Comparison of Different Semantic Negative Concepts Selection Methods in SVM Classifier Training for Image Annotation

Shan-Bin Chan† Shin'ichi Satoh‡ and Hayato Yamana† ‡

† School of Fundamental Science and Engineering, Waseda University 3–4–1 Okubo, Shinjuku-ku, Tokyo, 169–8555, Japan
‡ National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
E-mail: † {chrisjan, yamana}@yama.info.waseda.ac.jp, ‡ satoh@nii.ac.jp

Abstract When SVM is adopted for image annotation, most researchers randomly choose negative sample images for classifier training. Adopting different negative sample image datasets will vary annotation accuracy. This research discusses the accuracy and mean reciprocal rank (MRR) between different negative sample images selection methods. This research adopted ImageNet dataset for positive and negative sample images, and implement SVM for classifiers training. Then we adopted WordNet for building semantic hierarchical tree, and then propose six different negative sample images selection methods. The results show that the accuracy of baseline method (random sampling) is 0.48 and the best proposed method is 0.51.

Keyword Negative Sample Selection, ImageNet, WordNet, SVM, Image Annotation, Machine Learning

1. Introduction

SVM prediction is a very popular technique for image annotation. One-class SVM method is always adopted to annotate images. Researchers prepare positive and negative sample images for one-class SVM classifier training. There are several methods for positive sample image collection. Researchers can use pre-collected datasets or gather positive sample images form image search engines. Some researchers also use the famous online album service such as Flickr or Picasa for positive sample images collection. And the same collection methods are adopted for the negative sample images collection.

The qualities of the positive sample images are important in SVM classifier training. That's a given that better qualities of positive sample images will produce better SVM prediction accuracy. So that specialists select the positive sample images for SVM classifier training. But there are few studies discuss about the negative sample images collection.

In this study, we will discuss the accuracy and mean reciprocal rank between different negative sample images selection methods. A semantic hierarchical tree structure is built by using WordNet[5] for negative sample images selection. We adopted ImageNet[7] for the positive and negative sample image collection. And One-class SVM is employed for classifier training and prediction.

The chapter 2 will discuss the precious studies and related works. The chapter 3 will introduce our propose methods for comparing different negative sample image selection methods. The chapter 4 will show the experiment results. And the final chapter will discuss the conclusions.

2. Related Works

In order to efficient access to large image database, content-based image annotation becomes popular research topic recently. The feasibility and effectiveness of automatic image annotation is also hot topic in computer vision research. Li et al. proposed a two-level hierarchical ensemble model composed of probabilistic SVM classifiers and co-occurrence language model to annotate images automatically.[13]

In the automatic image annotation process, Le et al. also proposed approaches to incorporate lexical semantics into image annotation. They proposed using semantics-constrained K-means clustering in combination with hierarchical ensembles that can enhance the quality of clustering to some degree and the hierarchical ensemble annotation architecture, and also provide more accurate and consistent annotations.[13]

In order to achieve high annotation accuracy, researchers adopted active learning for effective image annotation. They selected semantically ambiguous images for users to label, and then used the user feedbacks to retrain the image classifiers to improve the annotation accuracy.[6][10][11]

To label a image manually is laborious. WordNet is adopted by researchers to automatically enhance object labels.[1]

Image annotation by SVMs that simply employs one low-level image feature is insufficient. Le et al. proposed fusion of multi low-level image features for one-class SVM classifier training and prediction.[3][4]

There are very few studies discuss about the negative sample images selection methods in one-class SVMs when implementing image annotation. Most researchers randomly collected negative sample images for one-class SVM classifier training and prediction. Although two-class SVMs is also adopted for multiclass image annotation by some researchers. In this study, we will only discuss the one-class SVMs for image annotation.

In this research, we adopted ImageNet as the image dataset, and adopted WordNet to automatically label sample images into "positive" and "negative" for classifier training. Finally, we propose six negative sample images selection methods and compare the accuracy between different negative sample image datasets.

