

Research on the change of perception of abandoned dogs through big data analysis

Ji-Yun Jang*, Seok-Won Lee**

*Student, Graduate School of Information Communication and Technology, Ajou University, Suwon, Korea

**Professor, Dept. of Software and Computer Engineering, Dept. of Artificial Intelligence, Ajou University, Suwon, Korea

[Abstract]

This study aims to analyze the changes in public perception of abandoned dogs through big data analysis. Data from January 2017 to July 2020 were collected to analyze how the quantitative change in social issues with abandoned dogs as a keyword had an effect on public perception of abandoned dogs, and factors that influence positive/negative perceptions. As a result of the study, it was confirmed that the number of stray dogs and the number of documents related to stray dogs had a positive correlation, and specific time series changes were found through various analysis techniques such as text mining, network analysis, and sentiment analysis. This study will have significance as basic data that can be used for policy establishment or other research on abandoned dogs. we hope it will help to solve problems so as to improve awareness of abandoned dogs and develop a sense of responsibility.

▶ **Key words:** Abandoned dogs, big data, text mining, word cloud, network analysis, sentiment analysis, correlation analysis

[요 약]

본 연구에서는 빅데이터 분석을 통해 유기견에 대한 국민 인식 변화를 분석하고자 한다. 2017년 1월부터 2020년 7월까지의 데이터를 수집하여 유기견을 키워드로 한 사회적 이슈의 양적변화가 유기견에 대한 국민 인식에 어떠한 영향을 끼쳤고, 긍정/부정적인 인식에 영향을 주는 요인들을 분석하였다. 연구 결과, 유기견 수와 유기견과 관련한 문서 수는 양의 상관관계를 가지고 있음을 확인할 수 있었고, 텍스트 마이닝과 네트워크 분석, 감정 분석 등 다양한 분석 기법을 통해 구체적으로 어떠한 시계열적 변화가 있는지 알 수 있었다. 이 연구는 유기견에 대한 정책 수립이나 다른 연구에 활용될 수 있는 기본 데이터로써 의의를 가질 것이다. 유기견에 대한 인식을 개선하고 책임의식을 기를 수 있도록 문제를 해결해 나가는데 도움이 되기를 기대한다.

▶ **주제어:** 유기견, 빅데이터, 텍스트 마이닝, 워드 클라우드, 네트워크 분석, 감성분석, 연관성분석

• First Author: Ji-Yun Jang, Corresponding Author: Seok-Won Lee

*Ji-Yun Jang (quartzjiyun@ajou.ac.kr), Graduate School of Information Communication and Technology, Ajou University

**Seok-Won Lee (leesw@ajou.ac.kr), Dept. of Software and Computer Engineering, Dept. of Artificial Intelligence, Ajou University

• Received: 2021. 02. 08, Revised: 2021. 09. 06, Accepted: 2021. 09. 07.

I. Introduction

1인 가구가 급증하면서 반려동물을 키우는 가구가 늘고 있는 만큼, 버려지는 동물의 수도 해마다 늘고 있다. 우리나라의 경우, 1991년 처음으로 동물보호법이 제정되었으나, 사회적 분위기를 고려하여 제정된 법조항 대부분이 선언적인 내용에 그친 경우가 많았다. 그 이후, 정부와 동물보호단체 등의 노력으로 2008년 1월 27일부터 개정된 동물보호법이 시행되고 있다. 동물소유자의 사육 및 관리 의무를 강화하고 동물 학대와 같은 위법 행위 시 처벌 규정을 대폭 강화하였으며, 우리나라에서 동물을 유기할 경우, 동물보호법 제 47조에 의해 300만 원 이하의 과태료가 부과된다. 이러한 강화된 법이 있음에도 불구하고 버려지는 동물의 수는 매년 증가하고 있다. 동물 유기와 관련한 처벌 수준을 더욱더 강화하고 반려동물 입양 전 교육과 자격요건을 확인함으로써 입양절차와 제도 개선이 필요하다.

빅데이터가 중요하고 많이 활용되는 시대이다. 이러한 빅데이터를 활용해서 많은 연구가 이루어지고 있고, 소셜 미디어를 분석한 대중들의 인식변화 연구도 활발히 이루어지고 있다. 관광, 미세먼지 등에 대한 인식변화 연구를 포함해서 다양한 주제 분야에 대해 인식연구가 이루어지는데 필자는 위에 언급한 유기견에 대해서 연구하고자 한다. 유기견과 관련된 분야에서 빅데이터 연구가 존재하지 않으며, 본 연구가 유기견과 관련한 정책 수립, 다른 연구에 활용될 수 있는 기본 데이터로써 의의를 가질 것이다.

이를 위해서는 유기견에 대한 사회적 인식에 대해 자세히 파악해 볼 필요가 있다. 이에 본 연구는 2017년 1월부터 2020년 7월까지의 데이터를 수집하여 유기견을 키워드로 한 사회적 이슈의 양적변화가 유기견에 대한 국민 인식에 어떠한 영향을 끼쳤고, 이에 따른 유기견 수의 변화를 파악하고자 한다.

