

다중 모달리티 학습 기반 동영상 주제 분류
(Classification of Subjects of Videos
Based on Multi-modality Learning)

지도교수 : 장병탁

이 논문을 공학학사 학위 논문으로 제출함.

2008년 12월 23일

서울대학교 공과대학
컴퓨터공학부
김동희

2008년 12월

목차

01. 목차	01
02. 초록	02
03. INTRODUCTION	03
04. 관련 연구	05
05. 실험의 구성과 내용	07
06. 실험의 결과와 분석	
06.1. VISUAL WORD를 이용한 분류	09
06.2. 키워드를 이용한 분류	12
06.3. VISUAL WORD와 키워드를 모두 사용한 분류	15
06.4. 단일 모달리티만 사용하는 분류와 복합적인 분류의 비교	17
07. 결론	20
08. 참고문헌	21

02. 초록

여러 가지 모달리티로 표현될 수 있는 대상에 대해서, 각 분야에서는 고유의 모달리티를 기준으로 대상을 파악하고 측정하는 것이 일반적이다. 그러나 이것은 대상에 대해 우리가 얻을 수 있는 모든 정보를 활용하지 못하는 방법이다. 사용 가능한 정보를 모두 활용하기 위해서는 한 가지 모달리티만을 통해 대상을 인식하는 것보다는 그것들을 복합적으로 고려하는 것이 더 효과적일 것이다. 여기에서는 동영상 이미지를 분석하는 방법 중 Visual word를 사용하는 방법과 그 자막으로부터 얻는 키워드 정보를 이용하는 방법, 두 가지의 모달리티를 활용하여 이를 확인해 보았다. 각각의 방법으로만 동영상 이미지를 분류하는 경우와, 두 가지 모두를 사용하는 경우를 비교하여 성능의 향상을 확인하였다. 이로부터 여러 가지 관점으로 분석 가능한 대상에 대한 측정, 인식기를 개발하는데 있어서 분야 통합적 관점이 유용함을 알 수 있다.

03. Introduction

비록 그 대상이 같은 것일지라도 대개 서로 다른 분야에서는 제각각 다른 속성에 주목하여 분석 대상을 파악한다. 또한 대상에 대한 인식이나 구분을 하는 데 있어서도 고유의 모달리티를 사용한다.

그러나 여러 가지 모달리티를 통하여 식별 가능한 대상이라고 했을 때, 우리가 대상에 대해 가진 정보의 100%는 그 모든 모달리티에서 얻을 수 있는 정보를 포함한다. 반대로 말하자면 한 개의 모달리티만을 사용하는 경우 우리는 대상에 대해 우리가 얻을 수 있는 정보를 전부 활용하지 못하고 있는 것이 된다.

간단히 생각해서 사람이 세상을 인식하는 데 있어서 사용하는 정보는 음향 정보와 시각적 정보와 물리학적 정보 등 여러 가지로 구성되어 있다. 즉 사물은 여러 가지 모달리티를 가지고 있다. 인간의 손에 의해 창조된 어떤 측정기가 이 중에서 한 가지만을 사용하여 세계를 인식한다면, 그 측정기는 결코 인간의 관점에서 파악하는 세계를 이해할 수 없다. 우리가 측정기를 통해 구현하고자 하는 능력의 수준은 궁극적으로 인간의 인식과 동등한 것이 될 것이다. 그러나 각 모달리티를 복합적으로 사용하지 않는다면 인간이 사용하는 정보에 비해 매우 적은 양의 정보만을 사용하여 사물을 인식하려는 결과를 빚게 된다.

따라서 복합적인 모달리티를 활용하여 대상을 인식하는 것이 단일 모달리티를 활용하는 방법에 비해 좋은 접근 방법이라는 것을 생각할 수 있다. 여기에서는 단순한 예를 들어 이를 확인하고자 한다. 이를 위하여 복합적 모달리티로 확인할 수 있는 대상이 필요한데, 데이터화가 가능한 멀티미디어 정보가 유용하다고 하겠다.

동영상 이미지는 시각 정보와 청각 정보를 포함한다. 이것들은 서로 별개의 접근 방법으로 분석될 수 있는 모달리티에 해당한다. 시각 정보를 식별하는 방법으로는 visual word를 사용하는 방법이 있다. 또 정보 전달이 목적인 영상인 경우 기교적 청각 신호는 거의 들어가지 않기 때문에, 이때의 청각 정보는 사실상 동영상의 자막을 활용함으로써 충분히 데이터화 할 수 있다.

여기 제시한 실험에서는 몇 가지 클래스로 분류 가능한 다큐멘터리 동영상 이미지를 사용한다. Visual word를 사용하여 동영상 이미지를 분류해보고, 또 자막의 키워드를 사용하여 동영상 이미지를 분류해보았다. 그리고 이 두 가지를 모두 사용하여 분류를 시도해 보았다. 그리고 두 가지 모달리티를 모두 사용한 경우의 결과와, 각각의 모달리티만을 활용한 경우의 결과를 비교하였다.

각각의 모달리티만을 활용한 경우에 비하여 이를 종합했을 때에는 분류 능력도 전반적으로 향상되었고, 또한 분류의 편차도 줄어들었다. 이때 특히 각각의 모달리티만을 활용했을 때 좋지 않은 결과를 내던 클래스에 대해 큰 향상을 보였다.

이에 따라 여러 가지 모달리티를 조합하여 사용하는 것이 측정기의 인식 능력에 도움이 된다는

것을 알 수 있다. 이러한 효과는 보다 복잡한 모달리티를 도입하면 아마도 그 효능이 증가할 것으로 생각된다. 또 개별 모달리티만을 활용할 때 좋지 않은 효능을 보이더라도 이를 조합함으로써 개선을 꾀할 수 있음을 나타내기도 한다.

이 논문의 나머지 부분은 다음과 같이 구성되어 있다. 4장에서는 논문에서 활용된 각종 방법론들에 대한 관련연구를 조사하여 서술하였으며, 5장에서는 이를 통하여 실험을 구성하고 사용할 데이터를 설계한다. 6장에서는 차례대로 visual word를 사용하는 방법의 결과를 분석하고, key word를 사용하는 방법의 결과를 분석하며, 또한 둘을 함께 사용한 결과를 살펴볼 것이며, 이들을 비교해 볼 것이다. 7장은 이로부터 얻을 수 있는 결론에 대해 설명하겠다.

04. 관련 연구

04.1. Text mining

텍스트 자료에서 나타나는 단어들을 위주로 그들간의 참조, 인접성 등의 관계나 출현의 성향을 데이터화 하고, 그것을 이용하여 정보 검색, 자료 분류, 키워드 추출 등의 작업을 하는 것을 말한다. 예를 들어 어떤 글에 출현하는 단어들에 대해, 그 단어가 관련된 분야(들)에 점수를 부여하여, 최대 점수를 받은 분야에 속한 글이라고 판단하는 등의 방법이다¹.

여기에서는 이미지에 대응되는 자막의 내용으로부터 단어를 얻고 데이터화하여, 이로부터 해당 자막이 어느 클래스의 이미지에 대응되는 것인가를 알아내는 Classification 작업을 한다.

