

나비에 스토크스 방정식은 그 해의 존재성이 확인되었지만 정확한 해를 구할 수 있는 경우는 매우 드물다. 따라서 최근 컴퓨터와 함께 발전하고 있는 CFD는 수치 해석을 사용하여 근사해를 구한 뒤 유동을 해석하는 추세이다.

본 연구에서는 혈관 내 콜레스테롤이 끼어 있는 관 내 유체 유동을 2D N-S 방정식과 CFD를 사용하여 해석을 실행하고자 한다. 본 논문에서는 수학적, 유체역학적 지식들과 CFD의 기초가 되는 이산화 방정식에 대한 지식들을 습득하고 N-S 방정식을 모델링하여 이상적인 관 유동에 대한 지배방정식을 수립하는 것까지 진행하게 되었다.

## Next Step

본 연구에서는 혈관 내 콜레스테롤이 끼어 있는 관 내 유체 유동을 2D N-S 방정식과 CFD를 사용하여 해석을 실행하려 하기 위해 혈관학을 학습하고 이를 바탕으로 혈액의 특징을 상사를 통해 모델링에 적용하므로 좀 더 사실적인 모델링을 구현할 것이다. 또한 코딩을 통해 상용 CFD의 오픈소스와 유사한 코드를 C언어로 개발하여 상용 CFD와 더불어 모델링 검토작업에 이용할 것이며 소스를 공개하여 유사한 연구를 하는 연구자들이 쉽게 사용할 수 있도록 도울 것이다.

## 참 고 문 헌

- [1] Adam Powell. (2004). "The Navier-Stokes Equations"  
[http://texmex.mit.edu/pub/emanuel/CLASS/12.340/navier-stokes\(2\).pdf](http://texmex.mit.edu/pub/emanuel/CLASS/12.340/navier-stokes(2).pdf)
- [2] Davit Martirosyan. (2016). "Large deviations for invariant measures of white-forced 2D Navier-Stokes equation" department of mathematics in cergy-pontoise university
- [3] Frank M. White, 유체역학
- [4] 김원석. (2000). "시뮬레이션을 통한 유체의 흐름과 퇴적물 이동의 수치 해석방법 해법 연구", 석사학위논문, 연세대학교 대학원
- [5] 김영희. (2005). "Navier-Stokes 방정식의 빠른 풀이법과 이를 이용한 실시간 유체시뮬레이션", 박사학위논문, 경북대학교 대학원
- [7] 김지태 외 2. (2013). "혈액 투석시 충혈방법에 따른 인조혈관 내 유동 특성에 관한 수치해석 연구" 한국전산유체공학학회지, v.18 no.3, pp.17-19
- [8] 김태우 외 2. (2011). "외부 유동 해석에 대한 오픈 소스 코드, OpenFOAM의 검증" 한국항공우주학회지, 2011년, pp.702-710

- [9] 명현국. (2004). "원형 실린더 주위의 비정상 이차원 층류유동 수치해석" 한국전산유체공학학회지, no.4 = no.27, pp.41 - 47
- [10] 박민규. (2014). "나비에-스토크스 방정식과 유체역학의 다양한 활용" 한국방재학회지, v.14 no.2, pp.58 - 63
- [11] 명현국. (2012), C F D, 문운당, 서울
- [12] 장태진. (2002). "2차원 비압축성 Navier-Stokes 방정식을 이용한 유체 유발 진동해석" 석사학위논문, 한국과학기술원
- [13] 장석규. (1986). "다공성 물체가 있는 계단수로에서의 유체유동 특성에 관한 전산 해석적 연구" 석사학위논문, 한국과학기술원

## 웹 뉴스의 감성분석과 기계학습을 이용한 주가등락예측 및 예측정밀도 개선

폴수학학교 : 류호성, 이윤수  
지도교사 : 이호현

Investment strategy and predicting technique appeared to analyze pattern of stock market to get economic gain. However, predicting the flow of stock index is quite difficult because stock market contains uncertain factors. To overcome this problem, varieties of methodology is going along. And with 'big data', varieties of atypical data come out with social media. Therefore, in this research paper we predicted fluctuation of stock price with using 'News data'. We used morpheme analysis and sentimental analysis to make digitalize it. Next with this data we applied machine learning and made predicting model. Finally, we got prediction rate and F1 score.

