

Classification of Videos using Semi-automated Tagging Techniques

M.Tech Project Phase I Report

Submitted in partial fulfillment of requirements for the degree of

Master of Technology

by

Shraddha Bhattad
Roll No : 13305R005

under the guidance of

Prof. Ganesh Ramakrishnan



Department of Computer Science and Engineering
Indian Institute of Technology, Bombay

Oct, 2015

Contents

1	Motivation	1
2	Automatically extracting a domain-specific taxonomy	3
2.1	Introduction	3
2.2	Wikipedia	3
2.3	ConceptNet	4
2.4	WordNet	6
2.4.1	WordNet as potential candidate for Knowledge Graph	7
3	Video tagging as a game	10
3.1	Reducing cognitive load on the player:	11
3.2	Enabling the player to focus on preferred videos:	12
3.3	Image Trailer of a Video	12
3.4	Using a fair scoring function :	15
4	Meta-Learner	17
5	Conclusion and Future Work	20

Abstract

Organizing online videos in a domain-specific taxonomy is an important and challenging problem. Defining such a taxonomy in advance (without even looking at the videos) is not prudent as it may suffer from the problem of over specification and/or under specification. In this project, we propose to create a data-driven, dynamically evolving taxonomy and refine it further with the help of crowdsourcing.

Specifically, we propose to extract important concepts from a domain-specific corpus. We then design a game where the users will be shown videos and asked to assign tags to them such that each tag corresponds to a concept extracted earlier. These tags and the correlations between them will eventually become nodes and relations in our taxonomy. To keep the user engaged, he/she will be given a score.

Cold start (when no user has played for that video): Scoring only based on the agreement between the tags assigned by her and the tags assigned by an automatic tagger which relies on features based on video textual meta-data and image meta-data. Otherwise more weightage will be given for ensuring consensus between fellow players who have already played for that video. A simple scoring mechanism is designed to cater the above need as we want to evolve and not rely on the automatic dumb tagger forever.

Our method of crowd-driven semi-automatic taxonomy creation and video tagging requires a combination of ideas from several research fields, such as (i) automatic extraction of knowledge graphs (ii) crowdsourcing (iii) gamification of annotation tasks and (iv) image, video and audio processing. In this report, we list down the work done till now and future work which will supplement it.

Acknowledgements

My sincere thanks to Prof. Ganesh Ramakrishnan for providing valuable feedback time and again which served to influence and improve this work.

I would also like to thank the entire team: Dr. Simoni S.Shah, Post-Doctoral Fellow IIT Bombay, and my fellow team mates Depen Morwani, Saketh Vadlamudi, Aditya kumar Akash, Deepak Dilipkumar for their extended discussions and brainstorming on various issues.

Special thanks to Ashish Kulkarni, Research Scholar at IIT Bombay and Mitesh Khapra, Researcher India Research Laboratory, Bangalore, India for providing us valuable insights.

I would also like to thank Ankit Vani & Pooja Ahuja, Interns at IIT Bombay for designing the initial game framework.

Chapter 1

Motivation

Video is a powerful avenue of reaching and influencing the lives of people. It is important for farmers and rural people to stay up to date on the latest techniques, possibilities and success stories from others all around the world, and video can be one of the most convenient sources of this information for them.

Our major goal behind using videos as the medium despite well developed text knowledge is that video is very intuitive. Text may or may not be intelligible to masses but video for sure is. Also, smartphones, which were an oddity in rural areas, have now become commonplace, so such a video library can be easily accessed and utilized by farmer groups using specialized mobile applications. If a farmer is having a problem, he/she can be recommended a video relevant to the problem being faced.

The video repositories currently in place, such as YouTube, provide a general purpose solution for the task of video categorization and searching. This means that a closed domain like that of agriculture falls into a small cluster, or is at best a union of a handful of clusters of all of their videos.

Every video on YouTube can be assigned exactly one category, which is chosen by the uploader from a fixed list of high level domains, such as Education, Comedy, Music, Sports, etc. Such a high level categorization is not adequate when dealing with a closed user group with more focused needs.

Along with a category, uploaders on YouTube also specify a list of tags for their videos. Although they are usually more specific to the video, we observed that such tags are often too fine grained, or too generic. fine-grain tags mostly do not reoccur. Sometimes, uploaders also tend to add many unrelated tags for their videos in ignorance or to gain 'views'.

Hence there is a need to have a good tagging system for the video repository, which in itself will be a dynamic one with several videos being uploaded each day. This will further enhance the way in which videos are classified, their search, and the entire learning experience of the users.