04.2. Visual word

시각 정보를 일종의 문장이나 단어와 같은 방식으로 인식하는 개념이다. 시각적 부분 요소들을 몇 가지의 Visual Lexicon으로 인식하고, 어떤 이미지로부터 그런 Lexicon들을 추출해 낸다. 이것들을 몇 단계로 조합하여 부분 이미지에 대한 word를 만들고, 이런 word의 조합으로 이미지에 대한 기술을 하게 된다².

아래는 이미지와 그것을 Visual word를 사용하여 인식한 예시를 나타낸다.



그림 1

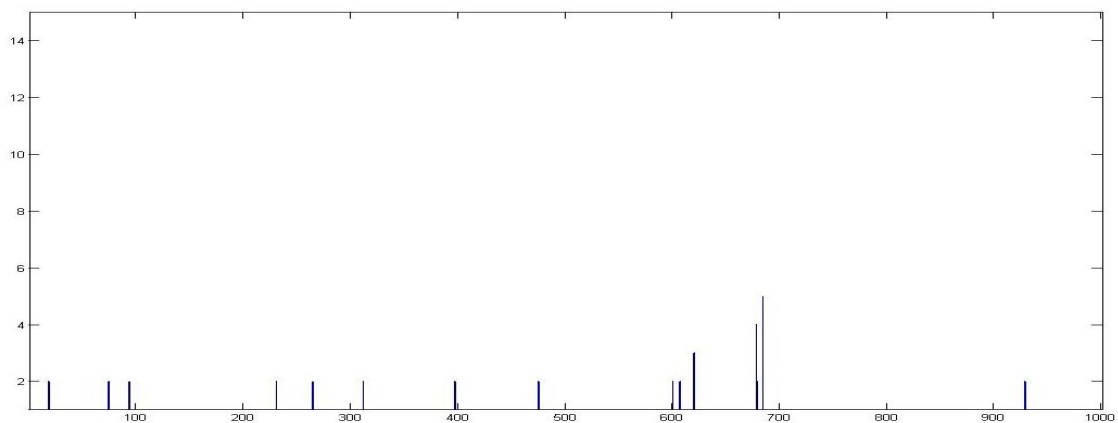


그림 2

04.3. Machine Learning

실험에 사용된 방법들은 기본적으로 Machine Learning 개념을 사용하고 있다. 이것은 중간 과정에 대해 사람의 직접 조작 없이 컴퓨터가 스스로 사고 과정을 구성하게 하는 방법이다. training 데이터가 주어졌을 때, 이를 이용하여 결과를 유도하는 함수의 가중치들을 바꾸어 가는 방법이다. 함수의 구조나 가중치를 변화시키는 방법이나 결과 유도의 개념 등은 구체적 방법론마다 다르다.

04.3.1. Support Vector Machine

Support Vector Machine은 속성값의 축으로 표현되는 다차원 공간상에 training set 데이터들을 배치시킨다. 그리고 거기에 classification 함수를 배치하는데, 이때 함수에 의해 공간적으로 구분되는 영역들이 각각 class에 해당한다. 함수의 가중치들을 바꿔감에 따라 training set 데이터들은 조금씩 다른 class 영역에 속하게 되는데, 이때 적절한 가중치를 택함으로써 보다 나은 분류가 가능해진다³.

04.3.2. KNN

KNN은 K-Nearest Neighbor 방법을 말한다. 주어진 패턴들과 새로운 패턴의 유사 정도를 측정하여 새로운 패턴을 적절한 위치에 배치시킨다. 이때 새로운 패턴이 위치하게 되는 위치는 유사한 것으로 판단된 기존 패턴들이 속한 클래스들이 많이 분포한 곳이 된다. 주위의 패턴들의 클래스를 확인함으로써 새로운 패턴이 어느 클래스에 속하는지를 판단하게 된다⁴.

04.3.3. Decision Tree

의사 결정 트리 기법이다. 트리 구조를 놓고 꼭대기에서부터 가지를 타고 내려가면서 점차적으로 판단을 해 가게 된다. 각 분기점에서는 데이터의 attribute를 이용한 분기 함수를 만나고, 이에 따라 현재 판단중인 데이터가 가는 길이 정해진다. leaf 노드에서는 어느 클래스인지의 정보가 판별된다. 이때 분기 함수들을 적절히 조정함으로써 최종적으로 여러 데이터가 각각 자신의 클래스를 잘 찾아가도록 조정할 수 있다. Weka에서는 J48로 구현되어 있다⁵.

04.3.4. Naive Bayes

조건부 확률 규칙에 의해 현재 데이터가 특정한 클래스에 속할 확률이 얼마인가를 계산한다. 모든 클래스에 대해 이 확률을 계산한 후 기대값이 높은 클래스를 택해 분류한다. classifier에서는 조건부 부분의 공통된 분모를 제거하고, 각 attribute가 독립적이라는 가정을 통해 계산을 단순화할 수 있다는 점을 이용할 수 있다.

04.3.5. Bayesian network

Naive Bayes에서 사용한 조건부 확률 문제를 좀더 확장해서 attribute들을 directed acyclic graph 구조로 표현된다. Bayesian network에서의 학습과정은 크게 두 가지로 구분되는 데 먼저 attribute들 간의 연관관계의 구조를 학습하여 이를 DAG로 표현하는 구조학습과 학습된 구조를 바탕으로 하여 확률테이블을 구하는 과정으로 구성된다. 구조학습을 통해 연관관계를 가진 노드들은 서로 연결되고 연결된 노드들에 대하여 training data를 기반으로 조건부확률 테이블을 계산을 통해 학습이 진행된다.

05. 실험의 구성과 내용

05.1. 문제 정의

단일 모달리티 정보를 가진 상태에서의 분류 능력과 복합적 모달리티 정보를 가진 상태에서의 분류 능력을 비교하고 다중 모달리티 사용의 효능을 확인하고자 한다. 이를 위해 여러 모달리티를 가진 분석 대상이 필요하며 또한 이를 분류하는 데 사용할 방법도 필요하다.

05.2. 데이터의 구성

여기서 선택한 대상은 동영상이다. 동영상 콘텐츠는 컷 이미지의 visual word를 가지고 있으며 또한편으로 그 내용에 대한 자막의 정보도 가지고 있다. 우리는 visual word를 통해 이미지가 어떠한 분류의 이미지인지 판별할 수 있다. 한편으로 자막의 내용은 동영상의 내용과 관련된 정보를 포함하기 때문에 이를 통해 컷 이미지가 어느 분류의 정보인지 추론하는 것이 가능하다. 이것들은 서로 다른 모달리티의 정보이고, 별개로 혹은 함께 고려되어 컷 이미지를 분류하는 기준이 될 수 있다.

구체적 대상 동영상으로는 BBC 방송의 다큐멘터리를 이용하였다. 다큐멘터리는 비교적 일관적이고, 형식주의 영화 같은 비현실적인 변칙적 이미지가 덜 등장한다. 이는 현실에서 인식의 대상으로 고려되는 것들의 양상과 일치하며 따라서 좋은 실험 예가 될 수 있다.

사용한 데이터에서 한 개의 컷 이미지의 정보는 한 레코드에 정리된다. 각 모달리티 정보는 분류 정보와 쌍을 이루어 레코드를 형성한다. 데이터 레코드의 순서는 무작위로 되어 있다. 컷 이미지의 분류(class)의 종류는 세 가지로, 정글(Jungle), 바다(Sea), 우주(Space)가 있었다.

