

## 텍스트 마이닝을 활용한 고객 리뷰의 유용성 지수 개선에 관한 연구\*

이 홍 주\*\*

### A Study on Classifications of Useful Customer Reviews by Applying Text Mining Approach\*

Hong Joo Lee\*\*

#### ■ Abstract ■

Customer reviews are one of the important sources for purchase decision makings in online stores. Online stores have tried to provide useful reviews in product pages to customers. To assess the usefulness of customer reviews before other users have voted enough on the reviews, diverse aspects of reviews were utilized in previous studies. Style and semantic information were utilized in many studies.

This study aims to test diverse algorithms and datasets for identifying a proper classification method and threshold to classify useful reviews. In particular, most researches utilized ratio type helpfulness index as Amazon.com used. However, there is another type of usefulness index utilized in TripAdvisor.com or Yelp.com, count type helpfulness index. There was no proper threshold to classify useful reviews yet for count type helpfulness index. This study used reviews and their usefulness votes on restaurants from Yelp.com to devise diverse datasets and applied text mining approaches to classify useful reviews. Random Forest, SVM, and GLMNET showed the greater values of accuracy than other approaches.

Keyword : Customer Review, Classification, Usefulness Index

## 1. 서 론

전자상거래의 활성화로 거의 모든 상품들이 온라인으로 판매되고 있으며, 판매되는 제품에 대한 고객 리뷰가 구매 의사결정에 중요한 영향을 미치고 있다(Dellarocas et al., 2010). 고객들은 판매자가 작성한 제품정보 보다 다른 고객이 제품을 경험한 후 작성한 제품리뷰가 더 많은 정보를 제공하고, 객관적이며, 신뢰할 만하다고 생각하고 있다(Dellarocas, 2003). 이로 인해 제품리뷰의 평균 점수 상승이 제품 판매량 증가로 이어지며(Zhu and Zhang, 2006), 구매제품을 변화시켜 가격경쟁을 증대시킨다(Li et al., 2011).

하지만, 너무 많은 제품과 고객 리뷰가 존재하여 탐색비용의 절감이라는 전자상거래의 장점이 무색해지고 있다. 구매하려는 제품의 모든 리뷰를 읽고 제품의 장단점을 파악하는 것은 무척 힘든 일이 되었다(David and Pinch, 2006; Liu et al., 2008). 따라서, 전자상거래 업체들은 제품에 대한 고객 리뷰를 다양한 방식으로 제시하여 사용자들이 손쉽게 제품에 대한 유용한 정보를 얻을 수 있도록 돕고 있다. Amazon.com은 제품 리뷰에 대해 유용한 지 여부를 묻고 긍정적인 리뷰 중 가장 유용성이 높다고 투표된 리뷰와 중립적인 리뷰 중 가장 유용성이 높은 리뷰를 최상단에 제시하고 있다. 또한, 리뷰 자체가 가지고 있는 리뷰길이, 작성자, 사용단어 등의 속성을 활용하여 리뷰의 유용성에 미치는 영향을 분석하였으며(이상재, 최준연, 최진호, 2014), 이를 이용하여 유용한 리뷰만을 사용자에게 추천하기 위한 연구들이 진행되어 왔다(Cao et al., 2011; Mudambi and Schuff, 2010).

Amazon.com과 같이 리뷰에 대한 유용성 투표에 유용하다와 유용하지 않다라고 묻는 경우에는 전체 투표 수(유용 투표 수+비유용 투표 수)로 유용 투표 수를 나눈 비율 값이 리뷰의 유용도를 나타내는 지표로 활용된다. 이 비율은 0과 1사이의 값을 갖기 때문에 일정한 기준을 정해서 그 이상의 값을 보인 리뷰들을 유용한 리뷰로 구분해서 활용

한다. Choeh et al.(2015)와 Zhang and Tran(2011)의 경우에는 위 비율이 0.6 이상인 경우를 유용한 리뷰로 구분하였다. Yelp과 TripAdvisor같은 경우에는 리뷰에 대한 유용성 투표에 유용한 지와 유용하지 않은 지를 묻지 않고, 유용한 지만을 묻기 때문에 Amazon과 같은 비율 지표를 활용하지 못한다.

본 연구의 목표는 리뷰가 가지고 있는 속성들을 제외하고 리뷰 텍스트만을 가지고 유용 고객 리뷰를 분류하는데 있어서 좋은 성과를 보이는 알고리즘과 유용 리뷰 분류 기준을 파악하고자 한다. 리뷰의 텍스트 이외의 속성들은 사이트마다 수집 가능한 정보들이 상이하고 다양한 형태를 띠고 있기 때문에 데이터가 분야나 사이트에 제한적인 경우가 있다. 따라서, 고객이 리뷰를 작성할 수 있는 사이트가 모두 공통적으로 가지고 있는 리뷰 텍스트를 활용한 유용한 리뷰 예측을 수행하고자 한다. 다양한 데이터 집합과 분류 방안들을 활용하여 유용 고객 리뷰 집합의 분류기준을 파악하고자 한다. 이를 위해 Yelp의 음식점종에 대한 고객의 리뷰를 활용하여 유용한 리뷰 집합과 유용하지 않은 리뷰 집합을 분류하는 실험을 수행하였다.