다음으로 2017~2018년과 2019~2020년을 나누어 텍스트 마이닝(Textminig) 및 네트워크 분석(Network Analysis)을 진행하였다. 시간의 흐름에 따라 어떠한 키워드들이 나타나고, 인식이 어떻게 변화하고 있는지 알아보려고 하였다. 또한, 유기견과 관련한 문서들은 어떠한 감성이 나타나는지 확인하기 위해 감성분석(Sentimental Analysis)을 진행하였다. 긍정/부정적인 감성에 어떠한 키워드들이 영향을 주는 지 연관성 분석(Association Rule)을 통해 연구하고자 하였다. 본 연구는 다양한 빅데이터 기술을 활용하여 유기견에 대한 인식을 파악하고자 한 점에 의의가 있다. 이를 바탕으로 유기견 발생 원인을 파악하고, 유기견에 대한 인식을 개선해 나가는 데에 도움이 될 수 있을 것이다.

II. Preliminaries

1. Related works

1.1 Abandoned Dog

유기동물이란 주인의 실수나 의도적인 목적으로 인해 버려진 이국적인 동물, 혹은 반려동물을 의미한다. 보통 유기동물들은 주로 주인의 죽음이나 동물들이 너무 커지거나 질병에 걸린 경우 발생한다. 유기동물은 보통 병원비를 감당하기 어려워지거나 어렸을 때와 달리 늙고 병들게 될 경우 유기를 하는 경우가 다반사다. 농림축산검역본부에 따르면 2019년을 기준으로 구조·보호 및 유실·유기동물이 13만 5,791마리로 전년 대비 12.1%가 증가했다고 집계된다. 버려지는 반려동물이 꾸준히 늘어나면서 구조·보호 비용을 포함한 운영비도 많이 늘어나 운영에 어려움을 호소하고 있다[1].

1.2 Big Data

빅데이터 공통적인 특성인 3V(Volume, Variety, Velocity)를 살펴보면 다음과 같은데, 먼저 Volume(크기)은 시스템의 데이터 처리 용량, 물리적 크기, 볼륨, 네트워크를 포함하는 대용량의 데이터양을 말하는 것으로 기하급수적으로 데이터가 증가되고 이를 수집, 저장, 분석, 처리하기 위한 기술이 빅데이터의 핵심이 되고 있다[2].

다음으로 Variety(다양성)는 사진, 동영상 등 구조화되지 않은 다양한 형태의 데이터가 포함되어 있어 일괄적인 구조로 정리하고 분석하기 어려운 빅데이터를 설명하는 특징이며 정형화된 데이터뿐만 아니라 비정형, 반정형 데이터를 모두 포함한다[3].

세 번째로 Velocity(속도)는 사물정보, 스트리밍(streaming)정보 등에 대해 실시간으로 처리되거나 장기적 접근을 요구하는 빅데이터의 특징을 나타낸다. 실시간으로 데이터가 산출되고 단시간의 수집과 분석 및 처리, 다양한 경로를 통해 수집된 데이터를 다양한 방법으로 분석하고 처리하는 차분하고 장기적인 접근이 빅데이터와 관련된 데이터의 특징이다[4].

이러한 빅데이터 처리를 위해 여러 가지 분석 기법들이 등장했으며 정형데이터, 비정형 데이터 분석에 활용되고 있다. 그 중 소셜미디어 등과 같은 비정형 데이터의 분석에 효과적인 분석 기법으로 텍스트 마이닝, 오피니언 마이닝(감성분석), 소셜 네트워크 분석 등이 주목받고 있다.

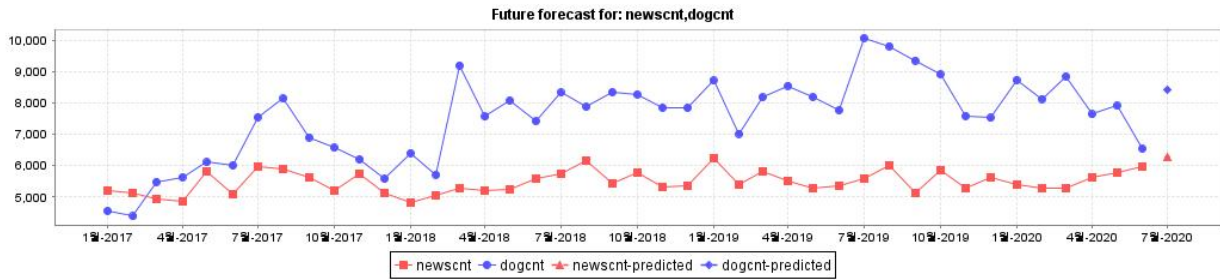


Fig. 1. Status of documents related to abandoned dogs and the number of abandoned dogs

III. The Proposed Scheme

본 연구의 목적은 빅데이터 분석 방법을 활용하여 소셜 미디어에서 유기견과 관련된 주요 키워드를 파악하고 사회적인 이슈와 주요 키워드의 관계를 분석하였다. 이러한 연구 목적을 달성하기 위해 본 연구에 대한 세부적인 문제를 살펴보면 다음과 같다.

연구문제 1. 2017년부터 2020년까지 유기견을 키워드로 다룬 사회적인 이슈의 양적 변화가 유기견 발생수와 어떠한 시계열적 변화를 보이는가? 이를 통해 향후의 유기견 수는 어떻게 예측되는가?

연구문제 2. 2017~2018년과 2019~2020년을 기준으로 유기견에 대한 인식은 어떻게 변화하고 있으며, 그에 따른 특징들은 어떤 관계를 가지고 있는가?

연구문제 3. 유기견과 관련한 주요 요인(피해, 대안)을 살펴보았을 때, 어떤 키워드가 유기견에 대한 긍정적/부정적인 감정에 높은 연관성을 보이는가?

본 연구에서 선정된 분석대상은 ‘유기견’을 주제로 한 온라인 데이터이다. 유기견에 대한 사회적 인식을 알아보기 위해 두 가지 기준을 바탕으로 데이터를 수집을 진행하였다.