05.3. 실험 설계

우선 이미지의 속성들에 대한 정보와 그 이미지가 해당하는 분류의 정보를 가진 데이터를 사용하여, 데이터 마이닝 기법의 여러 가지 방법들로 classification을 하였다. 그리고 똑같은 방법으로 자막 텍스트의 키워드에 대한 정보와 그 분류 정보를 가진 데이터를 사용하여 실험하였다. 이미지와 텍스트 정보가 함께 있는 데이터를 이용해서도 실험하였으며 이것은 여러 모달리티를 가지고 분류를 수행하는 경우에 해당한다. 이 세 가지를 비교, 분석하였다.

실험 도구로는 weka를 사용하였으며, 사용된 기계 학습 방법은 Support vector machine(weka에서는 SMO), KNN, J48 Decision Tree, Naive Bayes, Bayesian network 가 있다. 데이터의 70%를 training set으로 하고 나머지를 test set으로 사용하였다. random seed를 사용하는 경우 10가지 seed를 사용하는 경우를 각각 실험하여 평균을 구하였다. (KNN은 각각 10가지 K에 대해 평균을 구하였는데, image 의 경우 표준편차는 약 11, text와 merge의 경우 약 5.5였다.)

- 다음은 실험 구성의 요약이다.

데이터 셋 구성 (이름(attribute, ...) 형식)	image(visual word, class),	
	text(key word, class),	
	merge(visual word, key word, class)	
각 데이터 셋의 레코드 수	336	
Class 종류	Jungle, Sea, Space	
각 Class 에 해당하는 레코드 수	Jungle	120
	Sea	56
	Space	160
Training set size	70%	
Test set size	30%	
사용한 기계 학습 방법	SMO, KNN, J48, Naive Bayes, Bayesian network	

표 1

06.1. Visual word를 이용한 분류

다음은 이미지 정보만을 이용하여 분류한 실험 결과이다.

	BNet	J48	KNN	NaiveB	SMO
Correctly Classified Instances	61.61%	71.43%	49.70%	63.69%	74.70%
(stdev)		0	11.00		0
Incorrectly Classified Instances	129.00	96.00	169.00	122.00	85.00
Kappa statistic	0.42	0.52	0.22	0.45	0.59
Mean absolute error	0.25	0.20	0.34	0.25	0.29
Root mean squared error	0.49	0.41	0.45	0.48	0.38
Relative absolute error	0.62	0.01	0.01	0.60	0.71
Root relative squared error	1.07	0.01	0.01	1.05	0.83
Total Number of Instances	336.00	336.00	336.00	336.00	336.00

표 2

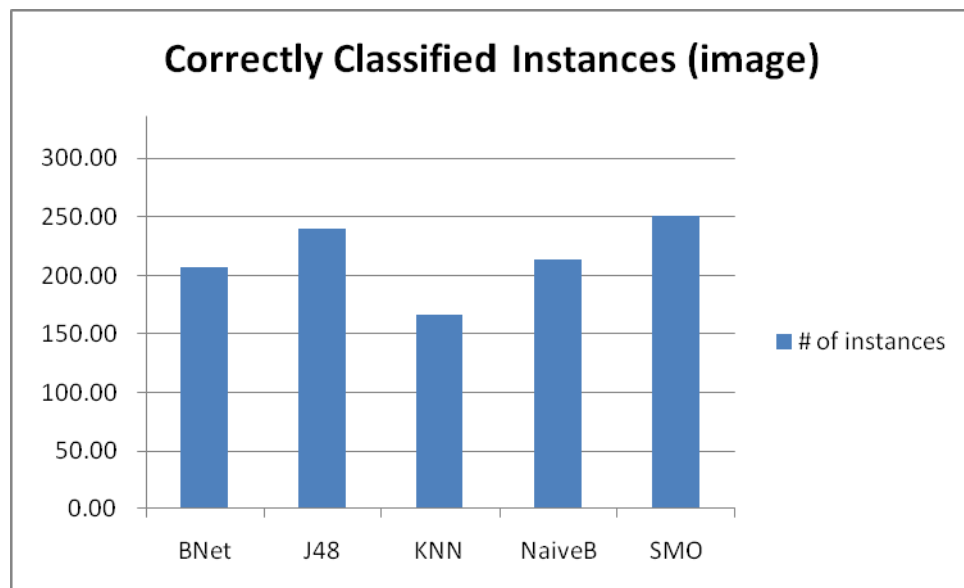


그림 3

일단 모든 분류의 이미지가 SMO 방식에서 가장 잘 분류되었다. 이때 결과에서 Correctly Classified Instances가 251장으로 나와서 옳게 분류된 이미지 수가 가장 많았다. 반면 Kappa statistic를 감안했을 때 KNN은 다른 방법에 비해 특별히 낮은 성과를 보였다.

SMO에서, 모든 경우에 TP Rate는 FP Rate에 비해 매우 높은 수치를 보였다. 이는 실제로 해당 클래스인 이미지를 그렇다고 판단하는 것이, 틀린 이미지를 옳다고 판단하는 것보다 잦다는 것을 보인다. 아닌 선택의 경우의 수는 2가지이고 맞은 선택은 1가지임을 감안했을 때 이러한 경향은 무작위적인 것이 아니고 의미가 있는 것이라고 볼 수 있다. 즉 SMO 방법에 의한 분류가 무작위

분류에 비해 의미 있는 효력을 가진다. 비슷하게 높은 분류 효과를 보인 다른 방법들에서도 이러한 양상이 보인다.

모든 분류에서 SMO의 경우에 F-Measure 값이 가장 컸다. F-Measure는 실제로 해당 클래스인 것과 해당 클래스로 분류된 것들이 일치하는 비율을 나타낸 것으로, 이것이 높으면 효과적 분류로 볼 수 있다.

J48의 경우 의사 결정 트리 방식인데, 두 번째로 높은 F-Measure값을 보였다. 이때 Precision과 Recall 값의 경향이 각각 다르다. Jungle이나 Space의 경우 Precision 값은 Recall 보다 약 7% 낮다. 그런데 Sea의 경우 Recall이 Precision보다 40% 정도나 낮다. 즉 실제 바다인 이미지가 바다로 분류되는 경우가 낮았다는 것이다. 다른 분류 방식을 통한 평균적인 양상 역시 그러하다.

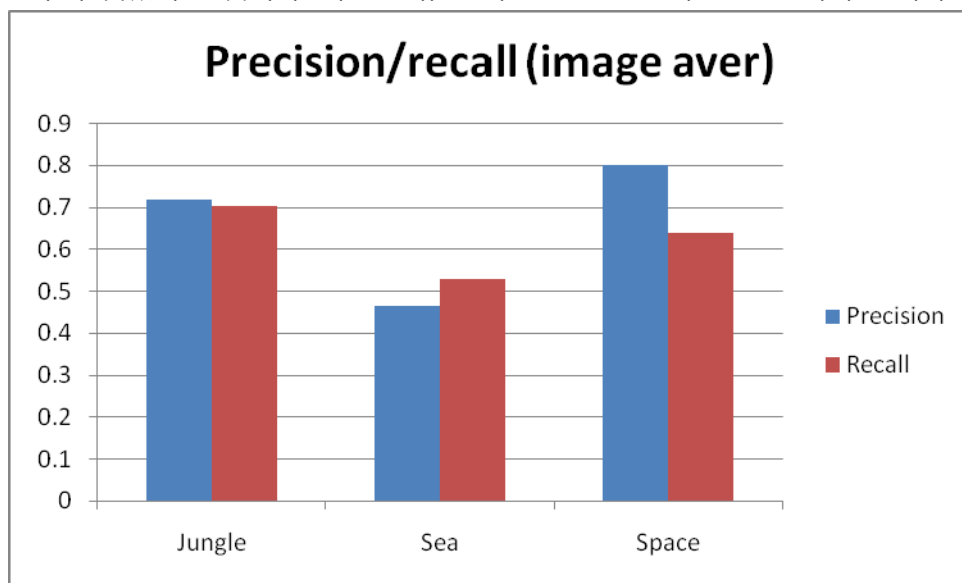


그림 4

그렇다면 바다 이미지가 어느 쪽으로 분류되는 경향이 있었던 것일까? J48 의 Confusion Matrix를 보면 다음과 같은 형태이다.