### 1. 서론

주식은 경제적 측면과 밀접한 관계를 맺고 있어 기업의 전문가뿐만 아니라 일반인들에게도 큰 관심의 대상이었다. 더불어 저물가, 저성장, 저금리 시대가 지속되면서 주식 투자 인구가 증가함에 따라 주가예측의 필요성이 점점 부각되어가고 있다. 현재, 주가예측 분야에 기존의 예측 방법론뿐만 아니라 많은 과학기술과 이론들이 도입되어 연구가 진행중인 실정이다. 주가예측 방법론은 크게 2가지로 기본적 분석(Fundamental analysis)과 기술적 분석(Technical analysis)이 있다. 기본적 분석이란 해당 기업의 재무 구조나 시장 전망 등 기업의 내재적/외

적 가치를 평가하여 주가를 예측하는 기법이며, 기술적 분석이란 거래량, 추세, 이동평균선 등 해당 기업의 통계적 자료를 통해 과거 흐름을 파악하여 미래의 주가를 예측하는 기법이다. 기술적 방법은 빅데이터의 빠르고 정확한 연산을 목적으로 사용되며, 활용되는 데이터의 종류는 그 수를 가늠할 수 없을 정도로 다양하다. 그 중에서도 소셜 미디어, 웹 등에서 기하급수적으로 생성되는 비정형 데이터가 주목받기 시작했다.

이에 본 논문에서는 비정형 데이터 중 웹 뉴스 데이터를 활용하여 주가를 예측하는 연구를 진행하였다. 우선 주가 예측의 대상이 될 표본 기업을 선정하고 그 기업과 관련된 뉴스 데이터를 수집하였다. 수집된 데이터는 형태소 분석과 감성분석을 거쳐 수치화 시켰고 이 수치화된 데이터를 기계학습에 활용되어 예측 모델을 생성하였다. 생성된 모형들을 통해 실제 주가지수 등락을 예측해 보았으며, 예측 모형 간 예측률을 비교하였다.

## 2. 이론적 배경

감성분석이란 반정형, 비정형 텍스트를 분석하여 실용적인 정보를 추출하는 텍스트 마이닝(text mining) 기술의 일부분으로, 어떤 브랜드, 이벤트, 상품에 대한 사람들의 주관적인 견해를 분석하는 자연어 처리 기술이다. 텍스트 마이닝은 주관적인 정보가 많이 내포되어있는 SNS, 객관적인 정보 위주로 작성된 뉴스에도 모두 적용이 가능하다.

형태소는 ‘언어의 형태론적 수준에서의 최소단위’로 정의되며, 이 이상 단위를 축소하면 언어의 의미를 잃어버린다. 형태소 분석은 주어진 문장을 형태소 단위로 분리한 후 용언, 체언, 형용사, 부사, 조사 등의 각 형태소에 맞는 품사 또는 의미를 부여하는 과정이다. 본 논문에서는 비정형 텍스트의 감성분석을 형태소 분석을 실시한 후 각 형태소 별로 극성을 부여하여 전체 텍스트의 극성을 파악하는 과정을 수행하였다.