3. Methods

Our approach is shown as Fig 1. And our research flow is as the following steps:

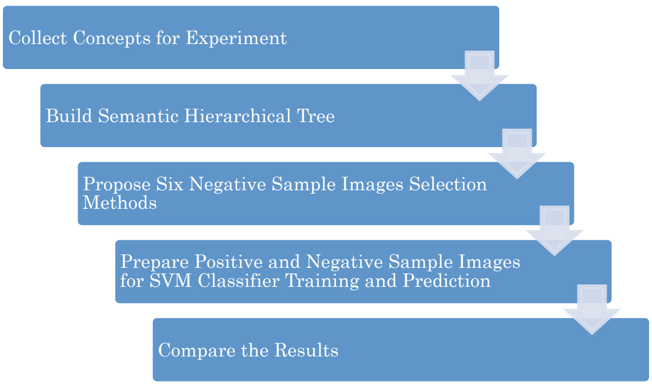

Fig 1. Research Procedure

- Collect popular concepts from recently uploaded photos on Flickr.
- Build semantic hierarchical tree by employing WordNet.
- Select 100 concepts for experiment.
- Retrieve sample images by employing ImageNet.
- Propose six negative sample images selection methods for the precision and accuracy comparison.
- For each pre-selected concept, prepare positive and negative sample images for SVM classifier training.
- Compare the SVM predict results.

The following subsections will introduce our approach in detail.

3.1 Collect Concepts

We collect concepts from the photos on Flickr by using Flickr API. We employ the "flickr.photos.getRecent" API to retrieve the recently uploaded photos randomly. We only retrieve the tags of recently uploaded photos and we use the POS (part-of-speech) application[14] to tag the retrieved tags. And we store the image tags, counts, and POS tags to MySQL database. We have already collected 176,303 concepts from July 2009 to July 2010. We will only use the noun as the concepts for sample images retrieval. Finally, the amounts of noun concepts are 6,838.

3.2 Build Semantic Hierarchical Tree

After the collection of concepts, we employ WordNet to build a semantic hierarchical tree. Apart of semantic hierarchical tree is shown as Fig 2. In this semantic hierarchical tree, we can understand the relationship and distance between each concept.

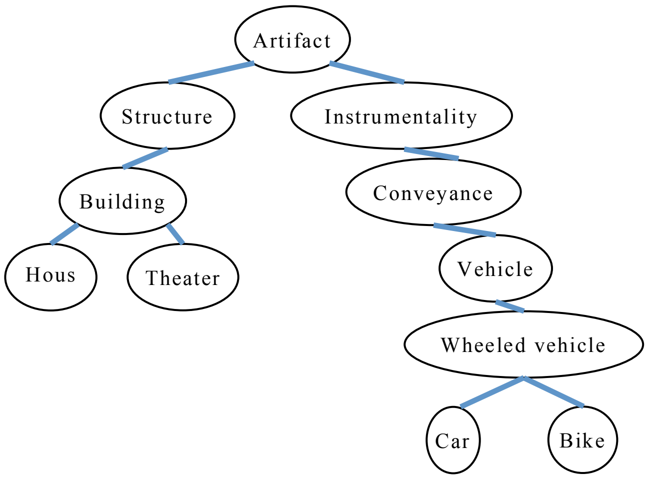

Fig 2. Apart of semantic hierarchical tree

3.3 Gather Sample Images

We use these noun concepts collected from Flickr as the queries and search the WordNet ID by using WordNet API. And then we use the WordNet ID to gather the sample image URL list from ImageNet. Then we retrieve the sample images from the sample image URL list. For each concept, we retrieve 2,000 images for maximum as sample images. And we delete the sample image whose image size is under 10Kbytes. We also skip the noun concepts if the number of their sample images is less than 50. Finally, the total number of collected noun concepts with sample images is 4,751. In this research we will only pick 100 concepts from 4,751 noun concepts for experiment. The pre-selected concepts are shown in Table 1.

