후 평화지수의 평균값으로 대체하는 것도 고려할 만하다.

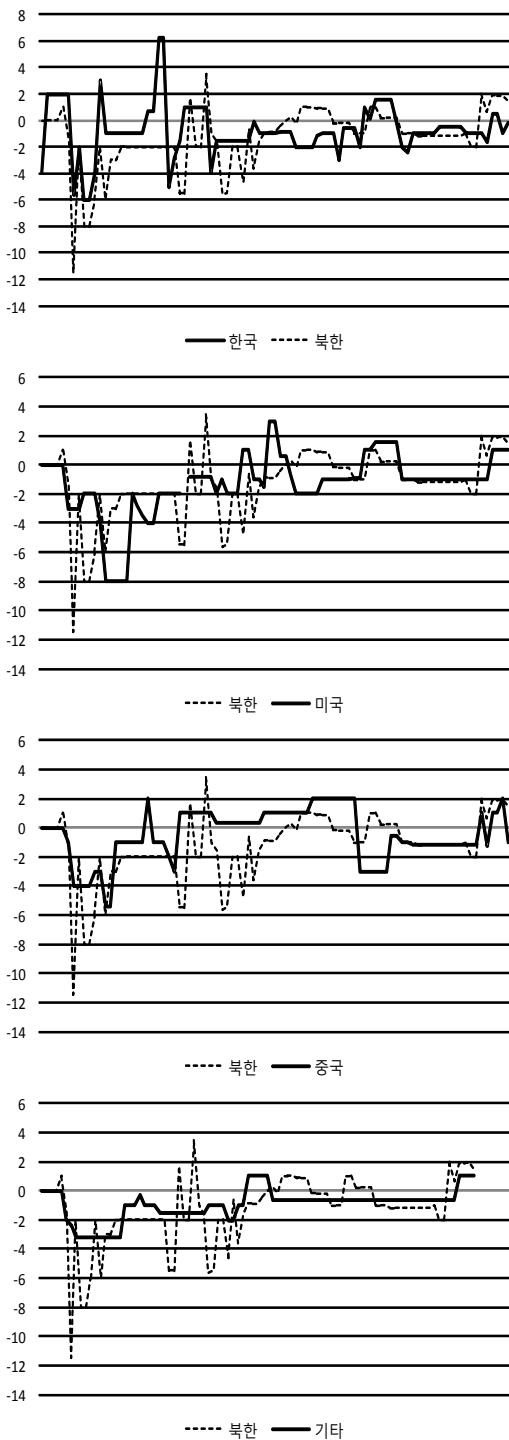
<표 1> 기사 분석의 예

날짜	대제목	보조 어휘	승수	평화지수 어휘	점수	평화지수
3/1	윤병세 “북핵 불용 확고 ...”	확고	2	북핵불용	-2	2.05
3/1	군사적 제재는 검토 안 해”	NEGATE	0.1	군사제재	-1	
3/2	정부, 북 핵실험에도 인도적 지원한다	없음	1	인도적 지원	2	2.0
중 략						
3/9	박 대통령 강경에	없음	1	강경	-2	-2.3
3/9	안보위기 고조되자 ...	고조	1.5	안보위기	-2	
3/9	핵공격 땀 김정은 정권 소멸	없음	1	정권 소멸	-2	
3/10	기사 없음					-2.3

3. 평화지수 기반 정세분석

3.1 시각적 분석

내용분석을 통해 확보한 사례를 통해서 먼저 [그림 1]과 같이 시각적 분석을 수행하였다. [그림 1]에서 가로축은 날짜, 세로축은 관련 기사의 평화지수가 일별로 합산된 값이다. 먼저 남북한은 전체적으로 추세가 비슷하게 흐르는 경향이 있다. 그러나 세 차례 정도는 정반대로 흐르는 경우도 보인다. 이중 두 차례는 한국이 완화 정책을 펼 때 북한은 강경책을 유지하는 북한공세 국면이었다. 이는 한국이 완화책을 펴도 북한은 핵을 고수하는 강경한 상태를 유지한 경우가 이때공씩 발생한다는 경험과 일치하는 결과이다. 하지만 전체적으로는 한국과 북한 사이에는 거의 시차 없이 대등한



[그림 1] 신문기사 제목에 근거한 각 주체간 정세 (가로축은 날짜, 세로축은 평화지수)

대응을 해오고 있는 것임을 보여주고 있다.

