

기계학습을 이용한 택배 고객의 소리 분류

홍성윤*

*고려대학교 컴퓨터정보통신대학원 디지털정보 미디어공학과

e-mail : syhong@hist.co.kr

Classification of V.O.C in The Door-to-Door Delivery Service Using Machine Learning Techniques

Seong-Yun Hong*

*Dept. of Digital Information-Media eng., Graduate School of Computer&Information Technology, Korea University

요 약

국내 택배시장 규모는 매출 3 조원 이상, 물량 13 억 상자 이상을 처리하고 있다. 2000 년 6 천억 원에서 불과 10 년 사이에 500% 이상 확대되었다. 그에 반해 소비자들의 불만 역시 증가하였다. 따라서 현재의 수작업 VOC 분류 방식으로는 적절한 대응에 한계가 있을 수 밖에 없다. 이 논문에서는 효율적인 택배불만 처리를 위해서 불만의 종류와 정도를 기계학습을 이용하여 자동분류 하는 과정 및 결과를 기술한다. 약 93,000 건의 VOC(voice of customer)를 대상으로 학습 데이터를 구축하고 여러 자질 선택 기법을 비교하였으며, 기존의 다양한 문서 자동 분류 방법들을 적용해 보았다. 실험 결과 지지벡터기계가 가장 좋은 성능을 보였고, 각각의 F-measure 값은 불만의 정도는 83.1%, 불만의 종류는 75.9% 로 측정되었다.

1. 서론

온라인시장의 발달과 B2C¹, B2B², C2C³, C2B⁴ 등 다양한 택배서비스의 등장으로 택배산업은 매년 20% 내외의 고성장을 이루는 등 국가경제에 있어서도 중요한 산업으로 자리매김하고 있다. 하지만 고객의 불만을 처리하는데 있어서는 아직 초보적인 수준이다. 대부분의 택배업체들은 지금껏 자사의 콜센터 또는 홈페이지를 통해 접수되는 VOC(voice of customer)에 대해서만 수작업으로 불만의 종류와 정도를 분류하여 관리하고 있다. 많은 비용과 적시 대응의 어려움, 오류로 인한 부적절한 대응 등 수작업이 지닌 한계로 인해서 포털사이트, 블로그 등 다양한 채널에서 들어오는 고객불만에 대해서 적절하게 대응하지 못하고 있고, 그로 인해 직접 경험하지 않은 이들에게도 회사의 이미지를 손상시키고 있다. 고객불만에 대한 대응은 고객과의 관계를 형성하고자 하는 기업목표 달성에 매우 중요한 요소가 된다.

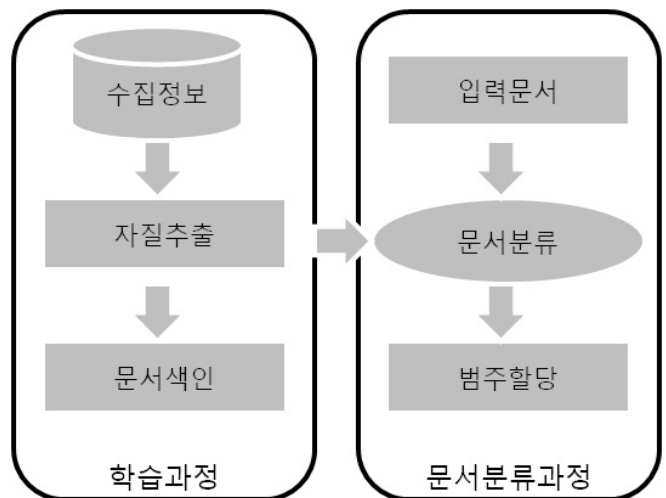
본 논문은 택배불만의 종류와 정도에 대한 분류를 문서 자동 분류의 문제로 접근하려고 한다. 기계학습 기법을 이용한 문서 분류를 하기 위해서는 문서를 자질로 표현하고, 추출된 자질을 적절한 기계학습 기법

을 이용하여 분류하는 과정이 필요하다[1]. 본 연구는 자질선택의 방법과 기계학습 방법 중 많이 쓰이는 방법들에 대해 비교 실험을 하여서 가장 좋은 방법을 찾는 것을 목적으로 한다.

2. 관련연구

2.1 기계 학습 기반 문서 분류

문서 분류는 (그림 1)과 같이 학습 과정과 문서분류 과정으로 나뉜다.