제 2장에서는 고객 리뷰 제시와 유용 고객 리뷰 분류에 대한 관련연구를 정리하였으며, 제 3장에서 활용한 자료를 제시하였다. 제 4장에서 활용한 분류 알고리즘과 성과를 제시하였으며, 제 5장에서 결과에 대한 토의과 결론을 제시하였다.

## 2. 관련 연구

### 2.1 전자상거래 업체들의 제품 리뷰 제시 방안

전자상거래 업체들은 고객에게 제품에 대한 정보를 제공하기 위하여 고객들의 제품 리뷰를 제공하고 있다. 구매의사결정에 제품 리뷰가 중요한 역할을 하기 때문에 제품에 대한 유용한 정보를 제공하여 줄 수 있는 리뷰를 선별해서 보여주는 것이 중요하다. 또한, 제품 리뷰가 많기 때문에 고객이 원하는 다양한 방식으로 정렬하여 리뷰를 제공할 필

요가 있다. <Table 1>은 국내외의 대표적인 온라인 쇼핑 사이트들의 고객 리뷰 평가 방안을 정리한 것이다.

<Table 1> Evaluations of Online Customer Reviews

Websites	Evaluation methods	Usefulness index
Amazon	Votes (Yes/No)	Total # of Useful votes/Total # of Votes
Yelp	Votes(Yes)	Total # of Votes
TripAdvisor	Votes(Yes)	Total # of Votes
Rakuten	Votes (Yes/No)	Total # of Useful votes
Auction (Korea)	Votes(Yes)	Total # of Views, Total # of Votes
emart	Votes(Yes)	Total # of Votes
cjmall	Votes (Yes/No)	Total # of Views, Total # of Useful votes, Total # of Not Useful votes

<Table 1>에 포함된 사이트들은 리뷰 평가를 유용한 지 만을 묻는 사이트와 유용한 지/유용하지 않은 지를 묻는 사이트로 나누어진다. Amazon은 리뷰가 유용했는지, 유용하지 않았는지를 묻고 있고, 옥션은 리뷰가 유용했는지 만을 묻고 있다. 리뷰의 유용성 지표도 마찬가지로 크게 두 가지로 구분해 볼 수 있다. 하나는 Yelp와 같이 유용하다고 투표한 수를 제시하는 것이며, 다른 하나는 Amazon처럼 전체 투표 수에서 유용하다는 투표수가 차지하는 비율을 제시하는 방안이다. 한국의 사이트들은 이외에도 고객들이 해당 리뷰를 얼마나 조회했는 지의 정보들도 추가적으로 제공하고 있다. 이는 전체 리뷰를 보기 위해서는 ‘전체 읽기’ 같은 버튼을 눌러야 하고 이 버튼을 누른 고객의 수를 제시하고 있는 것이다.

유용성 투표 수나 유용성 투표율을 활용하여 리뷰 중에서 높은 투표 수나 투표율을 가진 리뷰를 해당 제품을 탐색하는 고객들에게 제공하고 있다. 리뷰에 제품에 대한 선호도가 표시되기 때문에 제

품을 선호하는 고객들중에서 유용한 리뷰를 파악하여 제공하거나, 제품을 선호하지 않는 고객들중에서 유용한 리뷰를 파악하여 제공하고 있다.

## 2.2 고객 리뷰의 유용성 예측 방안

사용자에게 유용한 리뷰를 추천하거나 리뷰를 활용하여 사용자가 좋아할 만한 제품을 추천하는 방안이 연구되어 왔다. Cao et al.(2011)은 리뷰가 장점을 포함하고 있는지, 단점을 포함하고 있는지, 등록 후 얼마나 시간이 지났는지 등의 기본 정보와 단어 수, 문장 수 등의 리뷰스타일에 대한 정보, 그리고 리뷰에 포함된 단어들의 의미를 고려한 의미정보를 활용하여 리뷰의 유용성을 예측하는 경우가 각각 하나씩 만을 활용하는 경우 보다 정확함을 보였다. Kim et al.(2006)도 리뷰의 구조, 의미, 형태 정보를 활용하여 SVM Regression을 통해 유용성을 예측하는 방안을 제시하였다. Liu et al.(2008)은 리뷰어의 경험, 리뷰 작성 스타일, 리뷰 작성시기를 고려하여 리뷰의 유용성을 예측하는 방안을 제시하였다. Ghose and Ipeirotis(2011)은 제품 속성, 리뷰 속성, 리뷰어 속성, 리뷰어 경험, 리뷰 가독성, 리뷰 주관성을 고려하여 리뷰의 유용성을 예측하였다. 위의 방안들은 리뷰 유용성 지표를 예측하거나 리뷰 유용성의 순위와 비교하는 방식으로 결과를 검증하였다. Kim et al.(2015)는 댓글의 구조적인 정보, 의미정보와 초기 기대가 뉴스 댓글 공감에 미치는 영향을 분석하였다.