Motivation from **wikiHow**:

wikiHow [12] is an online wiki-style community consisting of an extensive database of how-to guides. Founded in 2005, now it has more than over 190,000 how-to articles.

Most how-to articles follow a similar format with steps, tips, and warnings, and are complemented with images to help a reader learn how to complete a task. wikiHow uses the wiki method of continuous improvement, allowing editors to add, delete, or otherwise modify content. Once an article is created, community members collaborate on it to improve its quality. wikiHow is designed for global settings, we want to preserve and evolve **Indian traditional practices** in farming and other expertise which are getting lost day by day. Our hope is to create a how-to-guide in *video format*. These would be very light weight videos formed by images and audio clubbed together. Our other team of interns working with Prof. Ganesh has developed an android application called Lok-Vidya[8] which creates such image slideshow stitched with audio and renders it in video format.

Apart from creating the videos, we also want to collect existing videos which may be of great use. We want the crowd-moderation for both of them.

We want to gamify this video tagging task so that user do not feel that they are working/volunteering for us, rather they are enjoying the game and learning while watching the videos and producing tags as a by-product.

Such a gaming system has two stake-holders (i) players/viewers and (ii) tag collectors (i.e. the person or organization interested in collecting these tags). The gaming system should try to serve the interests of both these stake holders. Specifically, from the point of view of the viewer it should ensure that he/she has maximum fun and learning while playing the game. Also it should be possible for him/her to quickly decide whether a particular video is of interest or not. From the point of view of the collector, the system should have good incentives to ensure that the tags are of high accuracy and high specificity (for example, where applicable, drip irrigation is preferred over irrigation).

For the purpose of this work, we shall assume that the videos of interest belong to a particular domain of interest., which we will quantify using our seed-set.

Specifically, we shall assume that the broad subject of the videos constitutes best practices of people in rural areas such as farmers, artisans, etc. Given this context, we now define important sub-problems that need to be addressed to achieve the stated goals.

Chapter 2

Automatically extracting a domain-specific taxonomy

2.1 Introduction

For videos to be organized in domain specific taxonomy, We first need a good domain specific category catalog. Manually identifying such a subset of concepts and the relations between them would be expensive, tedious and error-prone.

Hence we propose mining it from existing lexical databases like Wordnet, [17] encyclopedia like Wikipedia, or a hybrid Knowledge base like ConceptNet[3]. We have to ensure that regional language support is a must, for the game and tags should not be restricted to english. We start with a seed-set agreed upon by with the expert involvement. In our case from persons with farming and rural activities expertise.

2.2 Wikipedia

We have conducted an experiment on using only 'textual metadata' to classify Videos and give it potential candidate categories using AMN classifier which evolves based on user feedback and has Wikipedia as knowledge base. The results are analyzed in Section 4. With this experiment we became quite hopeful in what Wikipedia has to offer in the game settings. This paper "Mining Domain-Specific Thesauri from Wikipedia: A case study" [15] has done an analysis on the same. Let us follow their findings on how good is to exploit collaborative folksonomies like Wikipedia.

It has compared Wikipedia to Agrovoc [1], a manually-created professional thesaurus in the domain of agriculture, as the gold standard.

Wikipedia covers 69% of Agrovocs hierarchical relations, but only 25% appeared in the category structure, remaining 44% were found in redirects and hyperlinks between articles.

Wikipedia covers synonymy particularly well with its redirect structure. But hyperlinks tell nothing about the heirarchy.

On their study conducted on agricultural documents taken from the FAOs document repository., they compared Wikipedia with Agrovoc on 780 full text (not abstracts) documents. These were professionally indexed with at least three Agrovoc terms. Figure 2.1 tells about the coverage statistics of both. They also establish that the terms which Wikipedia didn't covered was very specific like *Margossa* etc.

Also, Wikipedia can help in extending the existing thesauri., as shown in Figure 2.2. We see later but what the results of ConceptNet(which uses Wikipedia and other knowledge bases) has to say about these. It is a good idea, but again we want to avoid high topic drift incurred after such extensions.