(그림 1) 문서분류 시스템 구성도

¹ 기업의 상품을 구매자 또는 일반 개인에게 배달하는 형태
² 부품 및 원자재의 운송 서비스 및 완성제품의 판매점배송
³ 개인이 개인에게 보내는 화물의 택배
⁴ 개인으로부터 기업체로 반품/반송 택배 및 폐기, 재활용품 회수

2.1.1 자질선정기법

자질 선정은 문서에 나타난 여러 단어들 중 범주화에 유용하게 사용될만한 단어들을 선택하는 과정이다. 그 방법으로는 문서 빈도(document frequency), 상호 정보(mutual information), 카이제곱통계량(χ^2 statistic), 정보 획득량(information gain)등의 기법이 있다.

2.1.2 문서색인

선택된 자질을 통해 문서를 표현하는 단계로서, 일반적으로는 벡터공간모델(vector space model)이 많이 사용된다. 이것은 한 문서를 단어의 벡터로 표현하고, 단어 마다 단어빈도(term frequency), 역 문헌빈도(inversed document frequency)등이 고려된 적절한 가중치 값을 가진다.

2.1.3 문서분류기법

문서 분류 기법은 미리 수집한 자료를 분석하여 문서 분류의 기준이 되는 사전을 작성한 후 이를 이용하여 새로운 문서가 주어졌을 경우 사전과 비교하여 해당 문서의 분류가 이루어지는 사전기반 분류(dictionary-based model)[2], 학습문서들에서 나타나는 범주간의 구별된 규칙을 전문가가 수동 구축하거나 학습을 통해 자동으로 추출된 규칙을 이용하여 문서를 분류하는 규칙기반분류(rule-based model)[3], 그리고 기계학습 기법이 있다. 문서 분류를 위한 대표적인 기계 학습 기법은 2.2 에서 소개한다.

2.2 기계학습기법

문서 분류를 위한 기계학습 기법은 나이브베이시언 확률모델(naïve bayesian probabilistic model), 지지벡터기계(support vector machines), K 근접이웃(k-nearest neighbor) 등이 있다.

2.2.1 나이브 베이시언 확률모델

대상 문서가 각 범주에 속할 확률을 구해 가장 큰 확률 값을 갖는 범주에 그 문서를 할당하는 기법이다[4]. 나이브베이시언 확률 모델은 쉽게 구현가능하고 다른 모델에 비해 적은 계산 양으로도 효과적인 성능을 기대 할 수 있다[5].

2.2.2 지지벡터기계

지지벡터기계(SVM)은 90 년대 중반에 등장하여 현재 까지 널리 쓰여지고 있으며 분류문제에 가장 좋은 성능을 보여주고 있는 기계학습 방법 중 하나이다[6]. SVM 은 기존의 경험적 리스크 최소화 원칙(empirical risk minimization) 대신 일반화 오류를 줄이기 위한 구조적 리스크 최소화 원칙(structural risk minimization)을 이용하였다. 패턴인식 등에서 우수한 성능을 보였으며, 문자인식, 얼굴인식뿐만 아니라 문서 범주화에서도 좋은 성능을 나타내고 있다[7].