리뷰 유용성 예측을 유용한 리뷰인지 유용하지 않은 리뷰인지로 구분하는 문제로 생각하여 해결하려는 연구들도 수행되어 왔다. Zhang and Tran(2011)은 특정 단어의 포함여부가 유용한 리뷰와 유용하지 않은 리뷰를 구분하는데 어느 정도나 기여하는 지를 정보이득을 통해 측정하여, 각 단어별 기여도를 측정한다. 리뷰에 포함된 단어들의 기여도를 합하여 리뷰의 유용성을 예측하여 추천하였다.

리뷰에 포함되어 있는 단어나 문장의 의미를 분

석하여 제품에 대한 리뷰가 긍정적인 지 부정적인 지를 분석하여 제품을 추천하는 연구들이 진행되었다(Hatzivassiloglou and McKeown, 1997; Pang et al., 2002).

### 3. 자료

고객리뷰를 활용하여 유용한 리뷰를 분류하기 위한 실험을 위해 Yelp.com의 고객 리뷰를 활용하였다.<sup>1)</sup> Yelp.com은 다양한 업종에 대해 리뷰를 작성하는 사이트이기 때문에, 업종에 음료(beverage)와 음식(food)가 포함되어 있는 업체에 대한 리뷰만을 실험에 사용하였으며 자료의 성격은 <Table 2>와 같다.

<Table 2> Data description

Total number of reviews	179,479
Total number of users	37,837
Total number of stores	5,709
Average review ratings	3.759(5점 만점)
Average review length	789.7
Review date	2005-03-07~2013-01-05



<Figure 1> Example of an Yelp review

Yelp.com에서는 사용자들이 음식점에 대한 리뷰를 작성할 수 있으며, <Figure 1>처럼 작성된 리뷰

를 읽은 다른 사용자들이 해당 리뷰가 도움이 되었는지, 재미있는지, 쿨한지 투표를 할 수 있게 되어 있다.

리뷰에 대한 사용자들의 유용성(useful) 투표 데이터를 사용하여 리뷰를 유용한 집합과 유용하지 않은 집합으로 구분할 수 있다. <Table 3>은 전체 리뷰 집합의 유용성 투표에 관한 정보와 유용성 투표 수가 1 이상인 리뷰 집합의 정보를 보여주고 있다.

<Table 3> Descriptions of Useful Reviews

	All reviews	Reviews (useful votes >= 1)
Minimum of useful votes	0	1
Median of useful votes	1	2
Mean of useful votes	1.359	2.326
3 Quartile of useful votes	3	2
Maximum of useful votes	82	82

리뷰 내용을 활용하여 유용한 리뷰를 분류하기 위한 실험이므로 유용하지 않은 리뷰로는 유용 투표에서 한번도 득표를 하지 못한 74,601개의 리뷰를 선정하였다. 유용성 투표의 득표 수의 분포는 평균이 1.359이지만 총 리뷰의 1/3에 해당하는 51,913개의 리뷰가 유용성 투표에서 1표를 획득하였기에 미디언 값은 1이다.

유용한 리뷰집합과 유용하지 않은 리뷰집합으로 구분하기 위한 실험에 사용되는 데이터 집합은 다음과 같이 설정하였다. 실험집합 I-1은 전체 리뷰 중에서 유용성 득표 수가 0개인 리뷰 5,000개와 1개 이상인 리뷰 5,000개를 무작위로 선정하였으며, 이중 90%를 학습집합으로 10%를 테스트집합으로 활용하였다. 10겹 교차 검증(10-fold cross validation)을 활용하여 분류에 활용된 알고리즘의 성과를 측정하였다. <Table 4>는 실험에 활용된 유용 리뷰 집합과 유용하지 않은 리뷰 집합에서 가장 많이 출현하는 단어들의 빈도를 보여주고 있다. 가장 많이 출현하는 10개의 단어들을 표시하였으며,

1) ACM RecSys Challenge 2013으로 공개된 데이터를 활용하였다.  
<https://www.kaggle.com/c/yelp-recsys-2013>.

두 집단에서 많이 사용되는 단어들은 매우 유사하나 출현빈도에서 차이를 보이고 있다.