감성분석은 주어진 텍스트 데이터가 본 글에서 긍정을 의미하는지, 부정을 의미하는지 판단하는 극성 분석 작업을 수행한다. 이 과정에서 각 단어의 빈도수와 긍정지수를 나타낸 감성사전을 이용하여 연구를 진행한다면 감성분석의 정확도를 더욱 향상시킬 수 있다. 송종성(2011)은 범용화된 감성사전이 아닌 각 분야별 전용 감성사전을 구축하여 연구를 진행하는 것이 감성분석의 정확도를 향상시킬 수 있다고, 김동성(2015)은 시간 경과에 따라 감성사전을 추가적으로 확장하면 더 높은 감성분석 정확도를 갖는 것을 확인하였으며, 반면 추가 확장 없이 감성사전을 방치하면, 그 감성사전은 시간 경과에 따라 효율성을 잃어버린다는 사실을 도출해 내었다. 본 연구에서는 김동성(2015)의 감성사전 구축 방법을 바탕으로 각 기업별 주가예측 전용 감성사전을 구축하여

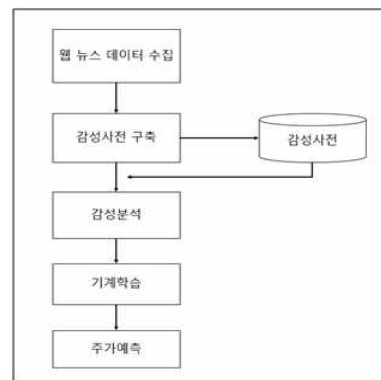
감성분석을 수행하였다.

지금까지 다양한 방면으로 주가예측을 위한 많은 연구가 꾸준히 진행되어 왔다. 조병수(2009)는 뉴스 종류별로 차이는 있지만 매년 발표되는 10대 뉴스가 유가증권시장의 주가 지수에 영향을 미친다는 사실을 알아내었고, 엄장윤(2015)과 최일지(2016)는 뉴스 기반 텍스트 마이닝을 통해 주가 등락 예측에 관한 연구를 진행하였으며, 김유신(2012)은 뉴스를 분석하여 투자정보를 생성하기 위해 감성분석 기법을 적용하였고, 이를 기반으로 주가예측 전용 지능형 투자의사결정모형을 제시하였다. 또한 R. Tushar and S. Saket(2012)는 대규모의 소셜 데이터를 통해 특정한 기업에 대한 감성분석을 기반으로 한 소셜 사이트와 단기 시장 성과와의 상관관계를 분석하였고, 긍정과 부정의 관점이 뚜렷하게 드러난 대중의 의견은 각 기업별 주가지수에 큰 영향을 미친다는 사실을 알아내었다. 이에 본 논문은 웹 뉴스 제목이라는 소규모 데이터가 주가예측에 영향을 미칠 수 있음을 밝혀, 규모의 크기와는 상관없이 뉴스라는 것이 주가 변동에 영향을 끼친다는 사실을 알아내고자 연구를 진행하였다.

## 3. 연구설계 및 실험결과

### 웹 뉴스 데이터를 이용한 주가예측

본 논문에서는 주가와 가장 밀접한 관계를 띄고 있는 비정형 데이터인 뉴스 데이터를 통해 주가예측을 시도하였다. ‘웹 뉴스 제목’이라는 범위 내에서 데이터를 수집하였고, 이 데이터들을 감성분석한 후 기계학습 시키는 데에 활용하여 주가예측 모형을 생성하고자 한다. 본 연구의 수행 과정은 <그림 1>과 같다.



<그림 1> 연구의 수행 과정

본 연구의 수행 과정은 김동성(2015)의 주가예측 모형을 활용한 것이며, 데이터 수집, 감성사전 구축, 감성분석, 기계학습의 총 4단계를 거친 후 주가예측모형을 구현하는 것이다. 데이터 수집 단계에서는 선정된 표본 기업당 하나의 키워드를 선정하여 그 키워드가 들어간 뉴

스 데이터를 수집한다. 수집된 데이터는 형태소 분석 과정을 거쳐 긍정 값과 빈도수를 계산한 후, 긍정지수가 가장 높은 상위 50개의 단어만이 감성사전에 실린다. 다음단계에서는 각 기업별로 구축한 감성사전을 토대로 수집한 데이터에 감성분석 기법을 적용한 후, 기계학습 시키기 위한 데이터로 사용하였다. 기계학습 단계에서는 정제된 데이터를 이용하여 각 분류기법의 기계학습 과정을 거쳐 다수의 예측 모델을 생성한 후, 최종적으로 모델들간의 정확도를 비교하였다.