첫 번째, TEXTOM을 활용하여 ‘유기견’이라는 키워드로 2017년 1월 1일부터 2020년 7월 7일까지의 데이터를 수집하였다. 네이버(블로그, 카페, 뉴스, 웹문서, 지식IN, 학술정보전체), 구글(웹문서, 뉴스, 페이스북), 페이스북, 트위터를 조사하였으며 많게는 43,000건 적게는 9,671건 정도로 총 85.84MB의 데이터를 수집할 수 있었다[5].

두 번째, 공공데이터 포털(data.go.kr)에서 동물보호관리시스템 유기동물 조회 서비스 open API를 활용하여 2017년 1월 1일부터 2020년 7월 7일까지 유기된 유기견 데이터 총 355,535건을 수집하였다. 유기동물조회 조회조건인 시·도나 시·군·구, 보호소, 품종, 유기동물 조건으로 데이터를 조회할 수 있다. 이 두 가지 유형의 데이터를 바탕으로 유기견에 대한 사회적 이슈들과 그에 따른 유기견 수 영향도 등을 연구해보려 한다[6].

IV. Experiment Results

4.1 Abandoned Dogs and Documents Related to Abandoned Dogs

Fig. 1은 수집한 유기견 관련 문서 수와 유기견 수를 나타낸 시계열 분석이다. Weka 프로그램에서 Forecast, Time Series Holt Winter를 통해 분석을 진행하였다.

유기견과 관련한 문서는 수집한 4년 치를 기준으로 월 평균 5,472건이며, 실제 유기견 수는 월 평균 7,505마리이다. 유기견 수가 가장 많았던 시기가 2019년 7월로 나타나는데, 유기견 수가 증가하면, 유기견과 관련한 문서수도 증가하고 있음을 보여주고 있다. 2020년 6월까지의 데이터 건수를 통해 시계열 분석으로 7월의 데이터를 예측해보았을 때, 유기견수는 8,415마리, 관련 문서 건수는 6,274건으로 유기견 수와 문서 건수가 증가하는 것으로 나타났다. 이를 검증해보기 위해 실제 데이터를 확인해본 결과, 7월의 유기견 수는 8,856마리였으며, 유기견 관련 문서는 6,194건으로 증가한다는 예측이 맞은 것으로 확인되었다.

이처럼 매년 7월이 다가오면 유기견 수가 급증한다. 가장 큰 사회적인 이슈는 여름휴가로 볼 수 있다. 농림축산식품부는 지난 7월 유실 및 유기동물 발생 마릿수가 1만3천700마리로 연중 최고치를 기록했다고 9일 밝혔다. 여름철에 가족 여행을 피서지에 데려가기 곤란하고 말길 곳도 마땅치 않은 경우 길에 내다 버리는 것으로 그간의 인연을 끊는 경우가 다반수이다. 이러한 이유로 7~8월의 유기견 수 및 유기견 관련 문서량이 많은 것을 볼 수 있다.



Fig. 2. 119 paramedics, dog bite accident hospital transfer patient graph

다음으로 ‘개 물림 사고’이다. 소방청에 따르면 최근 3년간 개 물림 사고로 병원에 이송한 환자가 6,883명이라고 2019년 2월에 밝혔다. Fig. 2처럼 최근 3년간 통계를 보았을 때, 2016년에는 2,111명, 2017년에는 2,404명, 2018년에는 2,368명으로 나타났으며, 꾸준히 2천명 이상의 사건 사고가 발생한 것을 확인할 수 있다. 계절별로 야외활동이 많은 5월부터 10월까지 월 평균 226명으로 연 평균 191명보다 18%(35명)가 더 많은 것으로 분석되었다[7].

상관관계란 확률론과 통계학에서 두 변수간의 어떠한 선형적인 관계를 갖는지 분석하는 방법이다. 상관관계 분석에서는 피어슨 상관계수를(Pearson correlation coefficient) 보편적으로 이용한다. 일반적으로 각 상관계수 값은 Table. 2와 같은 의미로 해석되는데 분석하는 사람에 따라 편차가 있다.

Table 1. Correlation between the documents related to abandoned dogs and the number of abandoned dogs

구분	문서	유기견 수
문서	1.000000	0.422440
유기견 수	0.422440	1.000000

Table 2. Pearson Correlation Coefficient Analysis Criteria

상관계수	기준
-1.0 ~ 0.7	강한 음적 상관관계
-0.7 ~ -0.3	뚜렷한 음적 상관관계
-0.3 ~ 0.1	약한 음적 상관관계
-0.1 ~ 0.1	상관관계 거의 없음
0.1 ~ 0.3	약한 양적 상관관계
0.3 ~ 0.7	뚜렷한 양적 상관관계
0.7 ~ 1.0	강한 양적 상관관계

피어슨 상관관계를 통해 본 유기견 관련 문서 건수와 유기견 수의 상관관계는 Fig. 3처럼 나타났고, 최종적으로 0.422440을 나타냈다. 즉, 뚜렷한 양적인 상관관계가 있는 것으로 판단되었다. 따라서 유기견 수가 증가하면 유기견 관련 문서 건수도 증가한다.

4.2 Text mining

텍스트 마이닝을 통해 단어 출현빈도가 많은 데이터들을 기준으로 나열하였다. 수집하는 과정에서 관형사나 명사 접미사 등 연구에 불필요한 단어들은 1차적으로 정제하였다. 이후, 의미가 중복되거나 국민 인식에 관련 없는 단어들을 2차 정제하였다. 유기견에 대한 국민의 인식 변화 흐름을 알아보려고 Table. 3 2017~2018년과 Table. 4 2019~2020년을 나누어 분석 진행 하였으며, 단어의 출현 빈도 순위를 매겨 40위까지 정리하였다.