〈Table 4〉 Term Frequency(Data set I-1)

Useful review set		Not useful review set	
Word	frequency	Word	frequency
place	3,974	good	3,022
good	3,492	place	2,992
food	3,364	food	2,825
like	3,121	great	2,094
just	2,497	like	1,954
get	2,406	get	1,594
great	2,251	time	1,592
time	2,153	just	1,567
one	2,103	order	1,409
order	2,050	tri	1,385

실험집합 I-2는 유용성 득표 수가 0개인 리뷰 5,000개와 2개 이상인 리뷰 5,000개를 무작위로 선정하였으며, 마찬가지로 90%를 학습집합으로 10%를 테스트 집합으로 활용하였다.

〈Table 5〉는 실험집합 I-2의 리뷰에 많이 출현하는 단어들을 보여 주고 있다. 〈Table 4〉와 〈Table 5〉를 보면 실험집합 I-1과 I-2에 많이 출현하는 상위 10개 단어들은 9개가 일치하며 집합 I-1에는 great이 집합 I-2에는 realli가 리스트에 포함되었다. 유용하지 않은 리뷰 집합은 실험 집합 I-1, I-2 모두 같은 모집단에서 무작위로 추출한 것이어서 큰 차이를 보이고 있지는 않다.

마찬가지 방식으로 실험집합 I-3은 유용성 득표 수가 0인 유용하지 않은 리뷰 집합과 유용성 득표 수가 3개 이상인 유용 리뷰 집합으로 구분하였으며, 실험집합 I-4는 유용성 득표 수가 4개 이상, 실험집합 I-5는 유용성 득표 수가 5개 이상, 실험집합 I-6은 유용성 득표 수가 6개 이상, 실험집합 I-7은 유용성 득표수가 7개 이상, 실험집합 I-8은 유용성 득표 수가 8개 이상인 리뷰들을 유용 리뷰 집합으로 구분하였다. 실험집합 I-9까지는 각각 5,000개씩의

리뷰를 무작위로 선정하였으며, 실험집합 I-5, I-6, I-7, I-8은 리뷰 집합의 크기 문제로 인해 3,000개, 2,000개, 2,000개, 1,000개씩 선정하였다.

〈Table 5〉 Term Frequency(Data set I-2)

Useful review set		Not useful review set	
Word	frequency	Word	frequency
place	4,414	good	3,004
good	4,135	place	3,002
like	4,127	food	2,824
food	4,030	great	2,122
just	3,386	like	1,951
order	3,056	get	1,630
get	3,021	time	1,589
one	2,789	just	1,561
time	2,646	one	1,403
realli	2,303	order	1,394

〈Table 6〉 Data Set I

Data set	Not useful sets	# of reviews	Useful sets	# of reviews
I-1	Number of useful votes = 0	5,000	1 >=	5,000
I-2		5,000	2 >=	5,000
I-3		5,000	3 >=	5,000
I-4		5,000	4 >=	5,000
I-5		3,000	5 >=	3,000
I-6		2,000	6 >=	2,000
I-7		2,000	7 >=	2,000
I-8		1,000	8 >=	1,000

실험집합 I은 유용하지 않은 집합은 모두 유용 득표 수가 0인 경우만을 활용하였다. 실험집합 II는 유용하지 않은 집합이 득표 수가 0인 경우가 아니고 유용 집합보다 득표 수가 적은 모든 리뷰를 유용하지 않은 집합으로 설정하였다. 실험집합 II-1은 유용하지 않은 집합은 득표 수가 0인 경우이고, 유용집합은 득표 수가 1 이상인 경우이다. 실험집합 II-1과 I-1은 같은 설정에서 추출된 집합이다. 실험집합 II-2는 유용득표 수가 1이하인 경

우가 유용하지 않은 집합이며, 유용 집합은 득표수가 2 이상인 경우이다.

<Table 7> Data Set II

Data set	Not useful sets	# of reviews	Useful sets	# of reviews
II-1	0	3,000	1 >=	3,000
II-2	1 <=	3,000	2 >=	3,000
II-3	2 <=	3,000	3 >=	3,000
II-4	3 <=	3,000	4 >=	3,000
II-5	4 <=	3,000	5 >=	3,000

#### 4. 분류 방안 및 결과

유용한 리뷰 분류방안으로는 분류 문제에 많이 적용이 되는 데이터마ining 기법들을 활용하였다.

Support Vector Machine(SVM)은 분류와 회귀 분석에 활용되는 기법으로 고차원의 특징공간으로 이동시킨 후 데이터의 특정 클래스로부터 가장 먼 값을 갖는 초평면(hyperplane)들을 찾아 이를 통해 데이터를 분류하는 방안이다(Hong, 2011; Min and Lee, 2005; Vapnik, 1998). 두 클래스로 분류하는 문제라면 최적 분리 초평면은 각 클래스에 속하는 데이터들 중에서 Support Vector를 지나는 두 개의 평행인 초평면들 사이의 거리 즉, 마진을 최대로 갖는 경우이다.