### 3.2 연구설계

본 논문에서는 KOSPI의 상장사 중 4개의 기업을 실험대상으로 선정하여 연구를 진행하였으며, 2016년 5월 1일부터 2016년 5월 31일까지의 웹 뉴스 데이터를 연구에 활용하였다. 실험 순서는 데이터 수집, 감성사전 구축, 감성분석, 기계학습 순으로 진행하였고 예측 모델과 함께 미래 주가지수의 등락의 예측률을 분석하였다.

### 3.3 데이터 수집

웹 뉴스 데이터는 실시간, 다양한 경로로 생성되는 것은 물론이고, 그 종류 또한 매우 폭넓다. 이러한 웹 뉴스 데이터를 일일이 읽어올 수는 없기에 웹 뉴스 데이터를 사용하기 위해선 자동화된 방식의 시스템을 통해 필요한 정보만을 추출해야 하는 작업이 요구된다. 본 논문에서는 컴퓨터 프로그래밍 언어 Python 을 통해 데이터를 수집하였고, 이를 통하여 필요한 데이터만을 추출해 정제하였다

### 3.4 감성사전 구축

텍스트 데이터는 형태소 분석을 통해 형태소 단위로 분류가 되고, 형태소 별로 품사가 부여된다. 이렇게 처리된 형태소는 각 단어마다 극성 값이 주어지고, 이를 통해 감성사전을 구축한다. 감성사전은 각 단어의 극성 값을 모아놓은 사전으로, 감성사전을 어떻게 구축하느냐에 따라 감성분석의 정확도에 큰 영향을 끼친다. 감성사전을 구축하는 과정의 순서는 다음과 같다.

첫째- 수집한 데이터를 형태소 분석한다.

둘째- 감성사전에 실릴 후보 단어들을 추출한다.

셋째- 단어들의 빈도수와 긍정 값을 계산한다.

넷째- 빈도수가 가장 높은 상위 50개의 단어들만을 추출하여 긍정 지수를 계산한다. 이때 빈도수는 선정한 키워드가 들어간 뉴스를 전체 뉴스 데이터로 나누어 계산하고, 긍정 값은 해당 키워드가 들어간 뉴스가 게재되었을 때 익일 주가가 상승한 경우를 전체 일수로 나누어 계산한다(김동영, 2015). 이때, 빈도수와 긍정값을 구하는 함수는 다음과 같다.

$$\text{include}(i, j) = \begin{cases} 1 & (\text{기사 } j \text{에 단어 } i \text{가 들어간 경우}) \\ 0 & (\text{그 외의 경우}) \end{cases}$$

$$\text{frequency}(i) = \sum_{j=1}^n \text{include}(i, j), \text{ } n \text{ 전체기사의 수}$$

$$\text{NSP}(j) = \begin{cases} 1 & (\text{기사 } j \text{가 게재된 후 익일 주가가 상승한 경우}) \\ 0 & (\text{그외의 경우}) \end{cases}$$

$$\text{positive}(i) = \sum_{j=1}^n \{\text{include}(i, j) \times \text{NSP}(j)\}, \text{ } n = \text{전체 기사 수}$$

데이터 수집 범위가 매우 좁기 때문에 감성사전에 실릴 단어들의 개수는 50개로 한정하였다. 최종적으로, 각 단어의 빈도수와 긍정 값을 통해 긍정지수를 계산하고, 감성사전을 구축한다. 긍정지수는 긍정 값을 빈도수로 나눈 값이며 수식으로 표현하면 다음과 같다.

$$P(i) = \frac{\sum_{j=1}^n \{\text{include}(i, j) \times \text{NSP}(j)\}}{\sum_{j=1}^n \text{include}(i, j)}$$

본 논문에서의 감성사전은 각 기업별 전부 50개의 명사 또는 명사추정범주에 해당되는 형태소로만으로 구성되어 있고, 긍정지수는 0에서 1사이의 값으로 0에 가까울수록 부정, 0.5에 가까울수록 중립, 1에 가까울수록 긍정의 의미를 나타낸다(김동영, 2015). <그림 2>는 본 논문에서 구축한 감성사전의 일부이다.