Table 1. 100 pre-selected concepts

airplane	Ant	ape	apple	ball	beach	bike	bird
butterfly	Car	cat	cow	crane	dog	dolphin	door
ducks	Eagle	elephant	enve-lopes	eques-trian	fabric	fish	foot-ball
fruit	garden	glow-worm	golf	gorilla	hall	heron	hook
horse	ice cream	ice tea	iceberg	insect	jacket	jasmine	jeans
jellyfish	Key	keyboard	kitchen	knife	lamps	leopard	lion
lobby	mango	maple	motor-cycle	mush-room	nails	nar-cissus	narthex
news-paper	onion	orange	orchid	owls	palace	pencil	penguin
people	pine-apple	quads	quahog	quesa-dillas	rabbit	rack	res-taurant
rocket	sea	shark	sheep	sky	swallow	temple	tiger
tree	turtle	umbrella	unicycle	utensil	vessel	video-tapes	violin
volleyball	whale	window	wolf	worms	yam	yard	yellow jacket
zebra	ziggurat	zinnia	zither				

3.4 Propose Negative Sample Images Selection Methods

In this research, we propose six negative sample image selection methods and one baseline method for experiment. We calculate the distance between each concept by employing the semantic hierarchical tree. The distance relationship is shown as Fig 3. For example, the distance between concept "car" and concept "bike" is 2, and the distance between concept "car" and concept "conveyance" is 3.

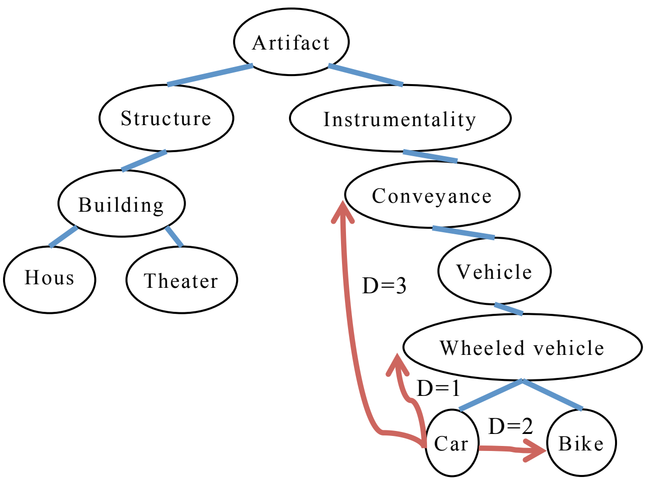

Fig 3. Distance relationship between each concept

According to the distance relationship between concepts, we propose six negative sample image selection methods.

3.4.1 Random Sampling (Baseline)

Most researchers collect the negative sample images by adopting random sampling method. They randomly gather the negative sample images from internet or datasets, and then employ the SVM classifiers training. In this research, we adopt random sampling method as the baseline method for experiment. We totally collect 3,511,137 sample images from ImageNet, and randomly choose negative sample images from these 3,511,137 sample images for SVM classifiers training.

3.4.2 Proposed Method 1

We choose fewer negative sample images for the concepts, which are near to the positive concept. And we choose more negative sample images for the concepts, which are far from the positive concept. The relationship between numbers of negative sample images and distance is shown in Fig 4.

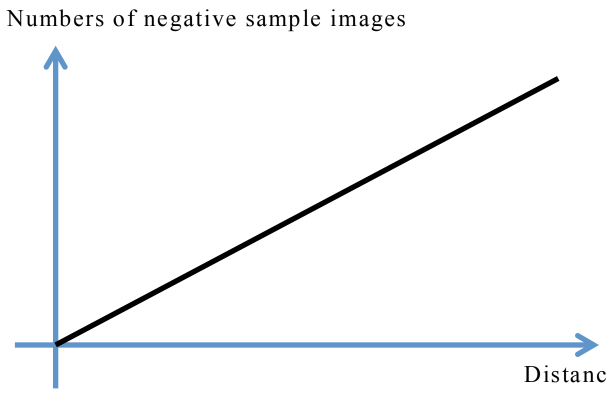

Fig 4. Graph of relationship between distance and numbers of negative sample images for proposed method 1

In this method, the total numbers of negative sample images will be almost equal to the numbers of positive sample images.