북한과 미국과의 비교를 보면 한국과는 달리 며칠씩의 시차를 두어서 반응이 일치하는 경향을 보인다. 그리고 북한의 반응이 크게 요동하는 것에 비해 미국은 상대적으로 일관된 태도를 견지해오는 것으로 보인다.

북한에 대한 중국의 반응에서는 전체적으로는 북한의 공격적, 호전적 입장에 비하여 주변 국가보다는 상대적으로 더욱 온건한, 즉 북한을 옹호하는 듯한 반응을 보이는 현상이 두드러진다. 그리고 세 차례 정도 양국 간 간격의 차이를 보이는데 이는 중국의 국가적 위상이 커짐에 다른 나라의 의견을 수용해주는 현 추세를 반영한 결과라 볼 수 있다. 그 예로 북한이 최룡해 특사를 파견하며 완화 정책을 폈을 때 중국 측에서 냉대했던 사례와 안보리 제재에 대한 동의 등을 볼 수 있다. 한편 북한과 기타 국가의 추세는 북한이 강경책을 펴면 적대적으로, 완화책을 펴면 우호적으로 받아들이는 분위기를 보이며 비슷한 추세를 띄고 있다. 그리고 다른 이해관계자들에 비해서 평화지수의 변동이 심하지 않은 것을 볼 수 있다. 이는 북한의 반응에 대한 영향이 상대적으로 적게 미치기 때문에 한반도 주변국들보다는 덜 민감하기 때문인 것으로 보인다.

[그림 4]는 국면의 변화 양상 분석을 남북한의 예를 들어 보여주고 있다. 요약하면 3월에는 북한 공세국면이었으나 4월 들어서 그 공세가 줄어들고 도리어 한국이 공세적이었으며 5월에 들어서는 전체적으로 화해국면으로 변경되는 것을 볼 수 있다.

3.2 집단 간 평화지수 비교

분석 결과 각 이해관계자 집단 간 평화지수의 차이가 있는지를 분석하기 위해 먼저 평균비교를 분석한 결과 <표 2>와 같이 전체적으로 기사는 부정적이었으며 부정의 강도는 북한이 가장 강한 반면 주변 나라들 중에서 북한에 대해 중국은 가장 중간적 입장을 견지하는 것으로 나타났다.

〈표 2〉 평균 비교

국가	평화지수 평균	표준오차
한국-북한	-.74	2.064
북한-타국가	-1.33	2.499
미국-북한	-1.30	2.213
중국-북한	-.48	1.752
기타-북한	-1.11	1.038

한편 각 이해관계자 사이의 평화지수 사이의 차이가 있는지를 통계적으로 검증하기 위해 T-검정을 실시하였다. 그 결과 <표 3>에서 확인되는 것처럼 한국과 중국은 북한에 대해서 평화지수가 높다. 이는 한국이 분석 기간 동안은 북한의 공격적 발언에도 불구하고 온건한 반응으로 응대했고, 중국의 경우에는 북한을 옹호하는 반응을 보였기 때문인 것으로 보인다.

〈표 3〉 T-검정을 통한 국가 간 평화지수의 차이 분석

비교 대상	차이 평균	표준편차	t
한국-북한	.584	2.540	2.170**
미국-북한	.022	2.615	.081
중국-북한	.843	2.301	3.455***
기타-북한	.213	2.161	.932

단, ** p < 0.05, *** p < 0.01.

3.3 관계 분석

먼저 각 국가 간 기사의 평화지수간 관계를 파악하기 위해서는 상관관계 분석을 시도할 수 있다. 그러나 각 국가의 북한 관련 발언을 토대로 한 평화지수 값이 같은 날 북한 발 기사의 평화지수의 값에 영향이 있는지를 파악하기 위해 먼저 일원배치 분석을 <표 4>와 같이 실행한 결과 통계적으로 유의하게 차이가 나는 것으로 나타났다. 일반적으로 이해관계에 놓인 국가들은 주변국의 반응에 대응하는 경향을 고려할 때, <표 4>의 결과는 북한을 둘러싼 발언에 대해서 동일한 날짜의 다른 나라의 발언 기사를 가지고 분석하는 것은 의미가

없음을 의미한다. 그래서 주변국의 발언에 대해서 시차를 두어서 반응할 것이라는 판단을 하게 되었다. 결국 같은 날짜에서 발생한 평화지수 값 사이의 상관관계 분석은 의미가 없으며 상대국의 반응에 대해 일정 기간이 경과한 후 대응하는 것으로 보여 시차 상관관계 분석을 실시하기로 했다.