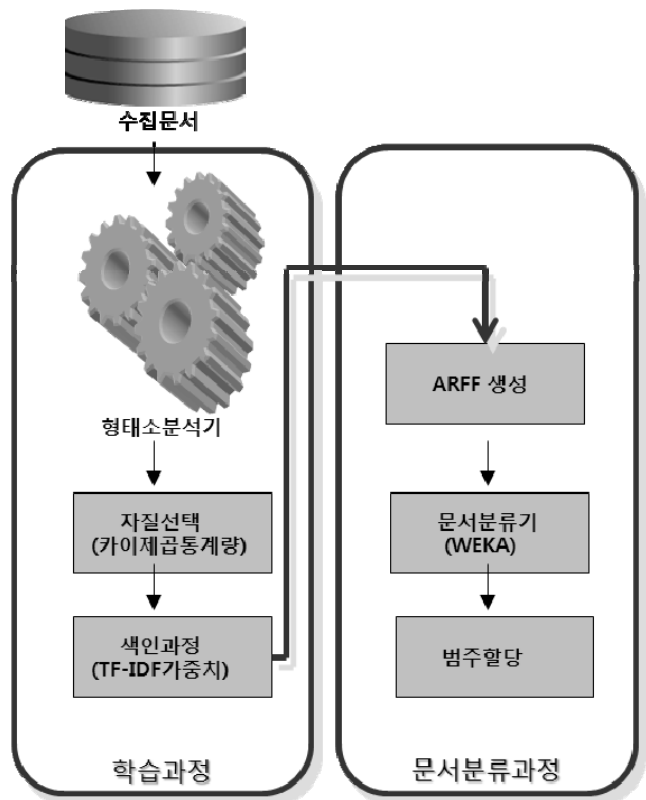
2.2.3 K-근접이웃

실험 문서가 주어졌을 때 학습 문서 중에서 테스트 문서와의 유사도가 가장 높은 k 개의 문서를 추출하

여 각 후보 범주의 순위를 정하는 방법이다. 모든 범주의 문서를 하나의 공간에 표시하기 때문에 범주의 영역을 나눌 필요가 없다. SVM 과 더불어 문서 범주화에 가장 좋은 성능을 보이는 것으로 알려져 있으나, 각 입력 문서에 대해 모든 학습 문서를 비교해야 하므로 수행 속도가 느리다[8].

3. 택배 불만의 종류와 정도 분류 시스템

본 논문에서는 택배 불만의 종류와 정도를 분류하기 위해서 (그림 2)와 같이 시스템을 구성하였다.



(그림 2) 시스템 구성도

3.1 학습 문서 집합 구축

학습 문서 집합 구축을 위해 H 사의 VOC(voice of customer) 약 93,000 건을 수집하여 불만의 정도, 불만의 종류를 수작업으로 분류를 하였고, <표 1>, <표 2>와 같이 정답 집합을 구축하였다. 불만의 정도는 아래 기준에 의해서 A, B, C 등급으로 분류하였다. 불만의 종류는 <표 2>에 나타난 바와 같이 지연 배송, 미집하, 분실 등 14 가지로 분류하였다.

- A 등급

- 집배송 업무 중 상호간의 귀책 불문 고객과의 반말, 욕설, 싸움행위
- 배송을 위한 고객 집안 무단 침입
- 규정 외 운송료 추가 징구를 통한 착복
- 허위배송완료, 허위예약처리
- 1차 고객 불만 유입 이후 미처리로 인한 재 유입 건

- B 등급
- D+1 일 이상의 미(지연)집하, 미(지연)배송 불만
- 임의 반송으로 인한 불만
- 임의 위탁 배송으로 인한 불만
- 해피콜 미실시 관련 불만
- 클레임 지연 처리 등 클레임 관련 불만
- 적정 대응 미조치 관련 불만
- 임의 전화 수신 거부 및 전화응대 불친절
- C 등급
- A, B 등급 관련 불만을 제외한 단순 불만

<표 1> 학습문서집합-불만의정도

정도	A	B	C	계
문서수	4,121	23,618	6,304	34,043
비율	12%	69%	19%	100%

<표 2> 학습문서집합-불만의종류

종류	문서수	비율
지연배송	28,815	31.1%
집배송불친절	7,816	8.4%
미집하	9,236	10%
분실	7,162	7.7%
기타	5,894	6.4%
전화응대불친절	4,392	4.7%
예약관련	3,934	4.2%
적정대응미조치	9,117	9.8%
클레임처리불만	3,665	4%
문전외배송	2,418	2.6%
파손	3,347	3.6%
영업장관련	2,598	2.8%
요구사항미수용	1,751	2%
임의위탁배송	2,429	2.7%
계	92,574	100%

3.2 자질추출

본 논문에서는 한국어 분석모듈 KLT version 2.3.0⁵를 사용하여 형태소 분석을 한 후 명사, 형용사, 동사 약 8,700,000 개의 자질을 추출하였다.

3.3 자질선택

형태소 분석을 통해 도출된 내용어 중에서 유용한 자질(feature)의 선별이 필요하다. 본 논문에서는 카이제곱통계량(χ^2 statistics)을 사용하여 자질을 선택 하였다.

카이제곱통계량은 용어 t 와 범주 c 와의 의존성을 측정하는 것으로 그 계산식은 아래와 같다[8].

$$\chi^2(t, c) = \frac{N * (AD - CB)^2}{(A + C) * (B + D) * (A + B) * (C + D)}$$

t : 내용어(Content Word)

c : 범주

N : 전체 문서 수

A : c에 속해 있는 문서 중 t를 포함하는 문서 수

B : c외의 문서 중 t를 포함하는 문서 수

C : c에 속해있는 문서 중 t를 포함하지 않는 문서 수

D : c외의 문서 중 t를 포함하지 않는 문서 수

3.4 문서색인

본 논문에서는 정보검색 분야에서 가장 기본적으로 사용하는 방법인 TF-IDF 가중치 방법을 적용하여 문서를 표현하였다. TF-IDF 가중치 방법은 문서에서 각 자질의 가중치는 해당문서에서 각 자질의 빈도와 역문헌빈도의 곱으로 나타낸다[9].