3.4.3 Proposed Method 2

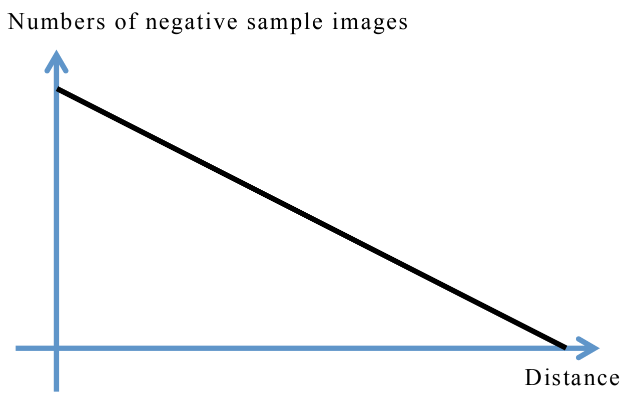

Fig 5. Graph of relationship between distance and numbers

of negative sample images for proposed method 2

We choose more negative sample images for the concepts, which are near to the positive concept. And we choose fewer negative sample images for the concepts, which are far from the positive concept. The relationship between numbers of negative sample images and distance is shown in Fig 5.

In this method, the total numbers of negative sample images will be almost equal to the numbers of positive sample images.

3.4.4 Proposed Method 3

We choose about 10 negative concepts, which are near to the positive concept. And we also choose about 10 negative concepts, which are far from the positive concept. The relationship between numbers of negative sample images and distance is shown in Fig 6.

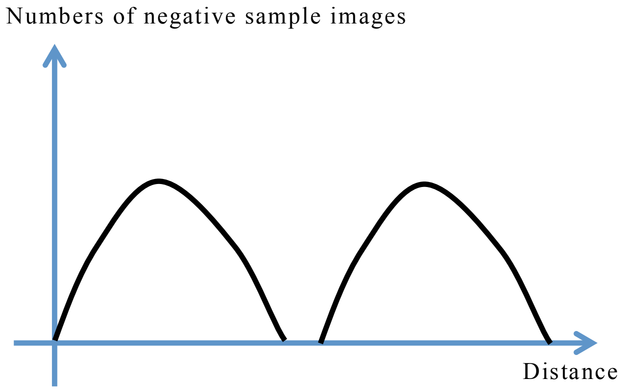

Fig 6. Graph of relationship between distance and numbers of negative sample images for proposed method 3

In this method, the total numbers of negative sample images will be almost equal to the numbers of positive sample images.

3.4.5 Proposed Method 4

We choose negative sample images which like normal distribution. We choose fewer negative sample images for the concepts both on near and far distance to the positive concept. And we choose more negative sample images for the middle distance to the positive concept. The relationship between numbers of negative sample images and distance is shown in Fig 7.

In this method, the total numbers of negative sample images will be almost equal to the numbers of positive sample images.

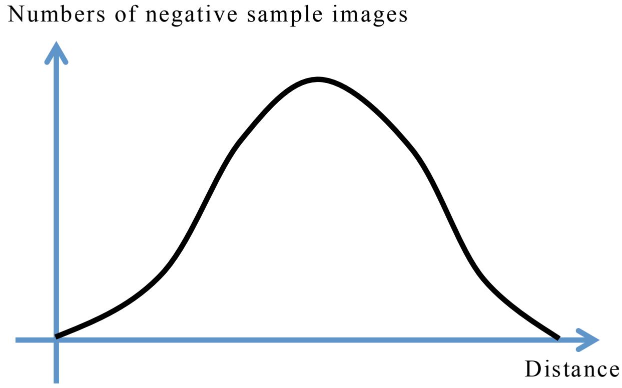

Fig 7. Graph of relationship between distance and numbers of negative sample images for proposed method 4

3.4.6 Proposed Method 5

We choose much more negative sample images for the concepts, which are near to the positive concept. And we choose very fewer negative sample images for the concepts, which are far from the positive concept. The relationship between numbers of negative sample images and distance is shown in Fig 8.