0.52558	현대	0.63975	스마트폰
0.62051	추천	0.66447	구조
0.58952	아이	0.60674	매수
0.45802	데이터	0.60526	중국
0.64767	사업	0.52597	업계
0.54651	화장품	0.6391	주간
0.49765	마감	0.61742	성장
0.61212	오늘	0.62238	그룹
0.61111	플러스	0.48276	제품
0.58333	화학	0.625	삼성전자
0.50314	금리	0.64052	조정

<그림 2> 감성사전 일부

### 3.5 감성분석

감성사전을 이용하여 웹 뉴스 데이터의 감성분석을 진행하였다. 우선 수집한 데이터는 형태소 분석 과정을 수행하여 명사와 명사추정범주의 형태소만을 추출한 후 감성사전의 단어들과 비교해 해당 데이터의 긍정지수를 계산한다. 텍스트의 긍정지수는 해당 데이터에서 추출한 형태소들의 긍정지수를 합해 그 개수로 나눈 평균값으로

분류 기법	척도	현대	LG	기아	삼성
로지스틱 회귀분석	정확도	0.75	0.90	0.85	0.80
	F1 score	0.78	0.90	0.83	0.80
의사결정 나무	정확도	0.80	0.75	0.75	0.55
	F1 score	0.80	0.83	0.55	0.60
인공 신경망	정확도	0.75	0.85	0.80	0.75
	F1 score	0.74	0.87	0.60	0.74
SVM	정확도	0.70	0.75	0.70	0.80
	F1 score	0.63	0.83	0.50	0.78

사용하였다. (김동영, 2015). 데이터의 긍정지수를 수식화하면 다음과 같다.

$$\text{match}(i, j) =$$

$$\begin{cases} 1 (\text{텍스트 } i \text{에 포함된 명사 } j \text{가 감성사전에 존재 할 경우}) \\ 0 (\text{그 외의 경우}) \end{cases}$$

$$PT(i) = \frac{\sum_{j=1}^n \{\text{match}(i, j) \times P(j)\}}{\sum_{j=1}^n \text{match}(i, j)}, n =$$

텍스트  $i$ 에 포함된 단어의 수

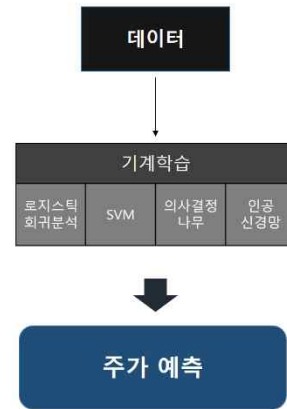
다음으로는 데이터의 긍정지수를 가지고 일별 긍정지수 또한 계산하였다. 일별 긍정지수는 해당일자에 게재된 텍스트들의 긍정지수를 합한 후 그 개수로 나눈 평균값으로 계산 하였으며, 일별 긍정지수를 구하는 함수는 다음과 같다.