i 번째 문서에 나타나는 자질 k 의 가중치는 아래와 같다[8].

$$a_{ik} = f_{ik} * \log\left(\frac{N}{n_k}\right)$$

N : 전체 문서의 수

n_k : 자질 k가 출현한 문서의 수

4. 실험 및 결과

4.1 실험 환경

실험을 위한 시스템 환경은 <표 3>과 같다.

<표 3> 시스템환경

CPU	Intel Core2 Quad CPU Q8400 2.66Hz
RAM	2GB
OS	Window7
TOOL	WEKA3.6.6

학습 및 실험은 기본 학습데이터와 테스트 데이터의 비율을 9:1 로 하는 10-fold cross validation 방법을 사용하였고, 평가방법은 문서 분류에서 일반적으로 사용하는 정확률(precision), 재현율(recall), F-measure 을 사용하였다.

정확률은 시스템에서 추출된 데이터 중 유효한 데이터의 비율을 의미하고, 재현율은 전체 정답 수 대 시스템에서 추출한 데이터 수의 비율을 의미한다. 또한 F-measure 는 정확도와 재현도 모두를 이용하여 시스템의 성능을 하나의 숫자로 평가하는 방법이다. 계산식은 아래와 같다.

$$F - measure = \frac{2rp}{r + p} \quad r:Recall, p:Precision$$

4.2 실험 결과

4.2.1 자질 선택 실험 결과

자질 선택 방법 비교 실험 결과는 <표 4>와 같다. 표에서 보이는 바와 같이 카이제곱통계량의 결과가 F-measure 85.8%로 가장 높게 나타났다. 따라서 이후 실험 결과는 카이제곱통계량을 이용하여 자질 선택을 한 것이다.

⁵ <http://nlp.kookmin.ac.kr>

<표 4> 자질 선택 방법 실험 결과

분류		불만의정도	불만의종류	평균
상호정보	Precision	83.1	80.7	81.9
	Recall	81.8	80	80.9
	F-measure	82.4	80.3	81.3
정보획득량	Precision	82.8	79.1	80.9
	Recall	81.5	78.7	80.1
	F-measure	82.1	78.8	80.4
카이제곱통계량	Precision	87.4	85.2	86.3
	Recall	86.1	84.8	85.4
	F-measure	86.7	84.9	85.8

4.2.2 기계학습 기법별 성능평가

<표 5>, <표 6>은 불만의 정도와 불만의 종류를 나이브 베이시언, 지지벡터기계, K-근접이웃의 기계학습 모델을 사용하여 각각의 성능비교의 결과를 나타내며, 지지벡터기계 모델이 가장 좋은 성능을 보여 주고 있다.

불만의 정도에서는 F-measure 값이 비교적 고른 분포를 보이고 있으나, 불만의 종류에서는 지지벡터기준으로 가장 높은 값이 83.4% , 가장 낮은 값이 65.4%로 18%의 편차를 보이고 있다.

낮은 값의 결과를 보인 L(영업장관련)의 내용을 보면 영업장에 위치나 전화번호에 대한 질문사항, 지연배송에 대한 불만, 집배송 불친절에 대한 내용 등 다른 범주에 속하여도 가능한 내용을 포함하고 있다. L(영업장관련), J(문전외배송), M(요구사항미수용) 등이 다른 불만의 종류에 비해서 내용어의 특징이 제한적이라고 보여진다.

<표 5> 기계학습 기법별 실험결과 - 불만의 정도

정도	나이브베이시언(%)			지지벡터기계(%)			K-근접이웃(%)		
	P	R	F	P	R	F	P	R	F
A	78.7	78.9	78.7	82.4	84.1	83.2	80.5	79.1	79.7
B	79.2	80.2	79.6	83.1	85.2	84.1	81.2	78.5	79.8
C	76.5	77.3	76.8	81.7	82.5	82	82.6	83.4	82.9
평균	78.1	78.8	78.3	82.4	83.9	83.1	81.4	80.3	80.8