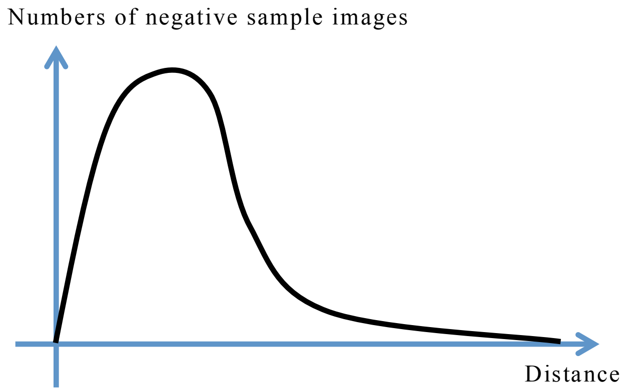

Fig 8. Graph of relationship between distance and numbers of negative sample images for proposed method 5

In this method, the total numbers of negative sample images will be almost equal to the numbers of positive sample images.

3.4.7 Proposed Method 6

For each distance, we choose uniform negative concepts, and then prepare the negative sample images. In the experiment we choose 2 negative concepts. The relationship between numbers of negative sample images and distance is shown in Fig 9.

In this method, the total numbers of negative sample images will be almost equal to the numbers of positive sample images.

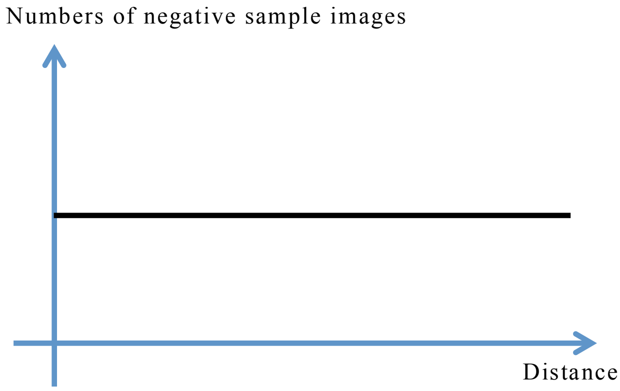

Fig 9. Graph of relationship between distance and numbers of negative sample images for proposed method 6

3.5 SVM Classifier Training and Predict

We extract low-level features from the entire sample images, and perform the SVM classifier training for each pre-selected concept. Our approach preform three feature extraction methods, Color Moments (CM), Local Binary Pattern (LBP), and Edge Orientation Histogram (EOH) on each positive and negative sample images. On CM feature extraction method, we use 6x6 grids and YCbCr as the parameters and the total extracted dimensions are 324. On LBP feature extraction method, we use 4x4 grids and 59 quantization as the parameters and the total extracted dimensions are 944. On EOH feature extraction method, we use 6x6 grids and 72 quantization as the parameters and the total extracted dimensions are 1,168.[3][4] After finishing the feature extraction of sample images, we implement SVM classifier training for each pre-selected concept. We use C-SVC as the SVM type and use Chi-Square as the kernel type for SVM classifier training.

3.5 Compare Experiment Results

We prepare 100 test images and predict with 100 pre-selected SVM classifiers. The main concept of 100 test images is the same as 100 pre-selected concepts. For each test image we also extract three features CM, LBP, and EOH. After predicting with SVM classifiers, we have three SVM predict results. Then we calculate the average of these three SVM predict results for ranking. For example, a test image airplane.jpg will match with 100 pre-selected classifiers. For each classifier will return three predict results: CM, LBP, and EOH. We average these three results as fusion score. Each fusion score belongs to one pre-selected concept. Then, there are 100 fusion scores for the test image airplane.jpg. Finally, we rank the 100 fusion scores (100 pre-selected concepts) and choose top-3 suggested concepts of test image airplane.jpg for accuracy comparison.