하나의 분류기(classifier)로 좋은 성과를 얻기가 어려워지자 여러 개의 분류기를 혼합하여(ensemble) 사용하는 방안들이 고안되었다. 대표적인 방안이 의사결정나무들의 boosting(Shapire et al., 1998)과 bagging(Breiman, 1996)이다. Random Forest(RF)는 Breiman(2001)와 Liaw and Wiener(2002) 여러 개의 의사결정나무를 독립적으로 생성한 후 결과를 취합하여 빈도가 더 높은 값을 취하는 bagging 방안에 무작위성(randomness)을 추가한 방안이다. 즉, 각 의사결정나무는 전체 데이터에서 각각 bootstrapping한 데이터를 사용하며, 의사결정나무에서 노드를 분리할 때 모든 변수들을

고려해서 가장 좋은 분리를 선택하는 것이 아니고 무작위로 선정된 부분 변수집합에서 가장 좋은 분리방안을 선택하게 하고 있다.

Multinomial logistic regression 혹은 maximum entropy classfier(MAXENT)는 로지스틱 회귀분석을 일반화하여 종속변수가 두 개 이상의 카테고리를 갖는 분류문제에 적용되는 알고리즘이다. 지도학습을 통한 텍스트 분류에 많이 활용되어 왔으며(Jurka et al., 2012), 텍스트 분류에 많이 활용되는 naive Bayes 알고리즘에 비해 높은 정확도를 보이는 것으로 알려져 있다(Nigam, et al., 1999).

Elastic-net 페널티를 활용한 일반 선형 모형(Generalized linear model)(GLMNET)은 Lasso 방안의 한계를 해결하기 위해 제시된 방안이다(Friedman et al., 2010). 서로 상관관계가 높은 변수들이 있을 때 Lasso 방안은 이 중에서 하나만을 활용하고 다른 변수들은 무시하게 되는 경향이 있다. 이를 극복하기 위해 elastic net은 Lasso에서 활용하는 페널티 외에 Ridge 회귀분석에서 활용되는 2차의 페널티를 추가적으로 활용한다.

실험은 통계분석 도구인 R의 RTextTools 패키지(Jurka et al., 2012)를 활용하였다. 제 3장에서 설명한 것처럼 전체 데이터 집합의 90%를 학습집합으로 10%를 테스트집합으로 활용하였으며, 10겹 교차 검증(10-fold cross validation)으로 각 알고리즘들의 성과를 분석하였다. 리뷰에서 숫자와 기호들은 제거하였으며, 불용어(stop words)들도 제거하였다. 단어를 어간 추출(stemming)하였으며, TF-IDF를 활용하여 단어의 가중치를 계산하였다. 단어의 단순빈도가 아니라 TF-IDF 가중치를 가지고 각 리뷰집합에 단어가 포함되는 확률과 정보 가치를 계산하였다.

제 3장에서 제시한 실험집합들을 활용하였으며, 실험집합 I에 대한 분류방안의 분류정확도는 <Table 8>과 같다. <Figure 1>은 <Table 8>을 가시화한 것이다.

실험집합 I-1, I-2, I-3까지는 RF 기법의 분류 정확도가 가장 높았다. 다음으로는 SVM, GLMNET,

MAXENT순으로 높은 정확도를 보였다. 실험집합 I-4부터 I-8까지는 위와는 다른 순위를 보인다. 가장 높은 정확도를 보인 알고리즘은 SVM이었으며, 다음으로 GLMNET, RF순으로 높은 정확도를 보였다. 위 세 개 알고리즘의 분류 정확도는 큰 차이를 보이지는 않았다. 다음으로는 MAXENT가 높은 정확도를 보였다.