3.3.1 시차 상관관계 분석

임의의 두 그룹 A, B의 평화지수인 ν_A 및 ν_B 사이의 상관관계 분석을 통해 두 국가 간 한 대상에 대한 견해 및 전략의 일관도($\rho(\nu_A, \nu_B)$)를 파악한다. 그런데 각 그룹의 반응은 당연히 상대 그룹의 입장에 영향을 받으므로 현재 시점의 ν_A 및 ν_B 는 각각 과거의 ν_B 및 ν_A 의 함수라고 볼 수 있다. 따라서 반응의 시차가 존재할 수 있으므로 시차를 둔 상관관계 분석을 수행한다. 즉, Ω_A 를 A국가의 평화지수 집합이라고 할 때,

$$\Omega_{A,L} = \{\nu_{A,t-L}, \nu_{A,t-(L+1)}, \dots, \nu_{A,1}\} \quad (2)$$

와 같이 표현된다. 이때 L은 시차(lag), 1은 관측의 최초시점이며, t는 현재시점을 가리킨다. 그러면 $\rho(\nu_A, \nu_B)$ 은 $\Omega_{A,0}$ 과 $\Omega_{B,L}$ 사이의 L만큼의 시차를 둔 기사들 사이의 상관관계를 의미한다. 이때 L값을 0에서부터 증가시켜가며 의미 있는 상관관계가 나오는 L값의 존재유무를 분석한다. 특히 그 값의 음양 여부에 따라 두 그룹의 대응 정책의 유형을 파악한다. 그리고 음양 여부의 변동은 정책구조의 변화라고 해석할 수 있다. 시차 상관관계를 분석한 결과는 <표 5>와 같다. 시차를 둔 것은 북한에서 언급한 내용을 다룬 기사로서 t = N은 N일 후 기사를 의미한다. 예를 들어서 t = 0은 동일 날짜의 기사, t = 1은 하루 후 기사가 된다. <표 4>의 내용을 보면 한국에 대해서는 매우 신속하게 반응을 보이고, 미국에 대해서는 비교적 긴 시차를 두어서, 그리고 중국과 미국에 대해서는 그 중간 정도의 반응을 보이는 것으로 나타난다.

〈표 4〉 동맹국간 및 북한과의 시차 상관관계

	한국	미국	중국	기타
한국	1	.142	.095	.236*
미국	.142	1	.261*	.633**
중국	.095	.261*	1	.488**
기타	.236*	.633**	.488**	1
북한 t = 0	.393**	.389**	.459**	.511**
북한 t = 1	.164	.378**	.413**	.517**
북한 t = 2	.130	.388**	.328**	.463**
북한 t = 3	.087	.380**	.327**	.440**
북한 t = 4	-.014	.377**	.251*	.354**
북한 t = 5	-.174	.333**	.115	.245*
북한 t = 6	-.090	.458**	.080	.195
북한 t = 7	-.027	.487**	.057	.087

단, * p < 0.1, ** p < 0.05, *** p < 0.01.