We consider three conditions "Top1", "Top2", and "Top3" for each proposed method. "Top1" means the suggested top-1 concept {airplane} matched with test image airplane.jpg. "Top2" means the suggested top-2 concepts {airplane, sky} content the main concept of test image airplane.jpg. "Top3" means the suggested top-3 concepts {eagle, airplane, sky} content the main concept of test image airplane.jpg.

Finally, we calculate the accuracy for each condition in each method. For example, in the "Top1" condition in method1, if only 30 suggested concepts are correct in the 100 test images (one test image with one suggested concept), the accuracy of method1 at "Top1" will be 0.30.

And in this research, we also discuss the mean reciprocal rank (MRR).

4. Results

We take positive concept "airplane" for example, the total positive sample images are 826. Table 2,3,4 shows the negative concepts and numbers of negative sample images that we have chosen in each distance for SVM classifier training for the proposed six negative sample images selection methods.

Table 2. Chosen negative concepts and numbers of negative sample images in method 1 and method 2

	Method1	Medhod2
D1	jet,3	seaplane,23 bomber,23 propeller_plane,23
D2	floatplane,6	glider,22 wide_body,22 airbus,22
D3	craft,9	craft,32 lighter,32
D4	hovercraft,12	vehicle,21 vessel,21 hovercraft,21
D5	sailing,15	conveyance,29 space_shuttle,29
D6	merchant_marine,18	bike,28 windjammer,28
D7	container,21	car_carrier,26 car_ferry,26
D8	empty,23	shoe_tree,25 telephone,25

D9	caddy,26	cue,23 holder,23
D10	lock,29	chimes,22 batting_cages,22
D11	suspender,16 red_carpets,16	airframe,21 handsaw,21
D12	cockpit,18 nightclub,18	winery,19 scimitar,19
D13	hatchlings,19 coif,19	ravine,18 washer,18
D14	winger,21 sedum,21	girl,16 rose_hips,16
D15	scallops,22 bus_boys,22	patriarch,29
D16	guan,23 freezer,23	leek,26
D17	popcorn,25 rana_temporaria,25	mimosa,23
D18	rye,26 hippos,26	candy_corn,21
D19	panthers,28 eel,28	american_copper,18
D20	haddock,29 sunfish,29	foie_gras,15
D21	pallone,21 bedlington_terriers,21 oryx,21	scad,12
D22	whippet,22 merino,22 bushbuck,22	cimarron,9
D23	rugby,22 angus,22 cavalla,22	soccer,6
D24	softball,23 professional_baseball,23 perfect_game,23	professional_baseball,3

Table 3. Chosen negative concepts and numbers of negative sample images in method 3 and method 4

	Method3	Medhod4
D1	bomber,2	biplane,2
D2	glider,17	jetliner,2
D3	craft,29	craft,1
D4	balloon,19 airship,19	airship,2
D5	bareboat,21 space_shuttle,21	lem,4
D6	ski,22 shipwreck,22	ski,10
D7	weaponry,21 rowboats,21	canoe,11
D8	magnets,19 hydroplane,19	rod,30
D9	canvas,29	mound,24 hangar,24 launcher,24
D10	loggia,17	flange,23 cap,23 loadstone,23 filament,23
D11	gallery,2	sports_arena,29 stoppers,29 apron,29
D12	anomaly,17	fancy_dress,22 jodhpur,22 light,22
D13	therapists,29	arthropod,23 beanie,23
D14	mammal,19 aquatic,19	coral_reefs,18 ref,18
D15	fruit,21 octopod,21	fizz,20 principal,20
D16	malt,22 tarantula,22	apple,23 litchi,23 crab,23
D17	mustard,21 flamingo,21	pudding,26 edam,26 weaver,26
D18	wild,19 snapping,19	champagne,24 star_anise,24 bluebirds,24
D19	martes_martes,29	marine_iguana,24 pieris_rapae,24
D20	flathead,17	saddle,17 siamang,17
D21	dive,2	pointer,30
D22	gordon_setter,2	clumber,12
D23	angus,2	rugby,3
D24	touch_football,2	professional_baseball,1