$$DP(k) = \frac{\sum_{i=1}^n PT(i)}{n},$$

$n = k$ 일에 게재된 텍스트의 수

### 3.6 기계학습

기계학습은 수집된 데이터의 특징을 추출하고 테스트하여 최적화 또는 자가 발전을 하는 일련의 과정으로, 예측 분야에서 자주 활용되는 기법이다(이호현, 2016). 기계학습을 시키기 위한 데이터는 속성 집합과 각각의 데이터가 시계열인지, 정수인지 실수인지 확인해주는 클래스 레이블로 이루어져 있으며, 이러한 것을 통해 독립변수의 데이터를 가지고 종속변수를 예측하는 것이 기계학습의 목적이라 할 수 있다(김동영, 2015). 본 연구에서는 <그림 3>과 같이 기계학습 이론을 바탕으로 하는 4개의 분류기법, 로지스틱 회귀분석, SVM, 인공신경망, 의사결정 나무 기법을 이용하여 예측 모형을 생성한 후, 주가예측을 진행하였다.



<그림 3> 기계학습의 흐름

### 3.7 결과 분석

뉴스 제목 데이터를 가지고 익일 주가 변동을 예측한 결과, 4개의 기계학습 기법 모두 모든 기업에서 70%가 넘는 예측률을 보였다. F1 score는 대체적으로 0.7에 가까웠으며, 산술평균값을 비교해 보았을 때 로지스틱 회귀분석이 0.81로 가장 높은 값을 가졌다. 각 분류기법에 대한 F1 score와 예측률은 다음 <표 1>과 같다.

<표 1> 4개 기업의 F1 score와 예측률

분류 기법	척도	현대	LG	기아	삼성
로지스틱 회귀분석	정확도	0.75	0.90	0.85	0.80
	F1 score	0.78	0.90	0.83	0.80
의사결정 나무	정확도	0.80	0.75	0.75	0.55
	F1 score	0.80	0.83	0.55	0.60
인공 신경망	정확도	0.75	0.85	0.80	0.75
	F1 score	0.74	0.87	0.60	0.74
SVM	정확도	0.70	0.75	0.70	0.80
	F1 score	0.63	0.83	0.50	0.78

본 논문의 결과는 주가예측을 위해 뉴스 데이터를 사용했던 기존의 연구들의 결과에 비교하였을 때 <표 1>에서 볼 수 있듯이 정확도 측면에서 개선되었음을 확인할 수 있다.

## 4. 결론 및 논의

### 4.1 연구결과의 요약

본 논문에서는 익일 주가 변동의 상승, 하락을 예측하기 위하여 뉴스 기사 데이터를 감성분석 한 후, 그 데이터들 기계학습 시키는 방법을 사용하였다. 데이터의 수집 프로그램은 컴퓨터 프로그래밍을 이용하여 자체적으로 개발하였으며, 수집한 데이터를 가지고 MS엑셀을 통하여 감정사전을 구축하였고 이 감정사전을 통하여 비정

형 텍스트 데이터인 뉴스 기사를 기계학습에 적용하기 위하여 수치화 시켰다. 다음으로는 수치화된 데이터를 가지고 기계학습을 통해 예측 모델을 생성하였다. 실험 결과로는, 예측률은 로지스틱 회귀분석(82.5%), 인공신경망(77.5%), 의사결정나무(75%), SVM(72.5%) 순으로, F1 score는 로지스틱 회귀분석(0.814), 인공신경망(0.737), SVM(0.701), 의사결정나무(0.700)순이었다. 실험결과로 보았을 때, 본 논문에서 구축한 예측모형 중 가장 높은 정확도를 보인 기법은 로지스틱 회귀분석으로, 기계학습 과정에서는 로지스틱 회귀분석을 사용하는 것이 가장 적합할 것으로 판단된다.

본 연구를 통해 최근 대두되고 있는 빅데이터 중 주식과 가장 밀접한 관계를 맺고 있는 비정형 데이터인 웹 뉴스를 통한 주가예측 방법을 하나 제시하였다. 또한, 기계학습 시키는 데에 사용할 데이터를 정제하는 과정에서 감성분석 기법을 적용 시킬 수 있었다. 형태소 분석을 이용해 각 기업별 맞춤형 감성사전을 구축하여 감성분석의 정확도를 향상시킬 수 있었다. 한편, 실험 대상인 4개의 기업은 전부 우리나라 10대 기업에 속하는 기업이었기 때문에 충분한 데이터를 얻는데 어려움이 없었지만 잘 알려지지 않은 중소기업, 신생기업들에 대한 감성분석을 진행하기에는 데이터의 불충분함으로 주가 예측이 어려울 것으로 예상된다.