3.3.2 평화지수 상의 일치성 수준 간 원인분석

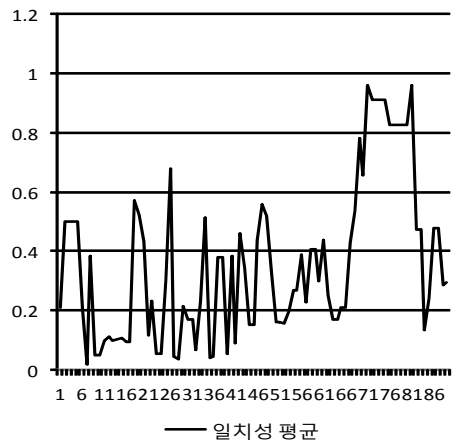
일치성 수준이란 임의의 두 그룹의 평화지수가 얼마나 유사한지의 정도를 의미한다. 임의의 두 그룹의 평화지수 쌍은 각 그룹의 평화지수가 ‘높다(H)’, ‘낮다(L)’로 양분했을 때, [그림 2]와 같이 2x2 행렬로 분류할 수 있다. 중간 평화지수 값은 0이다. 이때 (H, H)쌍은 높은 수준에서 일치하는 경우이므로 일촉즉발의 위기국면, (L, L)은 화해국면, (H, L)과 (L, H)는 각각 한쪽 국가의 공세 국면이라고 해석한다. 이때 네 국면에 해당하는 사건 또는 맥락들을 비교 분석하면 각 국가의 핵심 가치와 위기관리 전략을 이해할 수 있다. 또한 향후 특정 그룹의 정세 흐름에 대한 예측도 가능한데 예를 들어 북한과 중국 사이의 평화지수 행렬 상의 위치를 보아 만약 북중관계가 공고화되는 것으로 분석되면 향후 북한의 체제생존이 유리하다는 예측이 가능하다[4]. 또한 일치성 측면을 보면 화해국면과 위기국면은 일치성의 수치가 상대적으로 높은 상태를 의미한다. 일치성이 크면 장기적으로 지속될 경향도 있을 것으로 보인다. 반면 일치성이 작아지는 공세국면에서는 균형이 깨어졌으므로 일치성을 높이려는 외교적 정치적 노력을 할

것으로 예상할 수 있다.



[그림 2] 평화지수 일치성 행렬

일치성 분석 결과 3월 1일부터 5월 28일까지는 다음 [그림 3]과 같은 결과를 얻었다. 여기서의 일치성은 북한과 다른 네 주체들 사이의 일별 평화지수의 차이에 대한 RMSE(Root Mean Square Error)값이다. 그림에서 보면 일치성이 시간이 경과할수록 높아지는 것을 확인할 수 있다. 이는 3월 초에 북핵 실험 등 북한의 돌출적 행동에 의하여 국가 간 반응의 차이가 많이 났었고, 이에 대해서 주변 국가들의 대응을 통해 돌출적 상황이 줄어든 것을 볼 수 있다.



[그림 3] 일치성 변화

3.4 다중회귀분석

북한의 반응에 영향을 미치는 요인을 분석하는 또 다른 대안으로 회귀분석을 사용할 수 있다. <표 5>는 북한의 반응을 예측하기 위해 한국, 미국, 중국 및 기타지역의 반응을 독립변수로 보았다. 그 결과 북한의 반응은 한국과 중국, 그리고 기타지역의 반응에 대해서 민감하게 움직이는 것으로 나타났다. 재미있는 것은 미국에 대해서는 유의하게 반응 결과가 나오지 않았다. 이는 시차 상관관계에서 북한은 독특하게도 시차가 커질수록 미국의 과거 반응에 대해서 관계성이 높아지는 것과 관련이 있다. 즉, 미국의 반응에 대해 북한은 당장의 반응을 보이지는 않는다는 것이다. 그러나 한국과 중국에 대해서는 즉각적인 반응을 보이는 것으로 보인다.

<표 5> 북한 반응 예측 회귀분석 결과(t = 1)

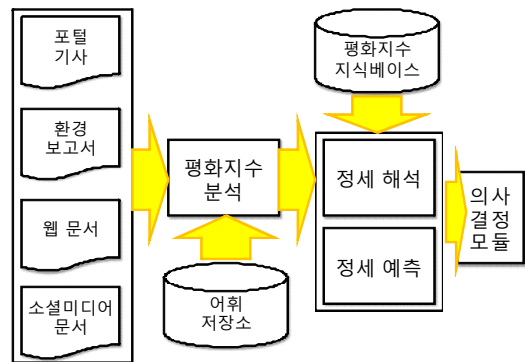
변수	계수	표준오차
(상수)	-0.089	0.313
한국반응	0.358***	0.104
미국반응	0.153	0.122
중국반응	0.416***	0.137
기타지역 반응	0.511*	0.293