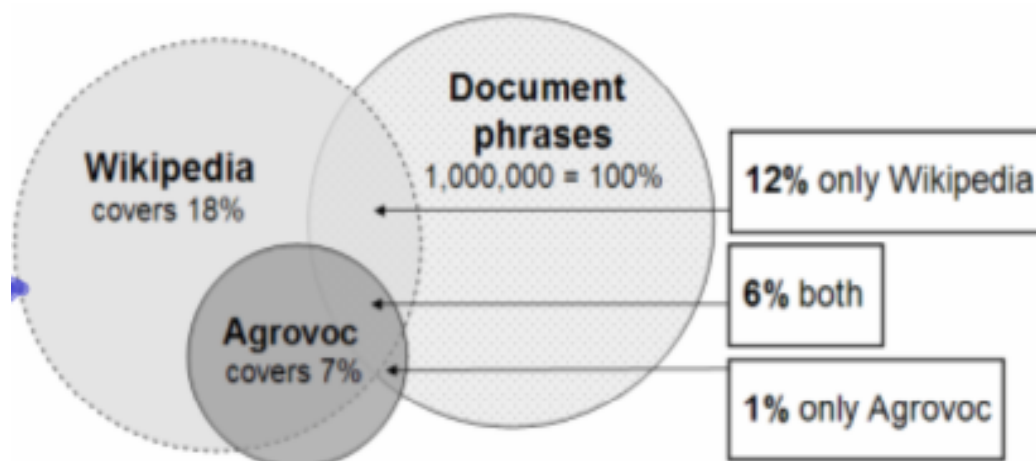


Figure 2.1: Wikipedia Vs Agrovoc on FAO documents

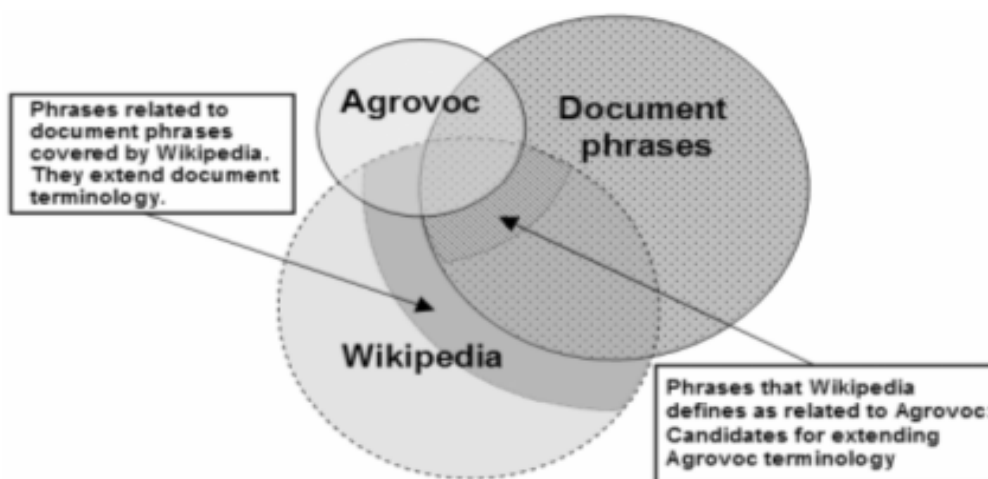


Figure 2.2: Extending Agrovoc with the help of Wikipedia

2.3 ConceptNet

ConceptNet[3] is a multilingual knowledge base, representing words and phrases that people use and the common-sense relationships between them. The knowledge in ConceptNet is collected from a variety of resources, including crowd-sourced resources (such as Wiktionary [13] and Open Mind Common Sense), games with a purpose (such as Verbosity and nadya.jp), and expert-created resources (such as WordNet and JMDict). It contains everyday basic knowledge. Several kind of relationships offered by ConceptNet are:

MotivatedByGoal, UsedFor, RelatedTo, IsA, PartOf, MemberOf, HasA, HasContext, MadeOf and many more.

Currently we just looked at associated words to a given word depicted by Figure 2.3.

We see a very good coverage in ConceptNet especially for western-english noun words. Seeing the results for threshing, we see it has given very high scores to womp, thrash, flagellation, all of them quite out of context, but at the same time, we got flail[a threshing tool consisting of a wooden staff with a short heavy stick swinging from it.] which seems to enrich the meaning

Biotechnology		Composting		Mowing		Threshing	
Similar	Score	Similar	Score	Similar	Score	Similar	Score
bioscience	0.8722	center of universe	0.6517	uncropped	0.8484	whomp	0.8681
natural history	0.8183	catch on fire	0.5857	reap	0.8053	flail	0.8666
agrobiology	0.8122	reaction mixture	0.5852	snip	0.8027	thrash	0.8640
system Science	0.7999	pick flower	0.5793	corncutter	0.7877	flagellation	0.8612
विज्ञान	0.5646	कुछ	0.5096	अश्रु	0.2121	मारना	0.5945
जीवित	0.5096	गायब	0.2815	आँसू	0.2117	दण्ड	0.3457
औषध	0.3505	जंगल	0.2797	तीखा	0.1750	अलग	0.3224
रहना	0.3356	परमाणु	0.2717	क्षेत्र	0.1593	जुदाई	0.3216