결과 \ 실제	Jungle	Sea	Space
Jungle	86	7	27
Sea	12	26	18
space	30	2	128

표 3

여기에서 Sea인 이미지들이 다른 쪽, 특히 Space쪽으로 분류되는 경우가 많았던 것으로 보인다. 해석을 하자면 바다의 이미지는 우주나 정글 이미지로 잘못 인식될 여지가 많다는 것을 나타낸다. 그럼에도 불구하고 Space나 Jungle인 이미지들이 Sea로 분류되는 경향은 매우 적었다. 클래스당

레코드 수를 보면 sea의 경우 56개로 다른 클래스에 비해 수가 적다. 데이터 특성상의 이러한 한계 때문에 sea는 좋지 않은 분류를 나타내고 있다.

class별 F-Measure의 평균값 비교나, TP/FP rate의 비교(특히 J48의 경우에서 두드러진다)에서도 sea의 경우 다른 것보다 특별히 낮은 수치를 보인다. 모든 이미지를 분류하는데 있어서 sea로 분류되는 경우 자체가 적었다는 것이다.

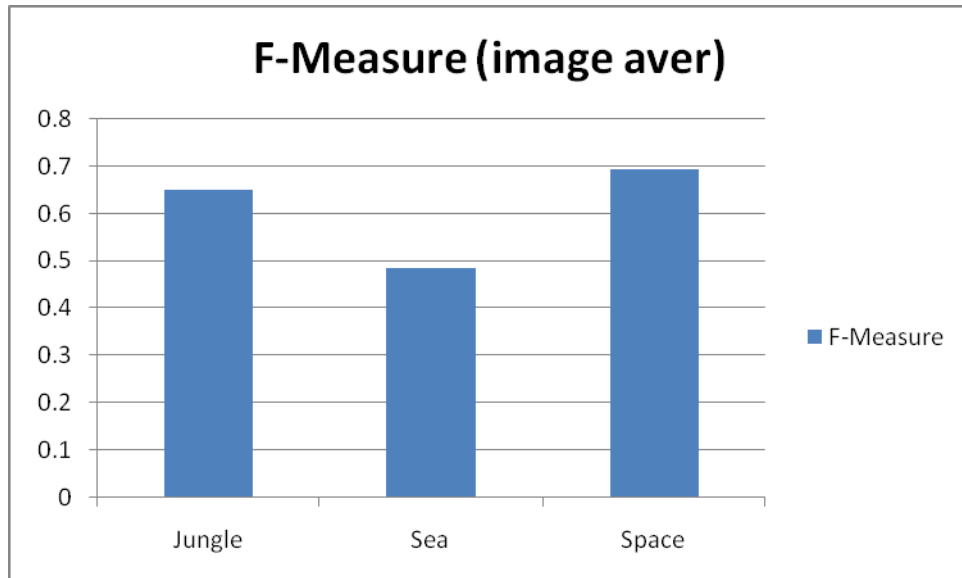


그림 5

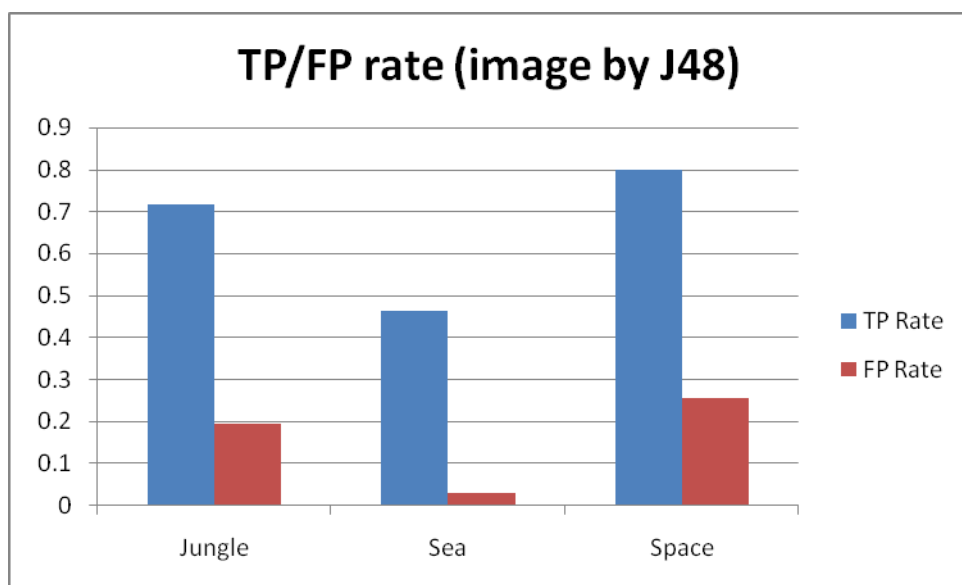


그림 6

KNN의 경우 전체 속성을 한꺼번에 놓고 비교하거나 유기적 관계로 놓고 비교하는 것이 아니라 국지적 성질들의 비교에 의해 분류가 결정된다. 이미지의 경우 이러한 방식의 분류가 좋은 결과를 내지 못했다. 반면 SMO가 높은 성과를 보여주었다는 것은 이미지의 각 속성들 간에 유사한 경향성을 가진 것들이 많음을 나타낸다.

06.2. 키워드를 이용한 분류

다음은 자막 키워드만을 이용한 분류 결과이다.