#### 4.2 연구의 한계 및 향후연구

본 연구에서는 웹 뉴스 데이터에서 제목 데이터만을 가지고 주가 예측을 진행함으로써, 소규모의 데이터만으로도 충분히 주가 예측이 가능하다는 것을 알아내었다. 하지만 데이터의 범위가 너무나도 협소하였기에 검증용 데이터(training set, test set)데이터 기간을 구축할 수 없었고 이에 본 논문의 모형을 실제 주식시장에 사용하기에는 부작용이 있을 것이다. 이에 향후 연구는 예측 모델의 성능 검증용 데이터와 예측 모형간의 성능을 비교하기 위한 데이터(validation set)셋의 범위와 수집 기간을 증가시켜 더욱 높은 정확도를 도출해내야 할 것이다. 또한, 뉴스 데이터뿐만 아니라 최근 빅데이터로 대두로 각광받고 있는 SNS 데이터도 주가예측에 활용하여, 최적의 주가예측 방법론을 제시하는 것을 지향하고자 한다.

#### 참 고 문 헌

- [1] 김동영. (2015). SNS와 뉴스기사의 감성분석과 기계 학습을 이용한 주가예측 모형에 관한 연구. 숭실대학교 소프트웨어특성화대학원 석사학위 논문.
- [2] 김유신. (2012). 주가지수 예측을 위한 뉴스 빅데이터 오피니언마이닝 모형. 국민대학교 박사학위논문
- [3] 이호현, 최은정, 정승현. (2016). 기계학습 응용 및 학

습 알고리즘 성능 개선방안 사례연구. 한국디지털정책 학회, 제 14권, 제 2호.

- [4] Cambria, E. (2013). New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, Vol. 28, No. 2, pp. 15-21.
- [5] Chen, H, Zimbra, D. (2010). AI and Opinion Mining. IEEE Intelligent Systems Volume 25, Issue 3, pp. 74-80.
- [6] Donald, F. (1991). A General Regression Neural Network. IEEE Transactions on Neural Networks, Vol. 2, No. 6, pp. 568-576.
- [7] Esuli, A and Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the 5th Conference on Language Resources and Evaluation LREC' 06, pp. 417-422.
- [8] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan and Claypool.
- [9] Quinlan, J. R. (1993), C4.5: Programs for Machine Learning. San Mateo, California: Morgan Kaufmann.
- [10] Vinodhini, G. (2012). Sentiment Analysis and Opinion Mining: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6.

\* 예측률, F1 score 모두 산술평균값을 이용하였다.

## 난수를 이용한 무리수의 추정

폴수학학교 : 이승우, 인태환  
지도교사 : 이준

### 1. 서론

흔히 무리수를 추정하거나 계산하고자 한다면 사소한 오차를 줄이고자 정확한 계산법으로 계산을 한다. 하지만 이 연구에서는 무리수를 다른 방법으로 추정하고자 하였고 난수와 해시함수를 통해서 무리수를 추정하는데 성공하였다. 본래 암호분야에서 사용되는 해시함수는 중요한 정보를 보호하기 위한 평문의 지문으로 사용된다. 해시함수는 현대 암호를 활용한 정보보호 과정에서 중요한 요소로 작용하며, 이를 난수로 사용하고자 한다. 해시함수에서 사용되는 난수는 예측이 불가능하고, 규칙성이 없어야 하므로 실질적으로 난수를 예측할 방법은 존재하지 않는다. 본 절은 이러한 난수와 함께 해시함수를 통해서 무리수를 추정하는 실험을 진행한다.