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4. 구현

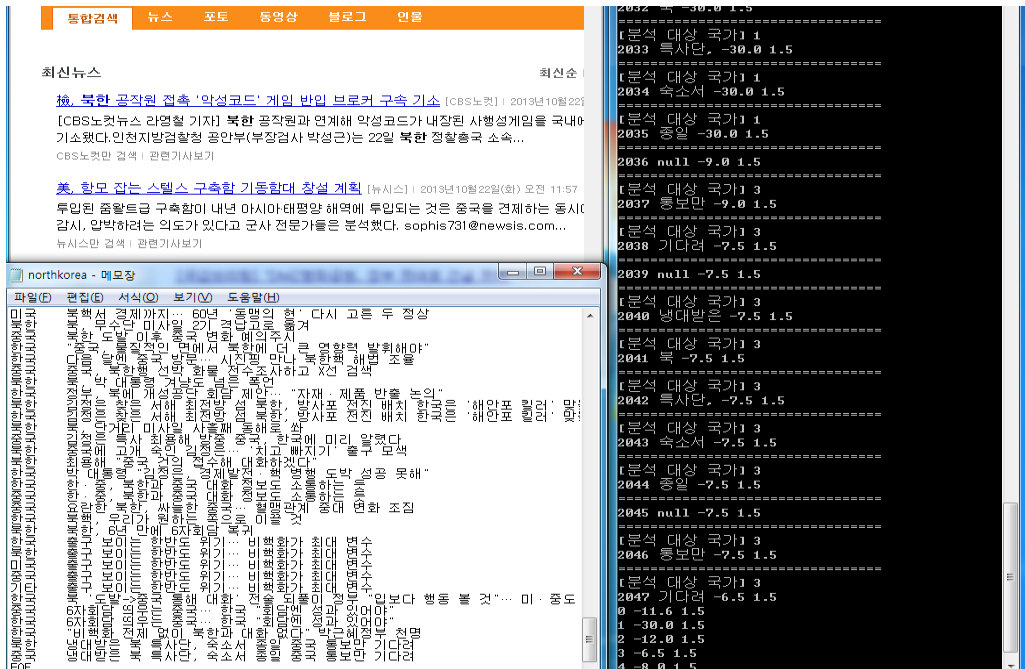
이상과 같은 기사 자료 수집 방법 및 정세 분석 방법의 성능을 자동 분석하기 위해 다음 [그림 4]와 같은 시스템 아키텍처를 제안하였다. 그 이유는 정세분석을 위한 평화지수 계산을 수동으로 수행하는 경우 노력의 양이 상당하며 정세분석의 주기가 짧아질수록 그 어려움은 가중되기 때문이다. 실제로 Hai의 붓스트래핑 접근법이나 Pang and Lee의 minimum cuts 방식은 문장 단위의 감성분석을 자동 수행하기 위한 노력의 산물이다[13, 16]. 먼저 현재 정세에 관련된 글을 포털 기사나 환경 보고서, 웹 문서, 소셜미디어 문서 등 공개적이고

전자화된 자료에서 획득하여 파싱(parsing)하게 된다. 파서(parser)는 글에 등장하는 어휘들을 인식하게 되는데 이들 중에서 어떤 어휘가 평화지수 계산에 관련된 어휘인지를 판단하는 것은 어휘저장소를 참고한다. 즉, 평화지수에 관련된 어휘들은 어휘저장소에 보관하게 되며 필요한 경우 검색하여 평화지수 계산에 사용한다. 평화지수 계산은 앞서 기술한 계산 방법을 따르되 정기적, 비정기적으로 승수와 용어의 평가지수 강도 등은 판별분석이나 다중회귀분석, 사례기반추론, 또는 이들의 혼합 분석 등 데이터마이닝 기법을 통해서 조정이 가능하다. 평화지수 지식베이스에는 앞의 회귀분석, 시차 상관관계 분석 등에서 나온 지식을 토대로 하여 한국 등 여러 나라의 반응에 대해서 북한이 언제 어떤 정도의 반응을 보이는지 예측하거나, 북한이 보인 반응이 과거의 어떤 국가의 어떤 반응에 의하여 발생된 것인지에 대한 원인 분석에 사용된다. 이렇게 추론된 정세 해석과 예측을 기반으로 하여 북한의 반응에 어떻게 대응할지에 대한 의사결정을 수행하게 된다.