Table 4. Chosen negative concepts and numbers of negative sample images in method 5 and method 6

	Method5	Medhod6
D1	bomber,9	bomber,18 monoplane,18
D2	glider,17 floatplane,17	floatplane,18 airbus,18
D3	craft,42 lighter,42	craft,18 lighter,18
D4	vessel,30 spacecraft,30 hovercraft,30 airship,30	balloon,18 airship,18
D5	rocket,26 ship,26 yacht,26	boat,18 spaceship,18

	shrimpers,26 space_shuttle,26	
D6	trailer,30 fire_boat,30 trawler,30 windjammer,30	public,18 hospital_ship,18
D7	mountain_bike,27 barrow,27 car_ferry,27	tracked,18 dredger,18
D8	bird_feeder,25 flowerpot,25	cosmetic,18 cars,18
D9	bomb,19 drawstring,19	fuse,18 ligament,18
D10	mask,16 elastic_band,16	digital,18 machine,18
D11	minivan,28	storeroom,18 military_uniform,18
D12	golf_tee,24	pocket_watch,18 touchscreen,18
D13	American,19	grub,18 argyle,18
D14	cannas,15	trilobites,18 swamis,18
D15	tiger_cub,13	dainty,18 road_runner,18
D16	curd,9	cuttlefish,18 gentian,18
D17	crab_apples,6	midges,18 new_england_aster,18
D18	cranberry_sauce,5	unai,18 long_legs,18
D19	chelonia_mydas,5	boar,18 sanderling,18
D20	fishing,4	bovine,18 ruddy_turnstone,18
D21	brahma,4	mountain,18 burro,18
D22	shire_horse,3	miniature_schnauzer,18 steer,18
D23	american_football,3	english_springer_spaniel,18 angus,18
D24	touch_football,2	touch_football,18 perfect_game,18

Fig 10 shows the accuracy results for “Top1”, “Top2”, and “Top3” between baseline method (random sampling) and proposed six negative sample images selection methods. The experiment results show that the accuracy of baseline method in “Top1” reaches 0.48, “Top2” reaches 0.64, and “Top3” reaches 0.73. And the experiment results also show that the proposed method 6 outperforms other methods. Method 6 is the method we choose uniform negative concepts (two negative concepts) from each distance and the accuracy in “Top1” reaches 0.53, “Top2” reaches 0.65, and “Top3” reaches 0.71.