Figure 2.3: ConceptNet results in english and hindi depicting the most closely associated words the given word.

further. So its a trade-off which we need to tweak according to our requirement. In order to use ConceptNet, we have to keep a check on high topic drift. We observed that verbs were not present in direct form. We converted verbs to its root form like compost for composting, and then curated the results. Still, the confidence and quality coverage for verbs is not good. We also saw the Hindi results for the same data. In general, the confidence for hindi is low as compared to its english counterpart. This may be due to lack of backbone knowledge bases in Hindi. Hence, its safer to reject building Hindi Knowledge base using ConceptNet. We will later look at prioritizing relationships, like giving higher priority to hierarchical relationships and the likes suiting to our needs. Similar to what the paper *Sentiment Aggregation using ConceptNet Ontology* [19] does. This may give a better confidence based on the relationship. for eg., a thresher *IsA* tool *UsedFor* cutting. We have to think on the line of relating roles to the concepts. This may prevent the topic drift to a great extent.

2.4 WordNet

Wordnet[17] is a lexical database which arranges words into sets of synonyms called synsets. It further captures various relationships between synsets (for example, meronymy/holonymy, hypernymy/hyponymy, etc.). Hopefully, many domain-specific concepts that we are interested in will already be present in Wordnet along with the relations between. We can use this information to enrich our domain-specific taxonomy graph by constructing deterministic edges between nodes/concepts based on the Wordnet relations defined between them. It is quite likely that the coverage of this method would not be of the order of Wikipedia/ConceptNet but it is still worth considering it to get an initial seed set of relations between concepts. as it would be very less prone to topic drift.

'जुताई'
 Sense Count is 2
 Synset [0] NOUN - [जुताई , जोताई, कर्षण, संकर्षण, प्रकर्षण]
 HOLO FEATURE ACTIVITY : NOUN - [खेती, कृषि, किसानी, खेतीबाड़ी, खेती-बाड़ी, खेती बाड़ी, खेतीबारी, खेती-बारी, खेती बारी, कृषिकर्म, कृषि-कर्म, कृषि कर्म, कृषि कार्य, फसली कर्म, किसनई, काश्त, काश्तकारी, एग्रीकल्चर, किश्त, गृहस्थी]
 Synset [1] NOUN - [जुताई, जोताई]
 HYPERNYM : NOUN - [मजदूरी, मजूरी, मज़दूरी, उजरत]
 There were no results for compound words like वृक्षविद्या, पशुविद्या but there were results for खरीफ, रबी which are Indian terminology for seasonal crops.

Figure 2.4: Hindi-Wordnet[6] results for Tilling

As we analysed the results, WordNet proved to be very helpful and relevant. Generated by a dedicated set of linguists, the terms and terminologies have great regional language support, and the work is going on till date. We analysed English WordNet[17] originated at Princeton University and Hindi Wordnet [6] originated at IIT Bombay. We feel that verbs/processes are very well covered in WordNet and little topic drift is there. Our Plan is to have a expert curated seed-set which will mainly cover broad topics in Agriculture and farming practices, Animal husbandry, Disaster Preparedness, Gardening, Recycling, Home Decorating, Energy, Sustainability and other rural art forms like Sculpture, Ceramics, Metalwork, Folklore, Painting, Printing, Metal-engraving and many more. This seed-set will be designed keeping rural masses in mind, their lifestyles, practices, living, entertainment and their traditional knowledge. Our final aim is to develop something like WikiHow keeping the regional taste and tutoring through videos. For more examples please visit this presentation. [9]

2.4.1 WordNet as potential candidate for Knowledge Graph

2.4.1.1 Relations in WordNet used in Knowledge graph

1. *Synonymy*: most important relation for WordNet is similarity of meaning,
2. *Hyponym/Hypernym*: subordination/superordination: A hyponym inherits all the features of the more generic concept and adds at least one feature that distinguishes it. e.g. maple/tree.
3. *Component Meronym* is a component part of *Component Holonym* e.g. branch/tree
4. *Substance Meronym* is a stuff that *Substance Holonym* is made from. e.g. aluminum/airplane
5. *Member Meronym* is a member of *Member Holonym*: e.g. tree/forest .
6. There are others like *Antonym* etc which we havent used.