이러한 TF는 수집된 데이터에서 단어의 출현빈도를 나타낸다. 2017~2018년과 2019~2020년을 나누어 단어 빈도를 분석해 본 결과, 공통적으로 ‘유기견’, ‘강아지’, ‘입양’, ‘아이’, ‘반려견’, ‘반려동물’, ‘기부’ 의 빈도가 가장 높게 나타났다. 순위에 가장 큰 변화가 있던 키워드는 ‘안락사’였다. 2017~2018년에는 16,111.2의 빈도로 순위 26에 위치하였는데, 2019~2020년에는 17위로 급상승하였다. 매년 늘고 있는 유기견 수에 따라 안락사 위기에 처한 유기견 수도 늘고 있다. 이는 사회적으로 큰 이슈이며, 이에 대한 관심도 증가하고 있음을 의미한다.

반면, 순위가 급격히 낮아진 키워드도 확인할 수 있다. ‘분양’은 2017~2018년에 26,893의 빈도로 5위에 위치하였는데, 2019~2020년에는 15,582의 빈도로 16위에 머물렀다.

다음으로 2017~2018년과 2019~2020년을 비교해보았을 때, 가장 큰 차이는 해피빈(happybean) 관련 키워드들그 재단에서 운영하는 온라인 기부 포털사이트 ‘해피빈’을 말한다[8]. 가상화폐인 콩(개당 100원)으로 기부하는 방식이며, 2019~2020년에는 ‘해피콩으로 유기견보호센터’(13,305.46)가 25위로 급상승하였다. ‘기부’ 키워드는 2017~2018년에 26,254.39로 6위였으나 2019~2020년에는 4위로 상승하였다. ‘답변’ 키워드는 2017~2018년에 21,059.94로 14위였으나 2019~2020년에는 8위로 상승하였다.

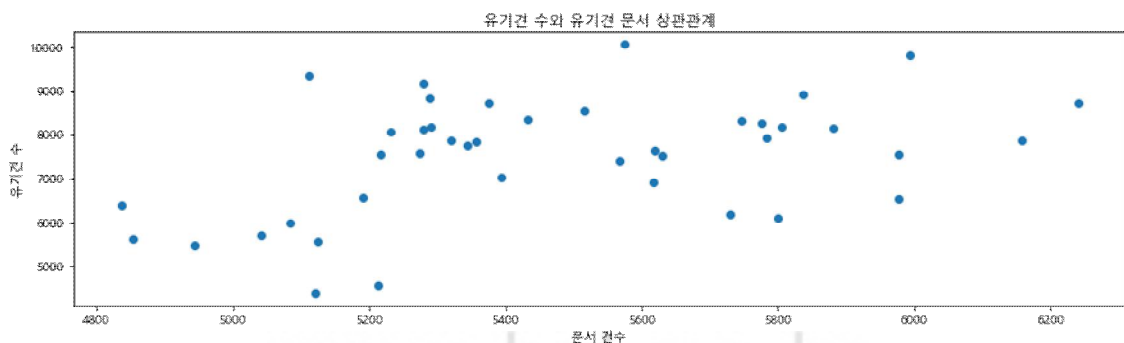


Fig. 3. Correlation between the documents related to abandoned dogs and the number of abandoned dogs

Table 3. Results of analysis of the frequency of the word 'abandoned dog' (2017~2018)

순위	단어	TF	순위	단어	TF
1	강아지	56688.34	21	유기묘	17362.18
2	유기견	44656.74	22	채택	17300.9
3	입양	41075.9	23	마리	17208.67
4	아이	29286.64	24	사진	16985.36
5	분양	26896.34	25	캠페인	16628.21
6	기부	26254.39	26	안락사	16111.2
7	반려동물	25020.93	27	토리	15694.88
8	반려견	24651.58	28	관심	15508.48
9	사람	23745.9	29	감사	15377.67
10	가족	23470.82	30	생각	15370.57
11	주인	23296.54	31	봉사활동	15275.97
12	개	22995.35	32	마음	15227.83
13	동물	21524.94	33	안녕	15186.56
14	답변	21059.94	34	부탁	15044.69
15	도움	20670.35	35	유기동물	14593.9
16	보호소	19567.33	36	대통령	14312.38
17	사랑	18450.36	37	진행	14153.61
18	집	18233.28	38	이야기	13813.75
19	오늘	17932.91	39	구조	13788.78
20	봉사	17880.37	40	행복	13471.87

Table 4. Results of analysis of the frequency of the word 'abandoned dog' (2019~2020)

순위	단어	TF	순위	단어	TF
1	강아지	44635.75	21	마음	13709.64
2	입양	35341.34	22	오늘	13591.89
3	유기견	34011.34	23	사랑	13367.21
4	기부	26120.22	24	유기동물	13325.18
5	반려견	25037.69	25	해피콩으로 유기견 보호센터	13305.46
6	아이	24232.56	26	집	12894.97
7	반려동물	22766.54	27	안녕	12727.69
8	답변	19363.79	28	봉사활동	12685.18
9	사람	19211.49	29	구조	12567.57
10	도움	18515.17	30	사료	12317.41
11	동물	18394.25	31	보호	12043.92
12	개	17891.39	32	부탁	11887.12
13	가족	17153.66	33	유기묘	11878.22
14	주인	16316.15	34	감사	11593.3
15	보호소	15592.79	35	후원	11576.47
16	분양	15582.74	36	마리	11479.99
17	안락사	14928.43	37	사진	11300.23
18	채택부탁	14065.74	38	추천	11052.38
19	생각	14015.04	39	진행	10351.81
20	봉사	13920.32	40	행복	10066.76