	BNet	J48	KNN	NaiveB	SMO
Correctly Classified Instances	77.68%	72.02%	50.77%	73.81%	77.68%
(stdev)		0	5.31		0
Incorrectly Classified Instances	75.00	94.00	165.40	88.00	75.00
Kappa statistic	0.62	0.52	0.08	0.54	0.63
Mean absolute error	0.17	0.24	0.35	0.19	0.29
Root mean squared error	0.35	0.39	0.48	0.37	0.37
Relative absolute error	0.42	0.59	0.85	0.48	0.70
Root relative squared error	0.76	0.87	1.05	0.82	0.82
Total Number of Instances	336.00	336.00	336.00	336.00	336.00

표 4

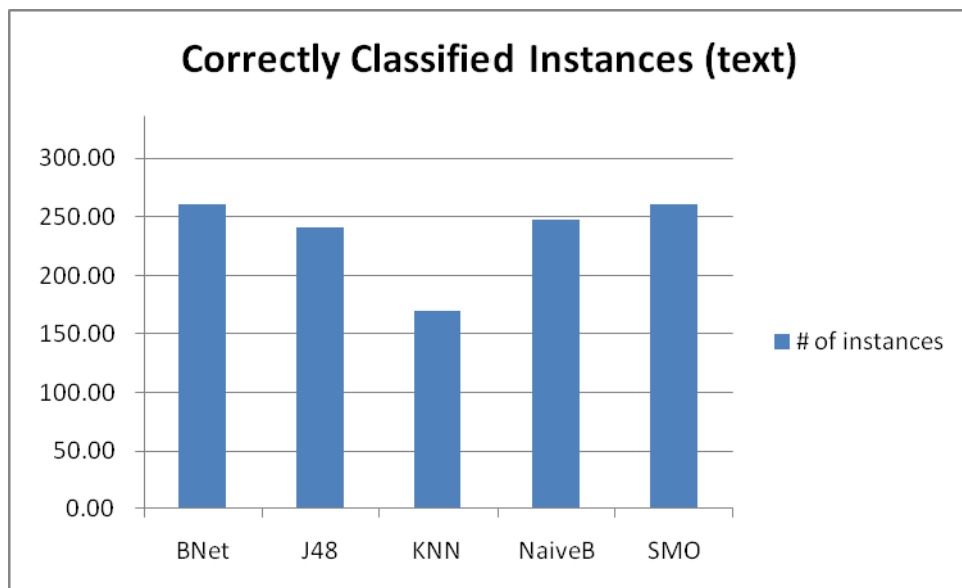


그림 7

Bayesian network 방식과 SMO 방식에서 똑같이 261개로 가장 높은 성공 수를 보였다. 이것은 70%이상의 비율에 달한다. 반면 Kappa statistic로 미루어 볼 때 KNN을 사용하는 것은 거의 신뢰성이 없다.

Bayesian network은 Jungle과 Space에 대해 SMO 방식보다 높은 F-Measure를 보여 더 효과적인 방법으로 생각되나, Sea에서는 SMO가 더 좋은 성과를 보였고 결국 같은 성공률을 초래했다.

각 클래스 별 Precision과 Recall을 비교했을 때 다른 클래스와 Sea는 양상의 차이가 두드러진다.

Jungle이나 Space는 Precision대 Recall의 비율이 10% 내외의 차이로 비교적 근소한데, Sea의 경우 높은 분류 정도를 보이고 또한 그 차이가 적은 SMO 방식에서도 200%가 넘는 차이를 보인다. 게다가 Sea에서 이 비율의 양상은 다른 클래스에 비해 반대이다. 바다로 분류되는 경우가 대단히 적은 반면 일단 선택되었을 때의 정확도는 높았다고 할 수 있다. 그러나 실제 이 선택의 개수는 기껏해야 5개 정도로 너무 적어서 그 정도에 대한 논의는 정확하지 않은 것으로 보인다.

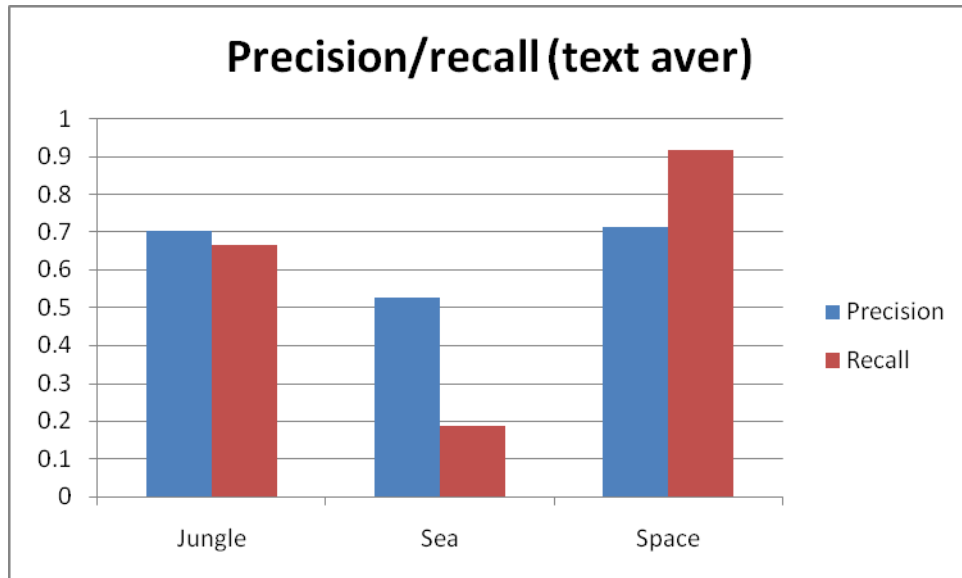


그림 8

Naive Bayes 방식에서도 특별히 Sea를 전혀 선택하지 않았다. 그럼에도 불구하고 종합적인 성공 횟수는 Bayesian network이나 SMO에 맞먹었다. 즉 바다 관련 자막 이외의 경우에는 SMO보다 Naive Bayes나 Bayesian network이 우세했다는 것이다.

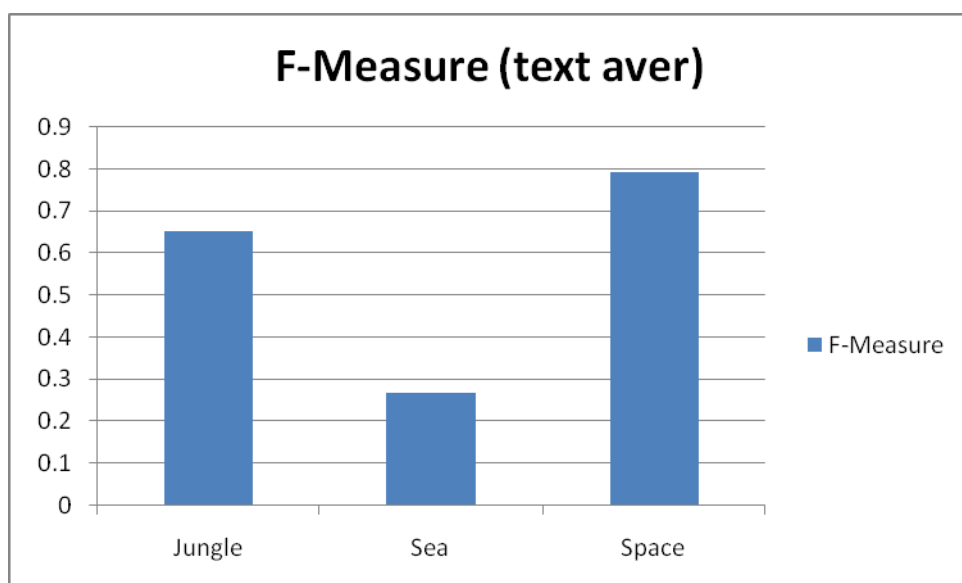


그림 9

여기에서 우리는 적은 양의 정보만 가진 클래스나 그 정보의 관련 정도가 희박한 경우에는 Bayesian network나 Naive Bayes가 SMO 방식에 비해 적합하지 않다는 것을 알 수 있다. 반대로 충분한 양과 복잡도가 있는 데이터라면 SMO 방식보다 Bayesian network나 Naive Bayes가 좋을 것이라고 예상할 수 있다.