2.4.1.2 WordNet Statistics and building preliminary Knowledge Graph

Table 2.1: WordNet 3.0 database statistics

POS	Monosemous words& senses	Polysemous words	Polysemous senses
Noun	101863	15935	44449
Verb	6277	5252	18770
Adjective	16503	4976	14399
Adverb	3748	733	1832
Total	128391	26896	79450

A good thing here is no attempt has been made to cover proper noun! which was prominently the case with Wikipedia and other Knowledge bases. This way we can get rid of the specific person, place, thing information and focus on common practices. As we clearly see, we need to take care of senses of words as number of polysemous words, specifically verbs are quite high. We explain how to tackle this next.

We built a preliminary Knowledge Graph using English WordNet. We will continue the same with Indo WordNet(for Indian languages.)

1. We procured a seed-set consisting of broad terms and processes in Agriculture, and other art and expertise of rural areas.
2. For now, we plan to generate a seed on our own and confirm it with them. We have picked relevant concepts from Farmer's Portal[4], India & Wikipedia.
3. For every entries in the seed set, We have manually chosen the desired synset.,so that there is the same context following and no sense drift.
4. We then automatically extract all the relations defined earlier recursively and keep adding them to KnG., till new nodes keep coming in.
5. We have stored the meanings, examples, synset, the edge type and have retained the original hierarchy.
6. Wordnet has no notion of scoring. We need to calculate scores dynamically based on graph position of a word relative to other word.
7. Right now, a trivial method using inverse of distance is used, with every kind of relation given same unit weightage. We will in future analyse and tune weightages based on kind of relations.

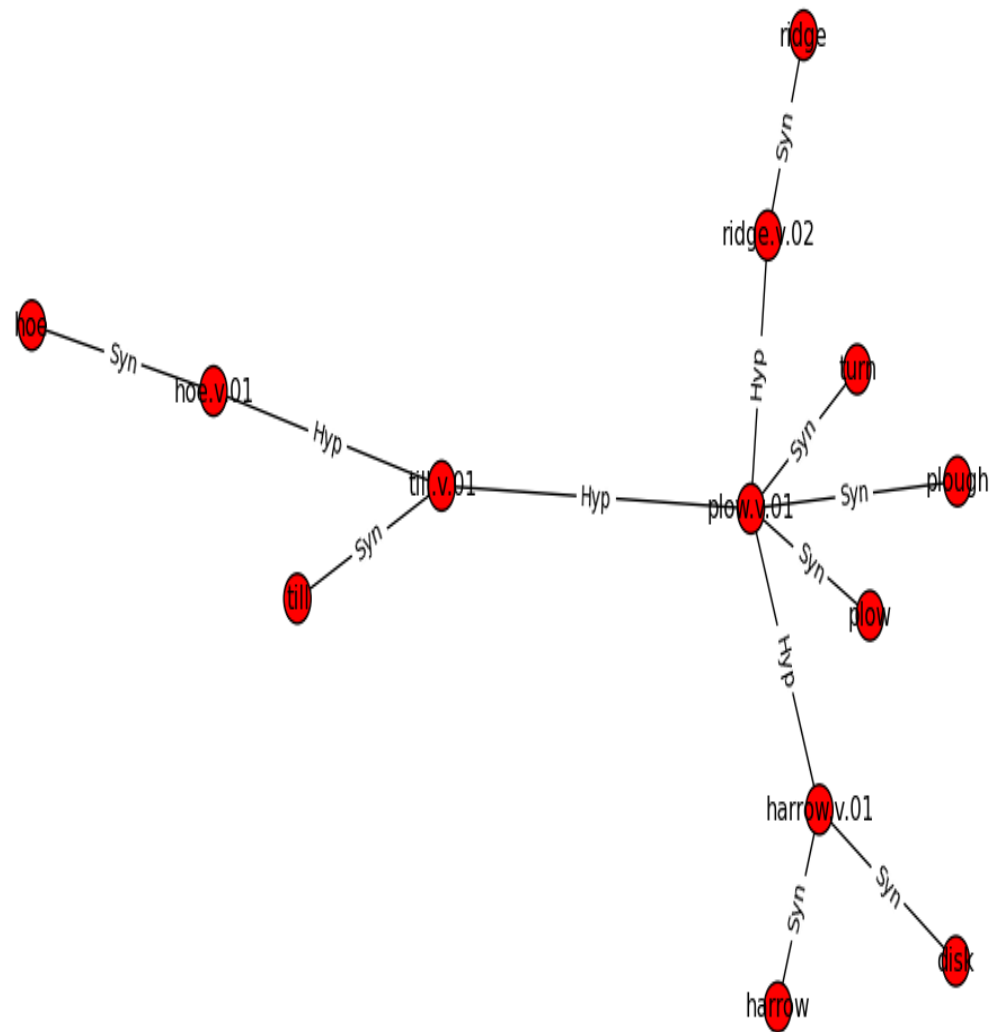


Figure 2.5: English WordNet results for Tilling(verb) generated using NLTK [16]