4.3 Network analysis results

N-gram이란, 입력된 문자열을 N개의 기준 단위로 자르는 방법이며 N개 단어의 연쇄를 의미하기도 한다[9]. 자르는 기준의 단위는 문자 단위가 될 수도, 단어 단위가 될 수도, 어절 단위가 될 수도 있다. N-gram으로 유기견 문서를 2017~2018년과 2019~2020년을 나누어 분석해보았고, 잘못된 단위로 잘린 단어들이나 유기견과 관련이 없는

단어들을 1차 정제하였다. 이를 통해 유기견과 관련한 네트워크상의 중요도가 높은 단어들을 파악하고자 하였다.

네트워크 분석 결과를 확인하고, 가장 빈도가 높았던 단어 조합을 순위 30위까지 정리하였다. Table.2와 Table.3을 보면 단어1과 단어2 그리고 빈도수를 확인할 수 있다. 단어1과 단어2의 빈도가 높다는 것은 두 단어가 나란히 등장하는 빈도가 높다는 것을 의미한다.

Table 5. Result of network analysis of the word 'abandoned dog' (2017~2018)

순위	단어1	단어2	빈도
1	도움	채택부탁	4,096
2	조금	도움	3,490
3	해피콩으로유기견 보호센터	기부	3,423
4	답변	조금	3,398
5	채택부탁	해피콩으로유기견 보호센터	3,396
6	감사	답변	2,352
7	애견샵	전문채널	1,610
8	분양	강아지	1,386
9	모두유기견아이들	기부	1,224
10	기부	좋은하루	1,224
11	좋은하루	질문	1,224
12	답변해피빈	모두유기견아이들	1,224
13	유기견	마리	1,220
14	유기견	입양	1,189
15	캠페인	진행	1,133
16	도움	채택	1,128
17	유기묘	유기견	1,117
18	유기견	가족	1,107
19	결혼	관심	1,083
20	남편	여부	1,082
21	이유	남편	1,082
22	여부	결혼	1,082
23	답변유기견	유기묘	1,048
24	답변감사	답변	1,039
25	퍼스트	도그	994
26	강아지	견종	982
27	채택	부탁	968
28	가족	캠페인	965
29	가정견	유기견	960
30	유기견	애견샵	958

먼저, 2017~2018년에 가장 빈도수가 높았던 단어들은 '도움&채택부탁'으로 4,096번 나타났다. '조금&도움'(3,490), '해피콩으로유기견보호센터&기부'(3,423), '답변&조금'(3,398) 순으로 빈도가 높았다. 다음으로 유기견 입양을 위해서는 다양한 절차를 통해 확인해야 할 사항들이 많은데, 그 중, 주거형태와 가족환경과 관련한 '결혼&관심'(1,083), '남편&여부'(1,082), '이유&남편'(1,082), '여부&결혼'(1,082)이 연관 있었다.

2019~2020년에는 2017~2018년에 비해 '기부'와 함께 나타나는 단어들이 많은 것을 확인할 수 있다. '해피콩으로유기

견보호센터&기부'의 빈도는 4,244번으로 2위를 차지하였고, '기부&감사'(1,587), '아픈유기견들&기부'(1,006), '불우이웃과유기견&기부'(614) 순으로 새로 나타난 것을 볼 수 있다.

Table 6. Result of network analysis of the word 'abandoned dog' (2019-2020)

순위	단어1	단어2	빈도
1	조금	도움	4,262
2	해피콩으로유기견 보호센터	기부	4,244
3	답변	조금	4,209
4	채택부탁	해피콩으로유기견 보호센터	4,208
5	도움	채택부탁	4,196
6	네임카드	참조	1,834
7	참조	답변	1,834
8	기부	감사	1,587
9	유기견	입양	1,379
10	아픈유기견들	기부	1,006
11	콩	아픈유기견들	991
12	채택	콩	938
13	유기견	유기묘	909
14	강아지	입양	905
15	2019	년	888
16	무료	답변	816
17	답변공유	무료	807
18	답변	결정	807
19	강아지	분양	750
20	답변	마음	618
21	클릭	친구	617
22	적립	콩	617
23	콩	불우이웃과유기견	614
24	프로필	클릭	614
25	불우이웃과유기견	기부	614
26	추가	부탁	614
27	부탁	적립	614
28	친구	추가	614
29	용인	시	578
30	강아지	주인	576

2017~2018년과 2019~2020년에서 공통적으로 나타난 단어는 '조금&도움', '해피콩으로유기견보호센터&기부', '답변&조금', '채택부탁&해피콩으로유기견보호센터'였다. 네이버의 해피빈 온라인 기부와 연관이 많은 단어들이 빈도가 높은 것으로 나타났다. 해당 단어들의 빈도수는 2017~2018년보다 2019~2020년이 더 높은 것으로 보았을 때, 해가 지날수록 유기견 입양에 대한 관심이 많아졌고, 유기견 보호센터에 도움을 손길이 많아지는 것을 확인할 수 있다.