[그림 4] 시스템 아키텍처

[그림 5]의 아키텍처에 맞게 프로토타입 시스템은 IBM PC에서 Java Application SDK로 구현되었다. [그림 5]에서 보면 왼쪽 상단과 같은 웹 문서에서 ReadURL 클래스를 통하여 HTTP 형태의 문장을 읽어서 텍스트파일의 형태로 저장한다. 그런 후에 기사의 발원지와 기사 제목만을 추출하여



[그림 5] 기사자료로부터의 평하지수 계산 구현 스크린

좌측하단과 같이 포맷화된 자료를 생성한다. 생성된 문서를 Java에서 제공하는 Scanner 클래스를 재사용하여 토큰 단위로 입력하고 어휘베이스에 있는 평하지수 계산 관련 형용사 부분과 정도부사를 읽어 들인 다음 각 기사 별로 평하지수를 계산한다. 계산 수행과정에 대한 시스템 스크린은 우측과 같다. 최종적으로 계산한 평하지수는 발언 국가별로 따로 집계가 된다. 그리고 집계된 결과를 활용하여 정세 해석과 예측을 수행한다.

5. 결 론

개성공단, 금강산 개발 이슈 등에서 볼 수 있듯이, 한국의 기업 경영 환경은 과거 한국과 북한, 그리고 주변 국가들 사이의 동적이고 복잡한 정세에 의하여 규정되는 경우가 빈번하다. 그러나 이해관계자들은 이들 국가들 사이의 정세변화를 정확히 수집, 예측하는 것은 매우 많은 비용이 들뿐더러 그러한 직접적인 정보는 접근 가능하지도 않다. 따

라서 기업 입장에서는 공개된 대규모의 데이터에서 정세 변화를 추론하는 방법을 고려해야 하는바, 본 연구에서는 평하지수 계산법을 제안하고 이를 바탕으로 다양한 정세분석을 수행하였다.

기존의 감성분석 연구가 대부분 개인 차원의 분석을 대상으로 한 것에 비해 본 연구의 가장 큰 공헌은 평하지수에 근간한 국가차원의 감성분석(nation-wide sentiment analysis)을 시도한 것이다. 국가 차원의 감성분석으로 국가 기간에 다양한 기사를 실시간으로 모니터링 하여 적절한 대응을 할 수 있는 능력을 제공할 수 있다는 점은 그간의 정세분석 방법과 차별화된다.

후속 연구의 필요성이 있는 부분은 다음과 같다. 첫째, 본 연구에서 제안한 비정형 데이터 기반의 정세분석 방법은 모형의 실효성을 파악하기 위해 실제 분석 및 진행한 데이터의 양은 3개월(2013년 3월~5월) 분량이다. 이보다 많은 양의 데이터를 대상으로 계속 데이터 모델을 구축한다면 더 유의미한 결과를 도출해 낼 것이라 생각한다. 또한 본

연구는 데이터 구분과정에서 데이터를 정치, 경제, 사회로 구분했다. 하지만 북한의 정치적 동향을 우선적으로 파악하기 위해 주로 정치 부분을 참고했다. 추후 연구로 다른 부문과 연계된 분석이 이루어진다면 다양한 시각으로 북한의 동향을 파악할 수 있을 것이다. 둘째, 본 연구의 분석 대상은 기사에 할당되어 있는 기사제목으로서 문장 단위 감성분석(sentence-level sentiment analysis)이고 감성 어휘를 사전에 내용분석 등을 통해 확보하는 것으로 지도학습(supervised learning)에 해당한다[12]. 하지만 향후에는 자율학습(unsupervised learning)으로 사전에 제공하지 않는 어휘에 대해서도 평화지수 계산이 가능하도록 해야 한다. 그럼에도 불구하고 본 연구에서 도출한 평화지수 기반의 정세 분석 방법은 향후 특정 대상국가군들에 대해 언급된 기사, 뉴스, SNS상의 언급 등 대규모 데이터를 자동으로 수집, 정세분석을 할 수 있게 되며 이를 통해 나오는 결과물로 기업의 위기관리 또는 새로운 기회와 위협을 고려한 경영전략을 적시적으로 수립하는 데 많은 기여를 할 것이다.