이외에, 클래스 별 Precision과 Recall 비율을 보면 Jungle과 Sea는 Precision비율이 높은 데 비해 Space는 이것이 역전되어 있다. 이것은 실제로 우주 클래스가 아닌데도 우주로 간주하고 택하는 경향이 높다는 것을 의미하고, 우주 클래스에 출현하는 많은 텍스트가 다른 클래스에도 출현한다는 것을 의미한다. space의 경우엔 선택 횟수 자체도 높은 편이다. 반면 우주가 다른 클래스로 오인되는 경우는 적은데, 그만큼 우주 관련 자막에 출현하는 텍스트들이 모호한 관련성을 지니는 범용적 언어라고 볼 수 있다.

06.3. Visual word와 키워드를 모두 사용한 분류

다음은 이미지와 텍스트 정보를 모두 사용한 분류의 결과이다.

	BNet	J48	KNN	NaiveB	SMO
Correctly Classified Instances	76.19%	80.65%	48.93%	76.79%	82.14%
(stdev)		0	5.64		0
Incorrectly Classified Instances	80.00	65.00	171.60	78.00	60.00
Kappa statistic	0.62	0.68	0.20	0.61	0.71
Mean absolute error	0.16	0.18	0.35	0.16	0.27
Root mean squared error	0.38	0.34	0.44	0.38	0.35
Relative absolute error	0.39	0.43	0.86	0.38	0.67
Root relative squared error	0.83	0.75	0.97	0.83	0.77
Total Number of Instances	336.00	336.00	336.00	336.00	336.00

표 5

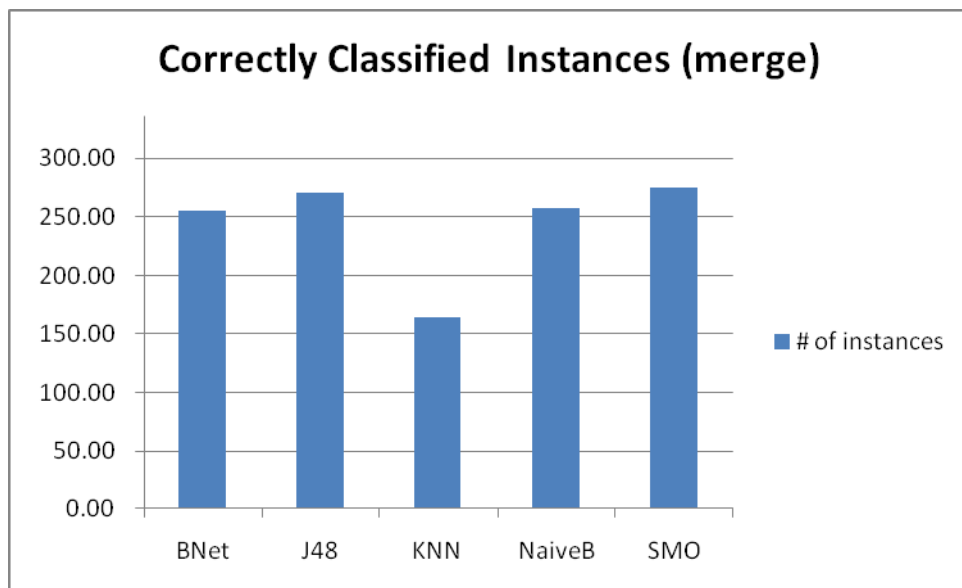


그림 10

SMO가 276개의 Correctly Classified Instance 수를 가지며 가장 높은 성공률을 보였다. J48이 근소한 차이로 두 번째 정확도를 보였다. KNN의 경우 여전히 좋지 않은 효과를 나타냈다.

KNN외의 데이터들에서, Precision 과 Recall 이 거의 같은 수치를 보였다. 너무 쉽게 선택되는 경우나, 선택한 것 중에서 잘못된 것의 비율이 줄었다는 것이다. 전체적으로 높은 TP Rate가 나왔으며, FP Rate는 매우 낮았다.

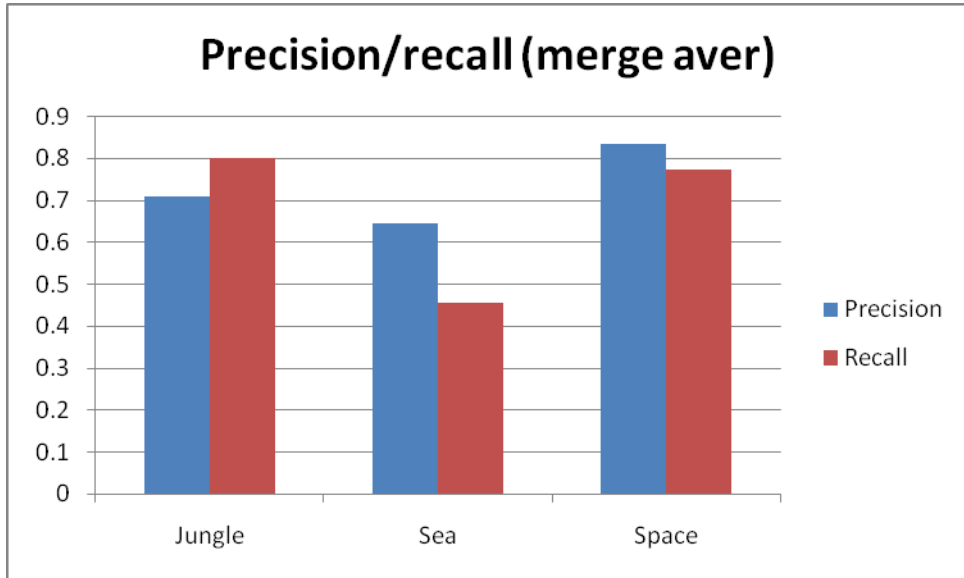


그림 11

정글의 경우 다른 클래스와 달리 Precision이 Recall보다 작게 나왔는데, 신뢰도가 낮은 KNN을 제외한 평균을 구하면 Precision은 0.7758, Recall은 0.7875로 1~2% 사이의 차이가 나타난다. SMO로 한정할 경우 차이는 1% 이하이며 실제 선택의 개수를 감안할 때 오차범위로 보인다.

이번에도 앞 테스트들과 같이 Sea의 경우 다른 것보다 낮은 F-Measure를 보였다. 다만 이미지 정보만 사용한 실험의 경우처럼 바다 클래스의 선택 자체가 적다거나, 텍스트 정보만 사용한 실험처럼 아예 없는 경우들이 나오거나 하지는 않았다. KNN을 제외한 F-Measure는 0.583으로, Jungle이나 Space에 비해 약 30% 낮은 정도였다.