4.4 Sentiment analysis

텍스트에서 감정 분류 분석은 베이저안 분류기(Bayes Classifier)를 통해 기계학습 기법의 감성분석 기능을 제공하고 있다. 필자는 이 감정 분류 분석을 통하여 해당 문서가 사회에 미치는 영향이 긍정적인지 또는 부정적인지에 대해 고민하여 학습 데이터를 구성하였다.

기존에 수집한 2017~2020년까지의 유기견 관련 데이터를 중심으로 감성분석을 진행하였다. “단독락스 먹이고 네 발 묶고... 천안시 유기동물보호소 충격적 ... 뉴스플러스 단독락스 먹이고 네 발 묶고... 천안시 유기동물보호소 충격적 ...” 과 같이 사회에 악영향을 미치는 문서는 부정으로 판단하였고, “한국마즈, 코로나19 의료진 응원, #덕분에 챌린지 동참 - 한국 ... 한국애견신문한국마즈, 코로나19 의료진 응원, #덕분에 챌린지 동참” 과 같이 사회에 좋은 영향을 주는 문서 내용은 긍정으로 판단하였다. 그 결과, 총 220,630건 중에 122,626건은 55.58%의 비율로 긍정을 나타냈으며, 97,080건은 44%의 비율로 부정을 나타냈다. 나머지 924건은 0.42%의 비율로 중립으로 나타남을 볼 수 있다.

Table 7. Abandoned dog keyword sentiment analysis frequency and rate

구분	빈도(건)	비율(%)
전체	220,630	100
긍정	122,626	55.58
중립	924	0.42
부정	97,080	44

다음으로 감정 단어를 분석하기 위해 유기견 관련 데이터 중 감성과 관련된 단어를 찾아서 빈도를 보여주는 기능을 활용하였다. 긍정·부정이라는 카테고리 안에, 긍정의 키워드는 흥미·호감·기쁨 3개의 단어가, 부정의 키워드에는 통증·슬픔·분노·두려움·놀람·거부감 6개의 단어가 있다. 이 기준으로 감정 단어를 분석한 결과, 유기견을 키워드로 한 감정 단어 분석은 Fig.4와 같았다.

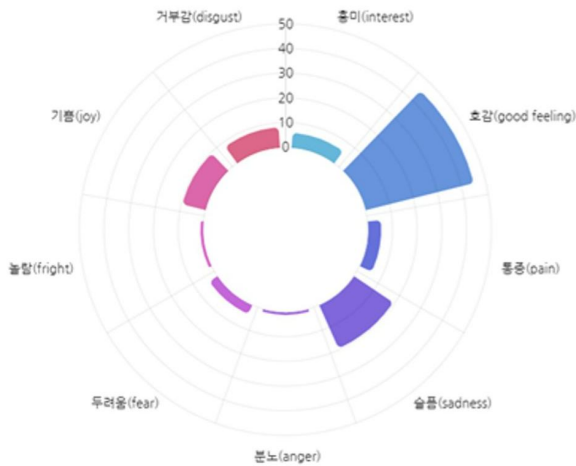


Fig. 4. Analysis of emotional words with keywords for abandoned dogs

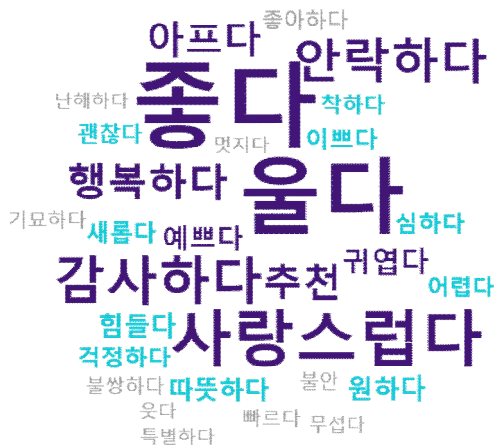


Fig. 5. Emotion word cloud with abandoned dog keywords

먼저 유기견에 대한 호감이 45.0374%로 가장 높게 나타났다. 이는 긍정적인 감정으로 많은 국민들이 유기견에 대해 좋은 인식을 가지고 있다 말할 수 있다. 두 번째로 높게 나왔던 감정은 슬픔이었다. 긍정적인 감정이 있는 반면, 안타깝고 슬프게 여기는 비율이 19.1137%로 나타났다. 다음으로 9.78035%의 비율로 기쁨이, 8.39893%의 비율로 거부감이 나타났다. 이는 긍정적인 감정과 동시에 부정적인 감정들도 모두 보이는 것을 의미한다. 이를 기반으로 세부적인 감정 빈도를 빈도순으로 15위까지 추출해 보았다. 가장 많이 나타난 감정 단어는 긍정적인 감성인 호감에 '좋다'로 26,395의 빈도를 보여주었다. 다음으로 부정적인 감성인 슬픔에 '울다'가 23,168건, 호감에 '사랑스럽다'가 16,411건, 기쁨에 '감사하다'가 12,696 순으로 나타났다. 이처럼 유기견과 관련한 긍정적인 세부 감성 단어들 많이 나타나지만, '아프다'(8,058건)나 '힘들다'(4,025건)와 같은 부정적인 감성 단어들도 존재한다.

4.5 Correlation analysis result for each factor

유기견 관련 문서를 중심으로 감정 분석을 진행하였고, 긍정/부정적인 감정 분석에 어떠한 키워드들이 영향을 주는지 분석해볼 필요가 있다. 이를 위해 유기견과 관련한 '피해' 키워드를 단어 빈도 순위에서 선정해본 결과, '안락사', '믹스견', '상처', '학대', '들개' 총 5개를 확인할 수 있었다. '안락사'는 9,281건(0.232%)으로 가장 많이 나타났다, '믹스견'이 3,294건(0.082%), '상처'가 2,241건(0.056%), '학대'가 2,236건(0.056%), '들개'가 2,200건(0.056%) 순으로 나타났다.