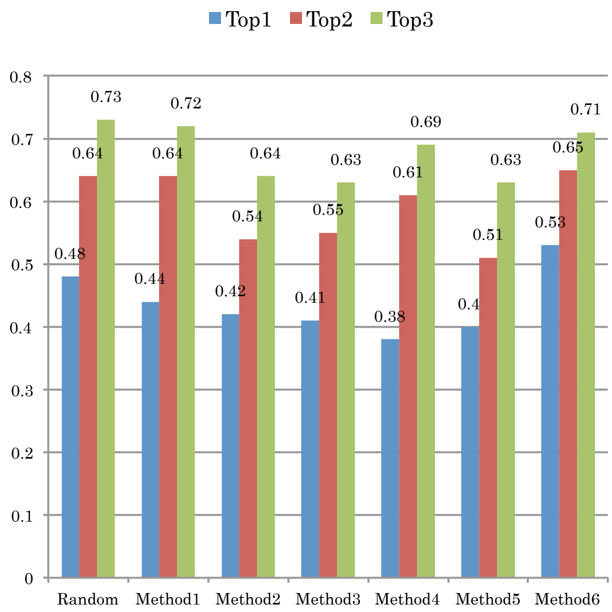

Fig 10. Accuracy results for six proposed methods

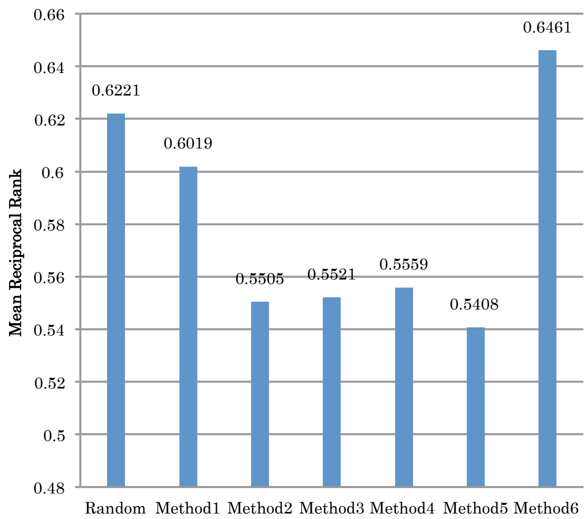

Fig 11. Mean reciprocal rank for six proposed methods

Fig 11 shows the mean reciprocal rank between baseline method (random sampling) and proposed six negative sample images selection methods. The proposed method 6 outperforms other proposed methods and also outperforms baseline method.

5. Conclusions

In this research, we discuss the accuracy of different negative sample images selection methods in one-class SVM classifier training and prediction. We adopted

WordNet for building semantic hierarchical tree and employ this semantic hierarchical tree to calculate the distance between each concept. We adopted the distance information for proposing six negative sample image selection methods for experiment. The experiment results show that different methods indeed generate different accuracy results. The method we proposed choose uniform negative concepts from each distance outperformed the baseline method and the other methods we have proposed.

Acknowledgements

This research was supported by grant from JST "Multimedia Web Analysis Framework towards Development of Social Analysis Software," Grant-in-Aid for Scientific Research (No. 18049068), and Ambient SoC Global COE Program of Waseda University" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Reference

[1] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation", International Journal of Computer Vision 2008, pp. 157-173

[2] Claudio Cusano, Gianluigi Ciocca, Raimondo Schettini, "Image annotation using SVM", SPIE 2003, pp. 330-338

[3] Duy-Dinh Le, Shin'ichi Satoh, and Tomoko Matsui, "High Level Feature Extraction", TRECVID 2007

[4] Duy-Dinh Le and Shin'ichi Satoh, "A Comprehensive Study of Feature Representations for Semantic Concept Detection", TRECVID 2011

[5] Fellbaum, C, "WordNet: An Electronic Lexical Databas," Bradford Books.

[6] Gerard Sychay, Edward Chang, and Kingshy Goh, "Effective Image Annotation via Active Learning", ICME 2002

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," CVPR 2009, pp. 248-255

[8] Marcus Rohrbach, Michael Stark, Gyorgy Szarvas, Iryna Gurevych, and Bernt Schiele, "What Helps Where-And Why? Semantic Relatedness for Knowledge Transfer", CVPR 2010, pp. 910-917

[9] Meng Wang and Xian-Sheng Hua, "Active Learning in Multimedia Annotation and Retrieval: A Survey", ACM TIST 2011

[10] Simon Tong, and Edward Chang, "Support Vector Machine Active Learning for Image Retrieval", ACM MM 2001

[11] Simon Tong and Daphne Koller, "Support Vector Machine Active Learning with Applications to Text Classification", Journal of Machine Learning Research 2001, pp. 45-66

[12] Vincent. S. Tseng, Ja-Hwung Su, Bo-Wen Wang, and Yu-Ming Lin, "Web Image Annotation by Fusing Visual Features and Textual Information", SAC 2007

[13] Wei Li, and Maosong Sun, "Automatic Image Annotation Based on WordNet and Hierarchical Ensembles", CICLing 2006, pp. 417-428

[14] Yoshimasa Tsuruoka and Jun'ichi Tsujii, "Bidirectional Inference with the Easiest-First Strategy for Tagging Sequence Data", Proceedings of HLT/EMNLP 2005, pp. 467-474