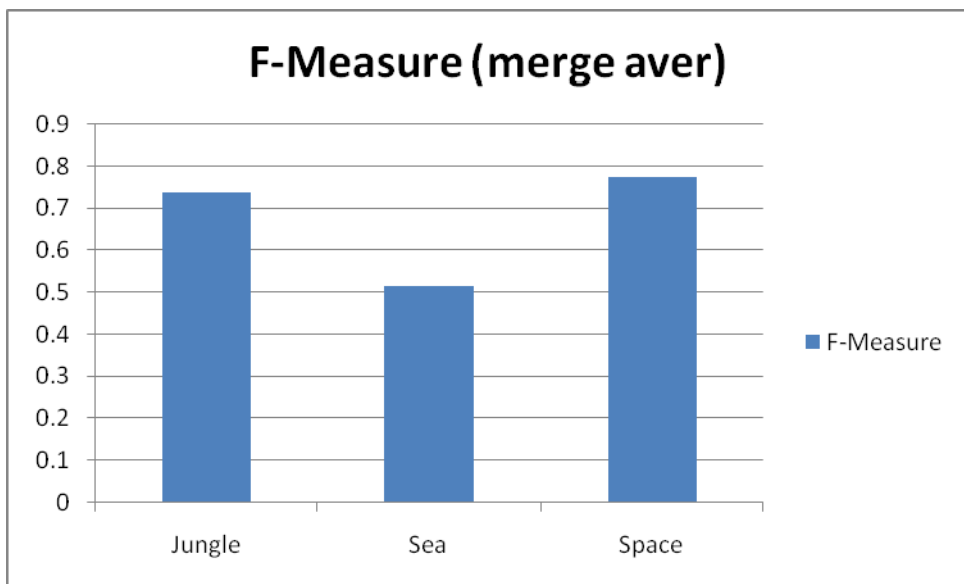


그림 12

06.4. 단일 모달리티만 사용하는 분류와 복합적인 분류의 비교

다음은 앞서 행한 세 실험의 비교를 나타낸다. M은 Image와 text의 평균이다. merge/M은 두 가지 모달리티를 모두 사용함으로써 변화된 비율을 나타낸다.

	image	text	merge	M	merge/M
Correctly Classified Instances	228.00	253.00	265.25	240.50	0.10
Incorrectly Classified Instances	108.00	83.00	70.75	95.50	-0.26
Kappa statistic	0.50	0.58	0.65	0.54	0.22
Mean absolute error	0.25	0.22	0.19	0.24	-0.19
Root mean squared error	0.44	0.37	0.36	0.40	-0.10
Relative absolute error	0.49	0.55	0.47	0.52	-0.09
Root relative squared error	0.74	0.81	0.80	0.78	0.02
Total Number of Instances	336.00	336.00	336.00	336.00	0.00

표 6

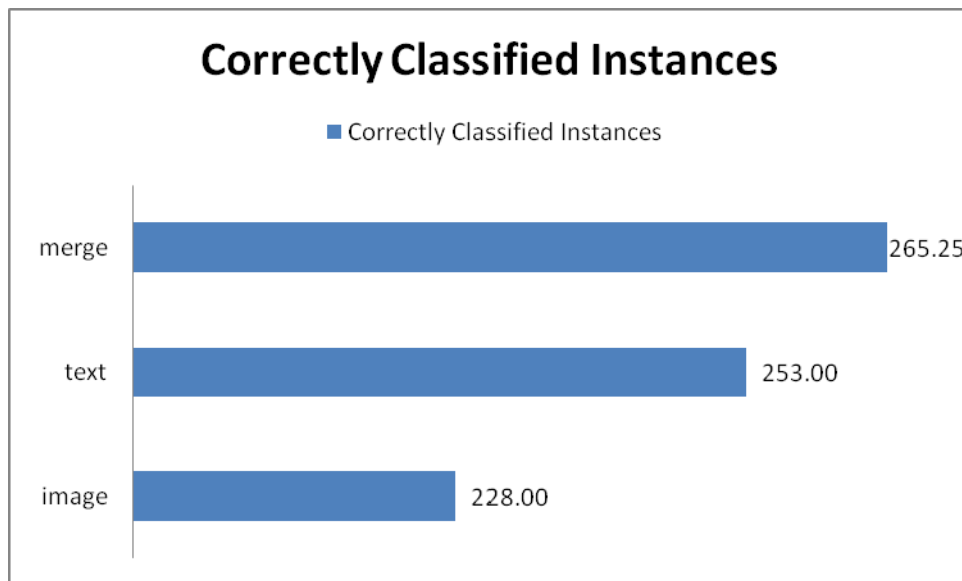


그림 13

Kappa statistics의 비교를 해볼 때 merge 데이터는 image only나 text only의 경우에 비해 대단히 높은 신뢰도가 나왔다. KNN을 제외하고 생각하면 평균 0.6이상이며, 가장 높은 SMO의 경우 0.7 이상에 달한다. KNN은 image text merge 세가지 모두의 경우에서 약 0.2나 그 이하의 매우 낮은 Kappa statistics를 보였기 때문에 이후로는 KNN을 제외하고 이야기하도록 하겠다.

우선 merge의 경우 merge의 경우 전체 레코드의 수는 336개로 image나 text에 대한 실험에 사용한 개수와 같다. 그리고 바르게 분류된 레코드 수는 image와 text의 평균보다 약 10% 증가하였다. 동시에 잘못 분류된 경우의 수는 25% 감소하여 최종적으로 Kappa statistics는 약 1.2배 가량 증

가하여 0.65%에 이르렀다. 그만큼 분류의 신뢰도가 증가하였다. error 경향을 나타내는 척도들도 대체로 감소하였다. 모든 클래스에서 TP Rate는 증가하고 FP Rate는 감소하고 있다.

특기할만한 것은 이런 평균값대비 증가뿐 아니라 최대값에 비해서도 증가 경향을 보였다는 것이다. text의 경우에 높은 분류 정도를 보여줬는데 merge는 그보다 많은 분류 성공 수를 보였다. F-measure의 비교에서도 모든 class에 대해 이와 같은 양상을 보인다. merge일 때에는 image와 text중의 최대값 이상으로 증가하고 있다.

여기에서 image 데이터와 text 데이터를 종합 분석하는 결과가 개별적인 분석의 결과보다 효과적임을 알 수 있다.

이때 변화의 정도는 image, text 각각을 개별 분석한 결과들에서 가장 난관을 보였던 class Sea에서 가장 두드러진다. 가장 높은 F-measure를 보였던 Jungle의 경우 F-measure가 7% 정도 증가하였으나, 가장 낮은 F-measure를 보인 Sea의 경우에는 37% 이상 증가하였다. Jungle의 경우에는 그 중간에 있었다. 개별 분석 시 낮은 신뢰도를 보이는 class일수록 image와 text 조합의 상승효과가 크다는 것을 알 수 있다.