Table 8. Analysis of emotional relevance of 'damage' keywords related to abandoned dogs

antecedents	consequents	support	confidence	lift
'안락사'	negative	0.429679 331	0.9576616 82	0.986457 72
'학대'	'negative'	0.220024 399	0.9464017 99	0.974859 263
'믹스견'	'negative'	0.164865 807	0.9288168 88	0.956745 589
'들개'	'negative'	0.109968 63	0.9805749 81	1.010060 002
'상처'	'negative'	0.091321 018	0.8799328 3	0.906391 63
'안락사', '학대'	'negative'	0.015510 631	0.7739130 43	0.797183 923
'안락사', '믹스견'	'negative'	0.009846 636	0.4977973 57	0.512765 682
'상처', '학대'	'negative'	0.008016 731	0.5317919 08	0.547782 419
'안락사', '상처'	'negative'	0.003311 258	0.4470588 24	0.460501 486
'안락사', '들개'	'negative'	0.002701 29	0.7948717 95	0.818772 886

연관성 분석을 위해 Python을 활용하여 Association rule을 적용하였다. 향상도는 생성된 규칙의 효용가치를 판별하기 위한 용도로 metric을 lift로 설정하여 진행하였다. '피해' 키워드를 대상으로 연관성 분석을 진행한 결과, 최소 신뢰도(min_support)는 0.0001로 설정하여 분석하였다. 그 결과, 부정적인 감성의 연관성 예측으로 비교적 높은 향상도(lift)를 보인 규칙은 '들개' → '부정' 이었다. 해당 키워드 간 연관성은 지지도 0.10996863, 신뢰도 0.980574981, 향상도 1.010060002로 나타났다. 이는 온라인 문서에서 '들개'가 언급되면 '들개'가 언급되지 않은 문서보다 '부정'적인 감성이 나타날 확률이 1.01배 높아진다는 의미이다.

반면, '안락사'나 '학대', '믹스견'과 같은 나머지 키워드들의 경우, 해당 키워드들이 언급되지 않은 문서보다 '부정'적인 감성이 나타날 확률이 낮은 것으로 확인되었다.

그러한 결과가 나온 이유를 분석해 보았을 때, ‘정부 지원금+후원받지 않고 사비로 안락사 없는 유기견 보호소...’, ‘주인에게 버림받아 상처받았다고는 생각할 수 없을 만큼 사람을 잘 따르는 모습이었다.’ 와 같이 부정적인 키워드 들이지만, 결론적으로 보면 긍정적인 내용들로 판단되는 문서들이 다수 존재하는 것으로 확인되었다.

다음으로 유기견과 관련한 ‘대안’ 키워드를 단어 빈도 순위에서 선정해본 결과, ‘입양’, ‘기부’, ‘분양’, ‘보호소’, ‘캠페인’, ‘후원’, ‘치료’로 총 8개를 확인할 수 있었다. ‘입양’은 37,705건(0.942%)으로 가장 많이 나타났고, ‘기부’가 22,027건(0.55%), ‘분양’이 13,520건(0.338%), ‘보호소’가 11,181건(0.279%), ‘봉사활동’이 7,749건(0.194%) 순으로 나타났다.

Table 9. Analysis of emotional relevance of ‘alternative’ keywords related to abandoned dogs

antecedents	consequents	support	confidence	lift
'봉사활동', '보호소', '후원'	positive	0.001407	0.988571	1.53728
'기부', '봉사활동', '후원'	'positive'	0.000635	0.987342	1.535368
'캠페인', '봉사활동', '입양'	'positive'	0.000407	0.980392	1.524561
'기부', '후원', '입양'	'positive'	0.000781	0.979592	1.523317
'캠페인', '후원', '입양'	'positive'	0.000626	0.974684	1.515684
'봉사활동', '후원', '입양'	'positive'	0.000602	0.973684	1.51413
'봉사활동', '보호소', '후원', '입양'	'positive'	0.000301	0.973684	1.51413
'캠페인', '후원', '보호소'	'positive'	0.000545	0.971014	1.509978
'기부', '캠페인', '입양'	'positive'	0.000521	0.969697	1.50793
'봉사활동', '후원'	'positive'	0.003474	0.963883	1.498888
'분양', '후원', '입양'	'positive'	0.000797	0.960784	1.49407
'후원', '입양', '치료'	'positive'	0.00074	0.957895	1.489576

‘대안’ 키워드를 대상으로 연관성 분석을 진행하였고, 최소 신뢰도(min_support)는 0.0001로 설정하여 분석하였다. 그 결과, 긍정적인 감성의 연관성 예측으로 가장 높은 향상도(lift)를 보인 규칙은 ‘봉사활동’, ‘보호소’, ‘후원’ → ‘긍정’ 이었다. 해당 키워드 간 연관성은 지지도 0.001407, 신뢰도 0.988571, 향상도 1.53728로 나타났다. 이는 온라인 문서에서 ‘봉사활동’, ‘보호소’, ‘후원’이 언급되면 ‘봉사활동’, ‘보호소’, ‘후원’이 언급되지 않은 문서보다 ‘긍정’적인 감성이 나타날 확률이 1.537배 높아진다는 의미이다.