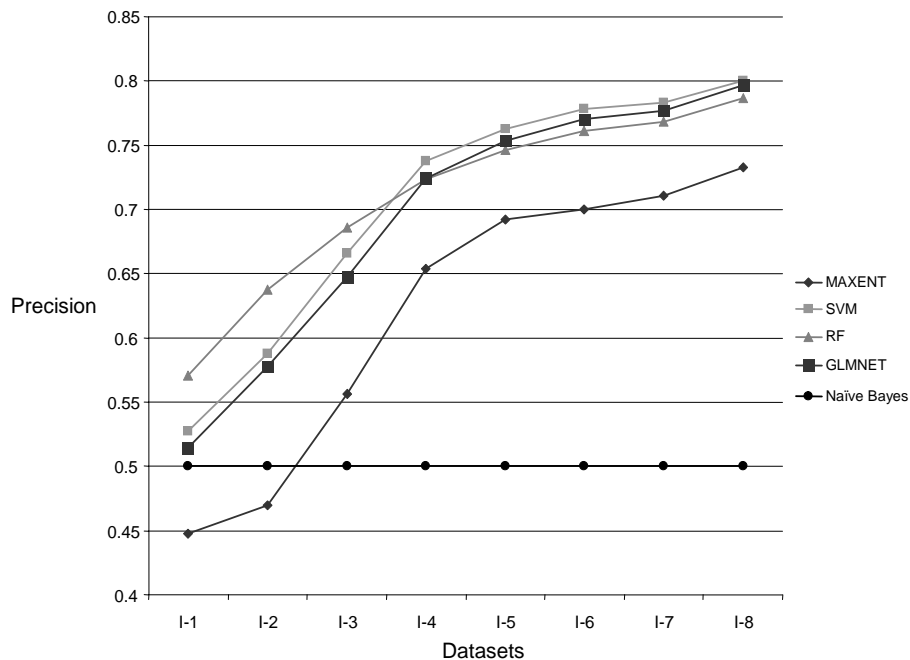
〈Table 8〉 Accuracy of Classifications

Data set	MAX ENT	SVM	RF	GLM NET
I-1	0.447	0.527	0.570	0.514
I-2	0.469	0.588	0.637	0.577
I-3	0.556	0.666	0.686	0.647
I-4	0.654	0.738	0.723	0.724
I-5	0.692	0.762	0.746	0.753
I-6	0.7	0.778	0.761	0.770
I-7	0.710	0.783	0.768	0.776
I-8	0.732	0.800	0.786	0.796

유용리뷰 집합과 유용하지 않은 리뷰 집합의 특징이 명확하게 구분되지 않는 상황에서는 RF가 가장 좋은 성과를 보였으나, 유용집합과 비유용집합의 특징이 좀 더 명확해진 이후에는 SVM이 더 정확하게 분류하게 된 것으로 볼 수 있다.

〈Figure 2〉에서 볼 수 있듯이 실험집합 I-1에서 I-4까지의 정확도 상승 폭에 비해 I-5부터 I-8까지의 정확도 상승은 크지 않다. SVM의 성과는 실험집합 I-1에서 I-4까지 0.21정도 상승한 반면 실험집합 I-5에서 I-8까지는 0.04정도만 상승했다. 마찬가지로 RF도 실험집합 I-1에서 I-4까지 0.15정도 상승한 반면 실험집합 I-5에서 I-8까지는 0.04정도만 상승했다.

따라서, 유용한 리뷰를 분류하는 실험의 정확도는 실험집합 I-5를 기점으로 정확도 상승정도가 낮아지는 것을 볼 수 있다. 텍스트 마이닝을 통한 분류 연구에서 벤치마킹 지표로 많이 활용되는 Naive Bayes 방안도 비교를 위해 수행하였으며, 〈Figure 2〉에서 볼 수 있듯이 Naive Bayes는 모든 집합에



〈Figure 2〉 Accuracy of Classifications

서 0.5의 분류정확도를 나타내었다.

실험집합 II에 대한 분류정확도는 <Table 9>와 같다. 실험집합 II-1부터 II-4까지는 RF가 가장 성과가 좋았으며, II-5는 GLMNET이 가장 성과가 좋았다. 하지만 II-1을 제외하고 II-2부터 II-5까지 SVM, RF, GLMNET의 정확도는 거의 유사하다고 볼 수 있다. 이는 실험집합 I의 알고리즘 별 분류 성과 순위와 유사하다.

실험집합 I과 II의 1부터 5집합은 유용 리뷰 집합에 속한 리뷰의 기준은 동일하다. 차이점은 유용하지 않은 리뷰 집합에 속한 리뷰의 기준인데, 실험집합 II가 득표 수 0부터 유용 리뷰 집합 기준 이하의 득표 수까지 섞여 있다. 실험집합 II가 분류 알고리즘에게는 더 어려운 문제라고 볼 수 있다. 집합 1부터 3까지는(I-1과 II-1은 같은 기준의 집합) 실험집합 I과 II에서 거의 유사한 분류 성과를 보이고 있으나, 집합 4, 5는 실험집합 I의 정확도가 더 높았다.

<Table 9> Accuracy of Classifications

Data set	MAX ENT	SVM	RF	GLM NET
II-1	0.447	0.527	0.570	0.514
II-2	0.555	0.632	0.642	0.639
II-3	0.583	0.661	0.666	0.661
II-4	0.608	0.678	0.683	0.667
II-5	0.631	0.702	0.702	0.704

## 5. 토의 및 결론

실험에 활용된 분류 기법들의 정확도가 대체로 우수하고 안정적이었다. 유용리뷰 집합의 기준이 되는 유용득표 수가 3개 이상인 경우는 RF가 제일 성과가 좋으나 기준이 3개 이상보다 높아 지게 되면 미세하지만 SVM과 GLMNET의 성과가 더 높았다.