※P:Precision, R:Recall, F:F-measure

<표 6> 기계학습 기법별 실험 결과 - 불만의 종류

종류	나이브베이시언(%)			지지벡터기계(%)			K-근접이웃(%)		
	P	R	F	P	R	F	P	R	F
A	84.8	87.5	86.1	86.2	80.8	83.4	73.6	77.1	75.3
B	76.5	85.4	80.7	83.2	81.8	82.4	82.6	78.2	80.3
C	80.5	70	74.8	84.2	75.2	79.4	70.2	64.8	67.3
D	83.4	85.1	84.2	85.7	80.7	83.1	74.1	76.8	75.4
E	70.2	71.3	70.7	75.2	78.5	76.8	72.2	73.3	72.7
F	73.1	70.5	71.7	78.1	76.9	77.4	74.2	72.2	73.1
G	71.3	68.3	69.7	76.9	74.7	75.7	72	70.2	71
H	69.2	67.5	68.3	62.1	81	70.3	61.3	72.7	66.5
I	71.1	69.4	70.2	76.5	77.1	76.7	72.7	70.9	71.7
J	63.7	61.5	62.5	64.1	72	67.8	61.5	72.1	66.3
K	76.1	83.6	79.6	83.9	80.2	82	80.4	79.1	79.7
L	61.7	58.4	60	66.2	64.8	65.4	62.4	60.7	61.5
M	64.1	68.1	66	69.5	70.1	69.7	68.1	62.2	65
N	65.1	63.8	64.4	71.2	75.8	73.4	68.4	61.7	64.8
평균	72.2	72.1	72	75.9	76.4	75.9	70.9	70.8	70.7

※P:Precision, R:Recall, F:F-measure

※A:지연배송, B:집배송불친절, C:미집하, D:분실, E:기타, F:전화응대불친절, G:예약관련, H:적정대응미조치, I:클레임처리불만, J:문전외배송, K:파손, L:영업장관련, M:요구사항미수용, N:입의위탁배송

5. 결론 및 향후 방안

본 논문에서는 기계학습을 이용하여 택배 불만의 종류와 불만의 정도를 자동 분류하였다.

명사, 형용사, 동사의 내용어를 자질로 추출하였고, 자질선택 시 비교, 실험을 통하여 카이제곱통계량이 가장 좋은 성능을 보여서 선택하였다. 선택된 자질을 이용하여서 문서 자동 분류에서 비교적 성능이 좋고 보여지는 나이브베이시언, 지지벡터기계, K-근접이웃의 기계학습 모델을 비교, 실험하였다. 실험 결과 지지벡터기계 모델이 가장 좋은 성능을 보였다.

향후 연구에서는 자질추출에서 세분화된 형태소분석을 하여서 다양한 조합으로 실험하고, 자질선택 시에도 선택 비율을 다양하게 설정하여 단순히 많은 자질을 선택하는 것 보다 좀 더 유용한 자질이 선택될 수 있도록 하여야겠다. 기계학습 모델 부문에서도 Bagging, Random Forests, RBFNetwork 등으로 확장하여서 비교 실험을 해 볼 계획이다.

참고문헌

[1] 이경찬, 강승식, “자질 중요도 계산 기법에 의한 자동 문서 범주화”, 한국정보과학회 봄 학술발표논문집, 제 30 권 제 1 호(B), pp. 537-539, 2003
 [2] 광동민, “의견어중심의 의존트리 패턴자질을 이용한 기계학습기반 한국어 블로그 문서 의견분류시스템”, 고려대학교 대학원 석사학위 논문, 2009
 [3] Chidnanand Apte, Fred Damerau, and Sholom M. Weis, “owards Language Independent Automated Learning og Text Categorization models” Proc. Of the 17th annual international ACM-SIGIR, 1994
 [4] 박진우, “문장 중요도를 이용한 자동 문서 범주화”, 서강대학교 대학원 석사학위 논문, 2002
 [5] 이탁희, “NewsML 기반의 뉴스 자동 분류 시스템에 관한 연구”, 고려대학교 대학원 석사학위 논문, 2008
 [6] V. Vapnik, “Statistical Learning Theory”, Springer, 1995
 [7] 임혜영, “SVM 분류기를 이용한 문서 범주화 연구”, 연세대학교 대학원 석사학위 논문, 2001
 [8] 고영중, 서정연, “문서관리를 위한 자동문서범주화에 대한 이론 및 기법”, 정보관리연구, 제 33 권 제 2 호, pp. 19-32, 2002
 [9] 이재연, “기계학습을 이용한 시놉시스 기반 영화 장르 분류”, 고려대학교 대학원 석사학위 논문, 2009