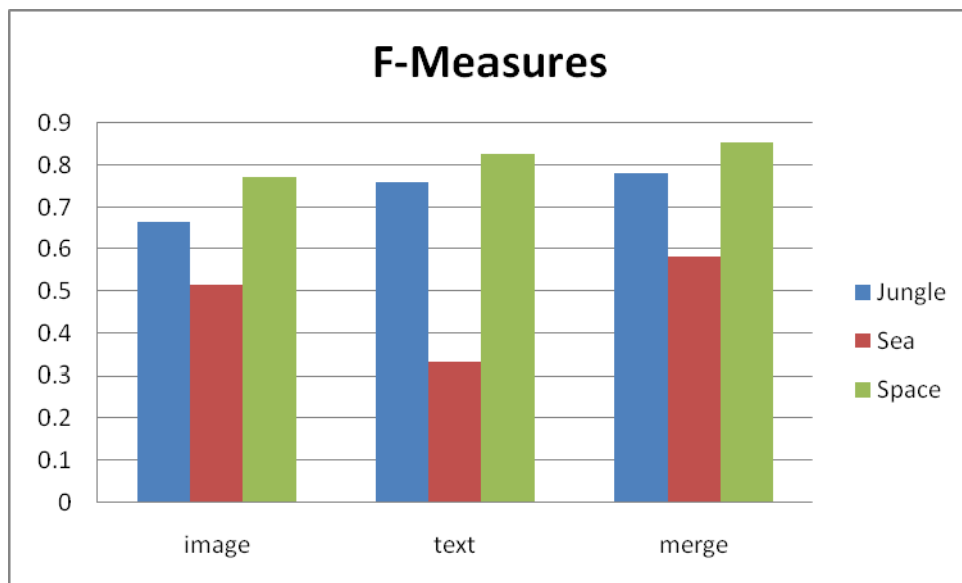


그림 14

클래스간 F-Measure들의 편차를 비교했을 때 image only 에서는 약 0.13 이 나왔고 text에서는 그보다 두 배 이상 높은 약 0.27 이 나왔다. 그런데 merge에서 조합 효과로 인한 영향은 이 편차를 image와 같은 0.13 이하로 줄이는 데에는 실패한 것 같다. merge에서는 0.14 정도가 나왔다. 그러나 text에 비하면 절반 이하이기 때문에 복합적 분석이 충분히 효과를 발휘했다고 볼 수 있다.

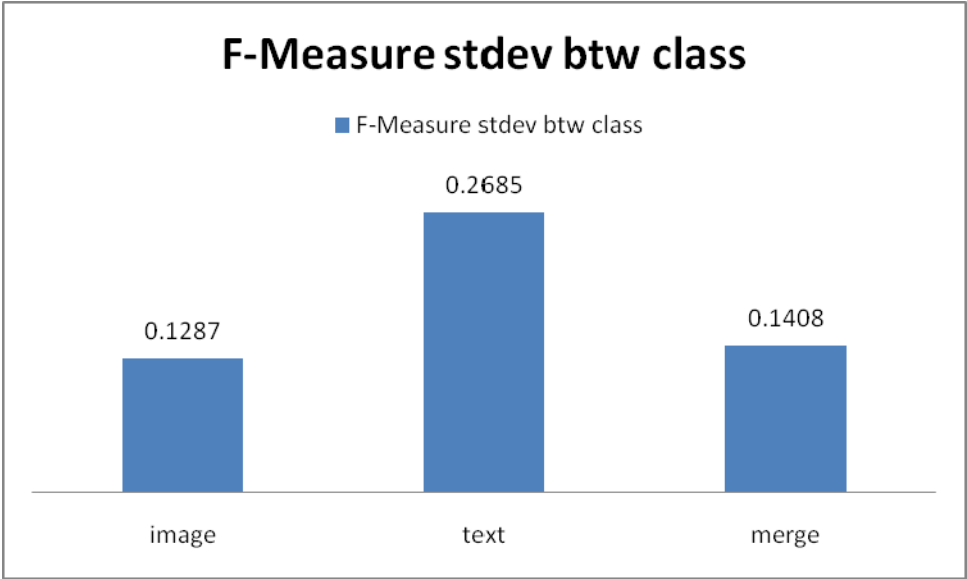


그림 15

07. 결론

이렇게 해서 Visual word만을 사용하여 분류를 하는 방법과 키워드만을 이용하는 방법을 분석했고, 이런 경우와 둘을 복합적으로 사용하는 방법의 효과의 차이를 알아보았다.

복합적 모달리티를 이용했을 때는 분류된 수만 해도 단일 모달리티를 이용할 때에 비해 10% 이상의 개선을 보였다. Kappa statistics 상으로도 SMO의 경우 그 이전에 달성하지 못했던 0.7을 달성해서, 상당히 신뢰할만한 정확도를 보여주었다. 이것은 모든 분류에 대해 공통된 개선으로 설령 단일 모달리티 상황에서 좋은 결과를 냈던 분류라도 이 경향이 달라지는 경우는 없었다. 특히 단일 모달리티 이용 시 불안정한 정확도를 보였던 분류일수록 그 상향 정도가 급격히 높아진다는 점이 주목할만하다. 또한 분류간의 F-measure도 줄어드는 경향을 보였으니 분류의 신뢰도 면에서 큰 상승을 얻었다고 하겠다.

이러한 현상은 분야 한정적인 대상 인식 방법을 사용하는 측정기들을 개선할 수 있는 방향을 제시해준다. 다각적 방면에서 변별 가능한 대상이라면, 그런 다각적 모달리티들을 함께 고려함으로써 인식 결과의 편차를 줄이거나, 인식률 자체를 높일 수 있다는 것이다.

이것은 개별 모달리티를 사용하는 경우보다 나은 결과를 낸다는 점에서 높은 인식률을 요하는 곳에서 차용할만한 방법이다. 또한 기존의 단일 모달리티를 이용하는 방법에서 별 성과를 내지 못하더라도 이러한 방법으로 기존의 것 이상의 인식률을 낼 수 있을 것으로 기대한다.

08. 참고 문헌

- Nils J. Nilsson, "Artificial Intelligence: A New Synthesis", Morgan Kaufmann, 1998.
- Ian H. Witten, Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, 2005.
- Weka program home page : <http://www.cs.waikato.ac.nz/ml/weka/>
- AH Tan, "Text mining: The state of the art and the challenges", Proceedings of the PAKDD, 1999.
- T Nasukawa, T Nagano, "Text analysis and knowledge mining system", IBM Systems Journal, 2001.
- Ronen Feldman, Moshe Fresko, Yakkov Kinar, Yehuda Lindell, Orly Liphstat, Martin Rajman, Yonatan Schler, Oren Zamir, "Text Mining at the Term Level", lecture notes in computer science, 1998.
- C. Apte, F. Damerau, and S.M. Weiss, "Text Mining with Decision Trees and Decision Rules", Conference on Automated Learning and Discovery Carnegie-Mellon University, 1998.
- ²Junsong Yuan, Ying Wu, Ming Yang, "Discovery of Collocation Patterns: from Visual Words to Visual Phrases", Computer Vision and Pattern Recognition, IEEE, 2007.
- Thorsten Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features", Springer Berlin, 1998.
- ³Nello Cristianini, John Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge University Press, 2000.
- ⁴Cover T., Hart P., "Nearest neighbor pattern classification", Information Theory, IEEE, 1967.
- 오승준, "범주형 시퀀스 데이터의 K-Nearest Neighbor 알고리즘", 한국컴퓨터정보학회, 2005.
- T Mitchell, "Decision Tree Learning", Machine Learning, Washington State University, 1997.
- ⁵Ian H. Witten, Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, and Sally Jo Cunningham, "Weka: Practical Machine Learning Tools and Techniques with Java Implementations", Department of Computer Science, University of Waikato, 1999.
- V Sugumaran, V Muralidharan, KI Ramachandran, "Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing", Mechanical Systems and Signal Processing, 2007.
- I. Rish, "An empirical study of the naive Bayes classifier", T.J. Watson Research Center, 2001.
- A McCallum, K Nigam, "A comparison of event models for naive bayes text classification", AAAI-98 Workshop on Learning for Text Categorization, 1998.
- N Friedman, D Geiger, M Goldszmidt, "Bayesian Network Classifiers", Machine Learning, 1997.
- D Heckerman, "Bayesian Networks for Data Mining", Data Mining and Knowledge Discovery, 1997.