실험집합 II의 경우에는 실험집합 I에 비해 분류 알고리즘에게 더 어려운 문제이다. 실험집합 I은

유용하지 않은 리뷰 집합이 득표 수가 0인 집단으로 좀 더 동질할 것으로 볼 수 있으나, 실험집합 II는 득표 수의 분포 수가 0부터 유용집합에 미달하는 득표 수까지이므로 좀 더 상이한 집합으로 볼 수 있다. 따라서, 유용한 리뷰를 분류하는 모형을 학습하기 위해서는 유용하지 않은 리뷰 집합을 득표수가 0이고, 유용한 리뷰 집합은 득표 수의 분포를 고려하여 높은 득표 수를 기준으로 선정하여야 한다. 본 실험에서는 5 이상의 득표 수부터 안정적인 분류성가를 보였다.

유용성 투표 수 기준 지표에서 유용 리뷰를 분류하기 위해서는 유용 리뷰와 유용하지 않은 리뷰를 구분하기 위한 기준이 필요하다. 비율형 유용 지표에서는 0.6이라는 기준이 널리 통용되고 있으나 투표 수 기준 지표에는 이와 같은 기준이 설정되어 있지 않다. 리뷰가 얻은 득표 수의 분포를 기준으로 설정할 수도 있으나 <Table 3>에 보이는 바와 같이 리뷰의 40%가 득표 수가 0이며, 30% 정도는 득표 수가 1이다. 득표 수의 평균도 1.3(0개 포함), 2.3(0개 제외)이며, 3분위수도 3(0개 포함), 2(0개 제외)이다. 따라서, 득표 수의 분포에 의해서 유용 리뷰의 기준을 설정하기에도 어려움이 있다.

본 연구에서는 유용 투표 수에 따라서 다양한 데이터 집합을 생성하였으며, 이를 텍스트 마이닝의 다양한 분류 알고리즘을 활용하여 분류하였다. 데이터집합에 따라 분류 성과를 파악하여 분류 정확도가 높은 분류 알고리즘을 파악하고, 어느 정도 이상의 득표 수를 보인 경우가 유용 리뷰 집합으로 선정해야 좋을 지를 해석하였다.

유용투표 수를 달리한 실험집합 I의 분류 정확도를 분석하여 본 결과 유용 투표 수가 5개 이상인 집단부터 분류정확도가 안정적인 것으로 나타났다. 실험집합 II의 경우에도 유용 투표 수가 5개 이상인 집단과 그렇지 않은 집단의 분류 정확도가 70%를 상회하여 두 집단간에 차이를 어느 정도 파악할 수 있게 되는 것으로 볼 수 있다.

득표 수 기준 유용성 지표를 활용하여 유용 리뷰 집합과 그렇지 않은 집합을 분류하는 경우에 그



기준이 명확하지 않은 측면이 있다. 유용 득표 수의 분포를 기준으로 선정할 수도 있으나 리뷰 텍스트를 활용하여 분류하는 알고리즘이 보이는 정확도를 기준으로 적합한 분류기준을 파악하였으며, 다양한 데이터 집합과 분류 알고리즘을 활용하여 분류 정확도를 파악하고 그 추세를 확인하였다. 단순히 유용성 투표 수만을 활용하는 것 보다는 리뷰에 대한 전체 조회 수 대비 유용 투표 수를 활용할 수 있다면 유용한 리뷰를 좀 더 명확히 파악할 수 있을 것이다.

## References

- Breiman, L., "Bagging Predictors", *Machine Learning*, Vol.24, No.2, 1996, 123-140.
- Breiman, L., "Random Forests", *Machine Learning*, Vol.45, No.1, 2001, 5-32.
- Cao, Q., W. Duan, and Q. Gan, "Exploring Determinants of Voting for The 'Helpfulness' Online Userreviews : A Text Mining Approach", *Decision Support Systems*, Vol.50, No.2, 2011, 511-521.
- Choeh, J.Y., H.J. Lee, and S.J. Park, "A Personalized Approach for Recommending Useful Product Reviews Based on Information Gain", *KSII Transactions on Internet and Information Systems*, Vol.9, No.5, 2015, 1702-1716.
- David, S. and T. Pinch, "Six Degrees of Reputation : The Use and Abuse of Online Review and Recommendation Systems", *First Monday*, Vol.11, No.3, 2006, Available at <http://dx.doi.org/10.5210/fm.v11i3.1315>(Downloaded November 28, 2015).
- Dellarocas, C., "The Digitization of Word of Mouth : Promise and Challenges of Online Feedback Mechanisms", *Management Science*, Vol.49, No.10, 2003, 1407-1424.
- Dellarocas, C., G. Gao, and R. Narayan, "Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?", *Journal of Management Information Systems*, Vol.27, No.2, 2010, 127-157.
- Friedman, J., T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Model via Coordinate Descent", *Journal of Statistical Software*, Vol.33, No.1, 2010, 1-22.
- Ghose, A. and P.G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews : Mining Text and Reviewer Characteristics", *IEEE Transactions on Knowledge and Data Engineering*, Vol.23, No.10, 2011, 1498-1512.
- Hatzivassiloglou, V. and K. McKeown, "Predicting Theoretical Orientation of Adjectives", *Proceedings of the eight conference on European chapter of the association for computational linguistics*, 1997, 174-181.
- Hong, E.S., "Early Software Quality Prediction Using Support Vector Machine", *Journal of Information Technology Services*, Vol.10, No.12, 2011, 235-245.
- (홍의석, "Support Vector Machine을 이용한 초기 소프트웨어 품질 예측", *한국IT서비스학회지*, 제10권, 제12호, 2011, 235-245.)
- Jurka, T.P., L. Collingwood, A.E. Boydston, E. Grossman, and W. van Atteveldt, "R Text Tools : Automatic Text Classification", *The R journal*, Vol.5, No.1, 2012, 6-12.
- Jurka, T.P., "Maxent : An R Package for Low-Memory Multinomial Logistic Regression with Support for Semi-Automated Text Classification", *The R Journal*, Vol.4, No.1, 2012, 56-59.
- Kim, J.H., O.B. Kwon, O.Y. Song, and Y.S. Jin, "Applying Text Mining to Identify Factors