참 고 문 헌

- [1] 김영주, “로동신문에 나타난 대남보도 논조 분석 : 2008년 이후를 중심으로”, 『언론과학연구』, 제13권, 제4호(2010), pp.80-121.
- [2] 김영주, 이범수, 현대북한 언론의 이해, 서울 : 한울아카데미, 1999.
- [3] 김영희, 섹슈말로서의 정도 부사, 한글학회, (1985), pp.133-162.
- [4] 김주삼, “북·중관계의 밀착과 한반도 통일정세 분석”, 『한국동북아논총』, 제61권(2011a), pp. 79-102.
- [5] 김주삼, “북중 경제협력관계 변화와 한반도 통일환경 분석”, 『동북아연구』, 제26권, 제1호 (2011b), pp.79-99.
- [6] 김태우, 백동현, “SNS의 스마트폰 게시물 내용 분석을 통한 사용자의 요구특성 분석”, 『한국전자거래학회지』, 제17권, 제4호(2012), pp. 197-208.
- [7] 선상신, 김위근, “북한의 핵무기 개발과 대외 프로파간다 전략 : <조선중앙통신>과 <조선신보>의 6자회담 보도를 중심으로”, 『커뮤니케이션 이론』, 제8권, 제1호(2012), pp.280-322.
- [8] 양병기, “북한 군사정책의 전개과정과 변화 전망”, 『군사문제연구』, 제13권(1998), pp.167-190.
- [9] 이완수, 배정근, “1994 vs. 2011 : 김일성과 김정일 사후 한국 언론의 북한문제 보도 프레임 변화”, 『커뮤니케이션 연구』, 제29권(2013), pp.149-183.
- [10] 허진, “한국의 정권 변화와 중국 주요 신문의 한반도 관련 보도”, 『언론과학연구』, 제11권, 제4호(2011), pp.557-589.
- [11] 최홍열, 정도부사의 유의어연구, 역락, 2005.
- [12] Feldman, R., “Techniques and applications for sentiment analysis”, *Communications of the ACM*, Vol.56, No.4(2013), pp.82-89.
- [13] Hai, Z., K. Chang, and J.-J. Kim. “Implicit feature identification via co-occurrence association rule mining”, *Computational Linguistics and Intelligent Text Processing*, (2011), pp.393-404.
- [14] Liu, B., “Sentiment analysis and subjectivity”, In N. Indurkha and F. J. Damerau (Eds.), *Handbook of Natural Language Processing*, Chapman and Hall : CRC Press, 2010, pp.627-666.
- [15] Ortigosa, A., J. M. Martin, and R. M. Carro, “Sentiment analysis in Facebook and its application to e-learning”, *Computers in Human Behavior*, to appear in 2013.
- [16] Pang, B. and L. Lee, “A Sentimental Education : Sentiment Analysis using Subjectivity Summarization based on minimum cuts”, In *Proceedings of the Association for Computational Linguistics*, (2004), pp.271-278.

◆ 저 자 소 개 ◆

**권 오 병 (obkwon@khu.ac.kr)**

현재 경희대학교 일반대학원 경영학과에서 교수로 재직 중이며 경영정보시스템 및 빅데이터를 전공하고 있다. 1988년과 1995년 서울대학교 경영대학 및 한국과학기술원 경영학과에서 각각 학사와 박사 학위를 취득하였고, 주요 관심분야는 유비쿼터스컴퓨팅, IT의료 융합기술 및 데이터분석 등이다.

**박 다 솔 (dahapu@naver.com)**

경희대학교 경영대학에 재학 중이며 현재 차세대정보기술연구센터에서 연구원으로 데이터 과학 분야로 연구하고 있다. 주요 관심분야는 이비즈니스, 빅데이터 등이다.

**최 지 혜 (qazwsx123455@naver.com)**

경희대학교 호텔관광대학에 재학 중이며 경제학을 복수 전공 중이다. 주요 관심분야는 관광 분야에서의 IT융합, 전자상거래 서비스 등이다.

**이 재 윤 (lee_jae_yoon@hanmail.net)**

경희대학교 경영대학에 재학중이며 현재 차세대정보기술연구센터에서 스토리텔링을 기반으로 한 역사탐방 모바일 앱개발, 기술수용모델을 중심으로 현 소비자 연구, 비정형자료 분석 등의 연구를 수행 중이다. 주요 관심분야는 빅데이터 등이다.