두 번째로 ‘기부’, ‘봉사활동’, ‘후원’의 키워드가 나타날 때, ‘기부’, ‘봉사활동’, ‘후원’이 언급되지 않은 문서보다 ‘긍정’적인 감성이 나타날 확률이 1.535배 높다. 해당 키워드 간 연관성은 지도 0.000635, 신뢰도 0.987342, 향상도 1.535368로 나타났다.

V. Conclusions

본 연구는 온라인상에서 유기견에 대한 국민들의 인식을 파악하고자 빅데이터를 수집하였고, 시계열 분석, 텍스트 마이닝, 감성분석, 연관성 분석 등을 활용하여 분석하였다. 이를 통해 최근 4년 동안의 유기견 수와 유기견 관련 문서 수는 어떻게 변화하고 있는지 시계열 분석을 통해 알아보고자 하였다. 2017년 1월 1일부터 2020년 7월까지 수집한 데이터 85.84MB를 분석한 결과, 유기견 관련 문서 수는 월 평균 5,472건이었으며, 실제 유기견 수는 월 평균 7,505마리로 나타났다.

이러한 연구를 바탕으로 다음과 같이 결론을 도출할 수 있다. 첫 번째, 유기견 수가 증가할수록 유기견에 대한 국민들의 관심도 많아져 온라인상 문서수도 증가한다. 유기견 수가 증가한다는 것은 사회적인 이슈와도 관련이 있으며, 이는 휴가철이나 개 물림 사고가 영향을 준다는 것을 알 수 있었다.

두 번째, 2017~2018년과 2019~2020년을 나누어 텍스트 마이닝과 네트워크 분석을 진행해보았을 때, 시간이 갈수록 유기견 수를 감소시킬 수 있는 ‘입양’과 ‘기부’에 대한 관심이 많아진 것으로 보아 긍정적으로 바라보고 있음을 확인할 수 있었다. 반면, 안락사의 단어 빈도수가 높아졌다는 것으로 보아 입양되지 못하는 유기견들의 안락사 시행에 대한 안타까운 소식들이 많아진 것으로 보인다.

마지막으로 사회적으로 순영향과 악영향을 주는지에 대해 기준을 삼고 감성분석을 진행해본 결과, 부정보다는 긍정적인 인식이 높았다. 이러한 긍정적인 인식에 영향을 주는 키워드로 ‘봉사활동’, ‘보호소’, ‘후원’의 연관성이 가장 높았고, 부정적인 인식에 영향을 주는 키워드는 ‘들개’였다. 이를 통해 유기견에 대한 관심은 온라인을 통해 꾸준히 나타나고 있음을 확인할 수 있었다. 따라서 긍정적인 인식에 영향을 주고 있는 키워드 관련 활동을 적극적으로 지원한다면 유기견 감소에도 순영향을 줄 거라 예상된다.

이 연구는 단순 설문 조사로만 연구되고 빅데이터 분석기법을 활용한 사례가 없기에 빅데이터를 통해 유기견에 대한 대중들의 인식을 분석적으로 살펴보았다는 점에서 의의를

가진다. 이는 4차 산업혁명 시대에 맞게 향후 이 분야에 대한 연구가 발전 가능하다는 것을 시사한다. 이 연구를 바탕으로 유기견에 대한 연구를 확장해 나간다면, 향후 유기견 관련 정책 수립 및 결정에 이바지할 것이라 기대한다.

ACKNOWLEDGEMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2020R1F1A1075605).

This work was supported by the BK21 FOUR program of the National Research Foundation of Korea funded by the Ministry of Education(NRF5199991014091).

REFERENCES

- [1] Organic animal definition, https://en.wikipedia.org/wiki/Organic_animals
- [2] Yang Hyeon-cheol, "The Effect of the Level of Application of Personal Information De-identification Technology on the Intention to Activate Big Data Use: From the Perspectives of Big Data Providers and Recipients", Kwangwoon University Graduate School Ph.D. thesis, 2015
- [3] Philp, R., "Big data analytics", TDWI Research, 2011
- [4] Park Ki-myo, "Analysis of National Intangible Cultural Heritage Performing Arts Trends Using Big Data," Chung-Ang University's Graduate School of Korean Traditional Music, Master's thesis, 2017
- [5] Textom, <https://www.textom.co.kr/>
- [6] Animal protection management system abandoned animal inquiry service open API, <http://data.go.kr/>
- [7] Reporter Jeon Jin-woo, 119 paramedics, dog bite accident hospital transfer patient graph, 2019
- [8] Definition of Happy Bean, <https://ko.wikipedia.org/wiki/%ED%95%B4%ED%94%BC%EB%B9%88#%EC%97%AD%EC%82%AC>
- [9] Textom Manul, https://www.textom.co.kr/home/video/TEXTOM_v5.0_manual.pdf

Authors



Ji-Yun Jang received the B.S. and M.S. degrees in Media, and Intelligent software of Information and Communication Technology from Aju University, Korea, in 2018 and 2021, respectively. She is specialized in high

performance computing. She is currently an Engineer in MIRACOM Inc Co., Ltd.



Dr. Seok-Won Lee is currently a Full Professor and Chair in the Dept. of Software & Computer Engineering and Dept. of Applied Artificial Intelligence, and Head of Graduate School of Software at Aju

University, Republic of Korea since 2012. His areas of specialization include software engineering with specific expertise in ontological requirements engineering and domain modeling, and knowledge engineering with specific expertise in knowledge acquisition, machine learning and knowledge-based systems, and information assurance with specific expertise in software security & privacy. He has published more than 180 peer reviewed articles. He is a professional senior member of IEEE, ACM and AAAI.