- Which Affect Likes and Dislikes of Online News Comments”, *Journal of Information Technology Services*, Vol.14, No.2, 2015, 159-176.
- (김정호, 권오병, 송영은, 진유선, “텍스트 마이닝을 통한 댓글의 공감도 및 비공감도에 영향을 미치는 댓글의 특성 연구”, *한국IT서비스학회지*, 제14권, 제2호, 2015, 159-176.)
- Kim, S.M., P. Pantel, and T. Chklovski, “Automatically Assessing Review Helpfulness”, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2006, 423-430.
- Li, X., L.M. Hitt, and J. Zhang, “Product Reviews and Competition in Markets for Repeat Purchase Products”, *Journal of Management Information Systems*, Vol.27, No.4, 2011, 9-42.
- Liaw, A. and M. Wiener, “Classification and Regression by Random Forest”, *R News*, Vol.2, No.3, 2002, 18-22.
- Liu, Y., X. Huang, A. An, and X. Yu, “Modeling and Predicting the Helpfulness of Online Reviews”, *Proceedings of the Eighth IEEE International Conference on Data Mining*, 2008, 443-452.
- Min, J.H. and Y.C. Lee, “Support Vector Bankruptcy Prediction Model with Optimal Choice of RBF Kernel Parameter Values using Grid Search”, *Journal of the Korean Operations Research and Management Science Society*, Vol.30, No.1, 2005, 55-74.
- (민재형, 이영찬, “Support Vector Machine을 이용한 부도예측모형의 개발”, *한국경영과학회지*, 제30권, 제1호, 2005, 55-74.)
- Nigam, K., J. Lafferty, and A. McCallum, “Using Maximum Entropy for Text Classification”, *IJCAI-99 Workshop on Machine Learning for Information Filtering*, 1999, 61-67.
- Mudambi, S.M. and D. Schuff, “What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com”, *MIS Quarterly*, Vol.34, No.1, 2010, 185-200.
- Pang, B., L. Lee, and S. Vaithyanathan, “Thumbs Up? : Sentiment Classification Using Machine Learning Techniques”, *Proceedings of the ACL-02 conference on empirical methods in natural language processing*, 2002, 79-86.
- Shapire, R., Y. Freund, P. Bartlett, and W. Lee, “Boosting the margin : A New Explanation for The Effectiveness of Voting Methods”, *Annals of Statistics*, Vol.26, No.5, 1998, 1651-1686.
- Vapnik, V., *Statistical Learning Theory*, Springer, New York, 1998.
- Zhang, R. and T. Tran, “An Information Gain-Based Approach for Recommending Useful Product Reviews”, *Knowledge and Information Systems*, Vol.26, No.3, 2011, 419-434.
- Zhu, F. and M. Zhang, “The Influence of Online Consumerreviews on The Demand for Experience Goods : The Case of Video Games”, *Proceedings of the Twenty-Seventh International Conference on Information Systems*, Paper No.25, 2006, Available at <http://aisel.aisnet.org/icis2006/25> (Downloaded November 25, 2015).

## ◆ About the Authors ◆



**Hong Joo Lee (hongjoo@catholic.ac.kr)**

Hong Joo Lee is an associate professor of business administration, Catholic University of Korea. He has a Ph.D. from KAIST Business School and was with MIT Center for Collective Intelligence as a postdoctoral fellow in 2006 and a visiting scholar in 2011. His research areas are utilizing intelligent techniques and harnessing collective intelligence in business settings, and analyzing effects of intelligent aids in e-commerce and m-commerce. His papers have been published in International Journal of Electronic Commerce, Decision Support Systems, Expert Systems with Applications, Information Systems Frontiers, and other journals.