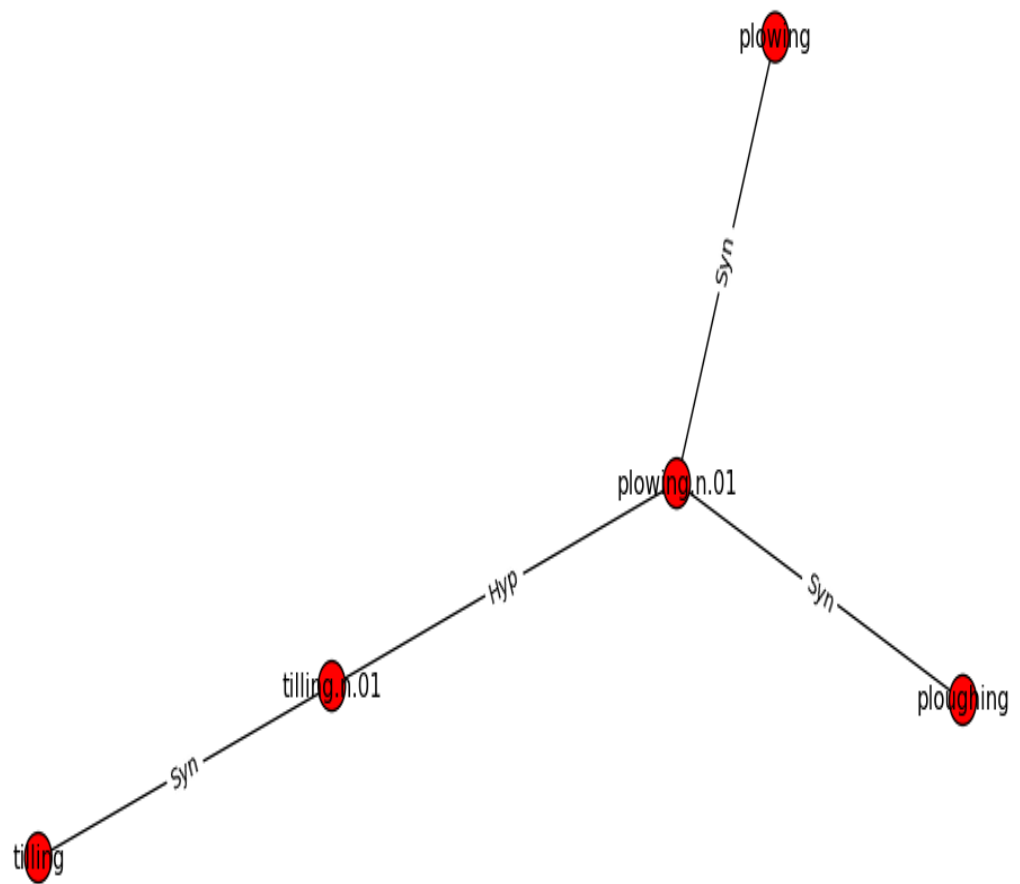


Figure 2.6: English Wordnet results for Tilling(noun)

Chapter 3

Video tagging as a game

Given a domain-specific taxonomy, as described in the previous section, we are interested in tagging all videos in our collection with appropriate concepts from the taxonomy. Instead of doing this using in-house annotators, we propose to design a game where multiple viewers/players can view a video and assign tags to them. The success of such a gaming system depends on keeping the player engaged which in turn depends on the following:

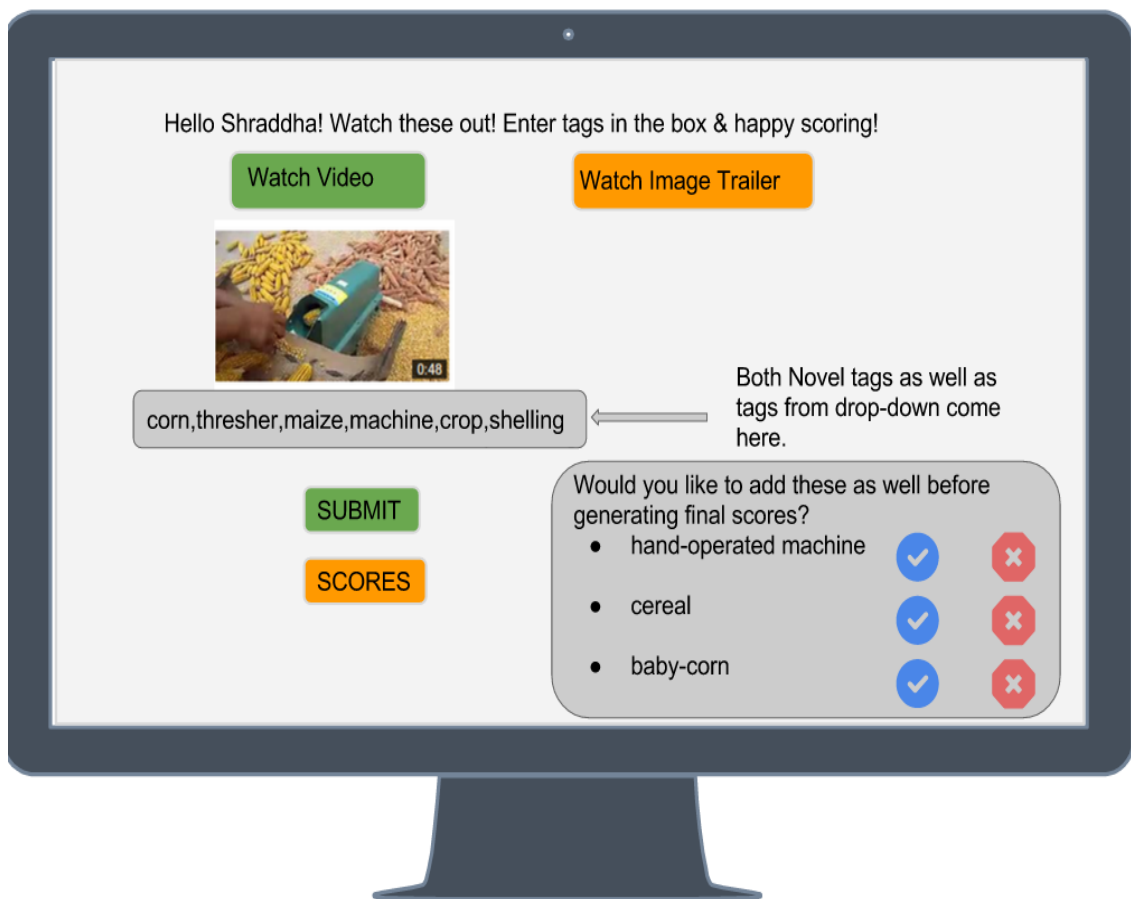


Figure 3.1: Basic Draft of Video tagging game

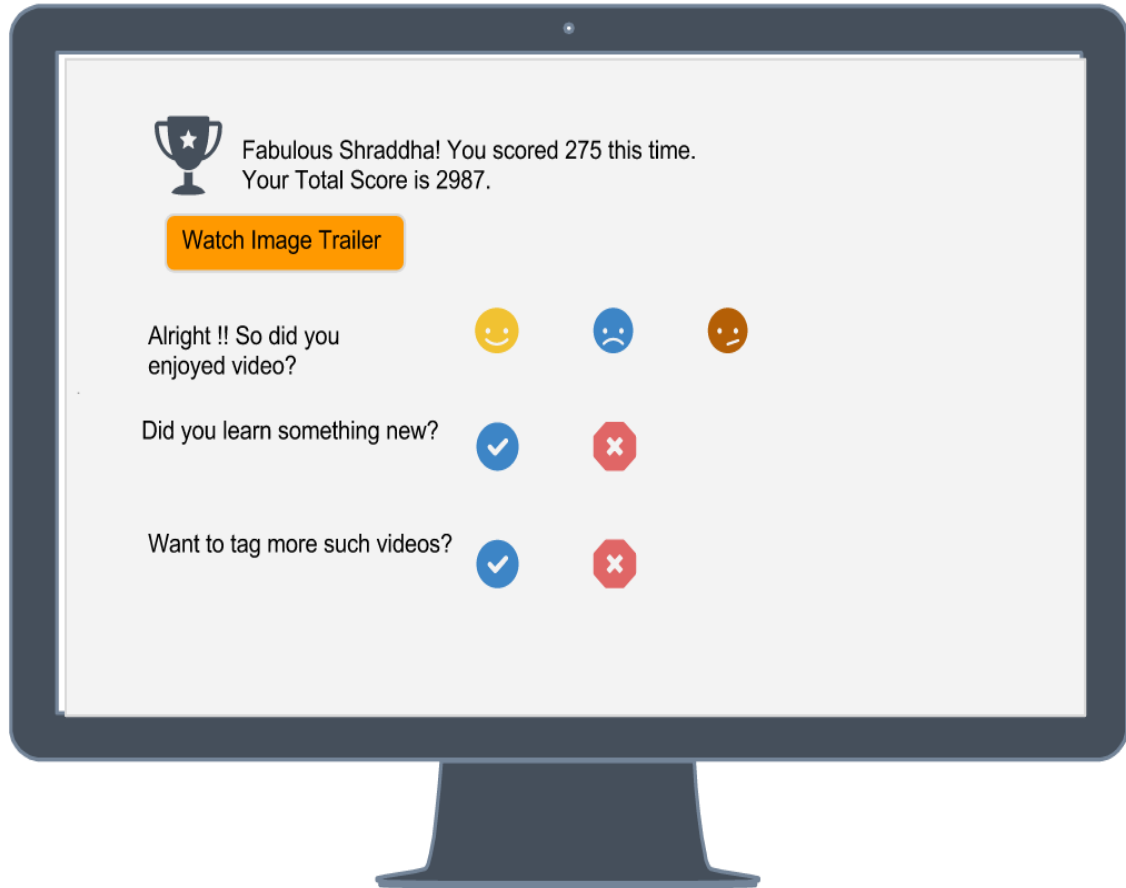


Figure 3.2: Additional feedback for the system to learn better

3.1 Reducing cognitive load on the player:

Ideally, we expect the user to simply watch the video for entertainment and/or education and the process of assigning tags should not hamper her viewing experience. To ensure this, firstly, the player can obviously not be expected to be aware of all the concepts in our taxonomy. He/She is expected to enter tags which seem natural to him/her and the system then display concepts which are closest to the tag suggested by the user. At the minimum, when the player starts typing a tag, (say, "gr") all concepts in the Knowledge Graph starting with this prefix (say, "green", "grass", "grape" etc.) are displayed in a drop-down and player can then select an appropriate concept from it. Secondly, if the tag entered by the player does not match any existing concept then also player can still enter tag marked as 'novel tag', which in turn help us to evolve. The 'scores' which makes a player hook to the game and explained in section 3.4

Above figure 3.1 and figure 3.2 shows a draft version of the game. We can think of this as a **User Story** which tells the game functioning overall.

Player watches the Video/Image Trailer, selects the tags from drop-down or gives his/her own tag. After clicking on submit, some system generated suggestions pops-up and asks players if he/she wants to include those as well before generating final scores.

Note that this is very less load for user, its just a yes/no thing, but of crucial information to us as these are the most uncertain tags for us which we want to gain confidence on. Designing scoring for these would be little tricky. We may instead give him some honor points for giving us his/her valuable feedback.

After clicking on Scores, we take a quick feedback on game, learning experience and the video to help us improve further.

3.2 Enabling the player to focus on preferred videos:

It is fair to assume that a player will feel more engaged if the system can assist him in quickly identifying relevant videos which match his interests. There are at least two ways of achieving this. We show video trailer consisting to budgeted Image representation of the video to player. Either player can tag by viewing images only or if he develops the 'feel' by the time he go through the image set, he can go ahead with watching full video.

Case Study:

We conducted a small study with 10 students of age group 19-25 to have user's opinion regarding Image trailer concept. We found that majority(8) of people preferred watching the image summary of the video to quickly decide their next step. 6 among them didn't prefer watching the video at all. When asked the reason, they said it can't be done quickly and parallel to some other work. Some also quoted that video buffering was heavy on internet. 4 people went ahead with watching the video.

This was a study done with people whom I knew personally. Amateur feedback may differ. Important thing to note here is we should try hooking up the user in first shot. We will rather prefer a user tagging 10 twenty minutes video rather that watching and tagging 1 twenty minute video.

We are happy with any option however as one of the following Tagging/Learning/Both is ensured in all the scenarios.

However this study is not complete as our subjects were only students, We will soon test this on non-academicians.

3.3 Image Trailer of a Video



Figure 3.3: Whats going on here? Give me some hint!

Use cases of Image summary in our game settings. A player can:

1. Tag the images itself without proceeding towards video
2. Tag the images and then proceed towards video tagging
3. Do not tag the images and proceed towards video tagging
4. Skip the entire tagging based on Image summary not appealing him/her.

Also, these images are a potential candidates for image tagger., as they are diverse and represent the entire video in a certain budget. Ours being 25 images.

Our Algorithm for obtaining diverse & representative set of budgeted images from a video:-

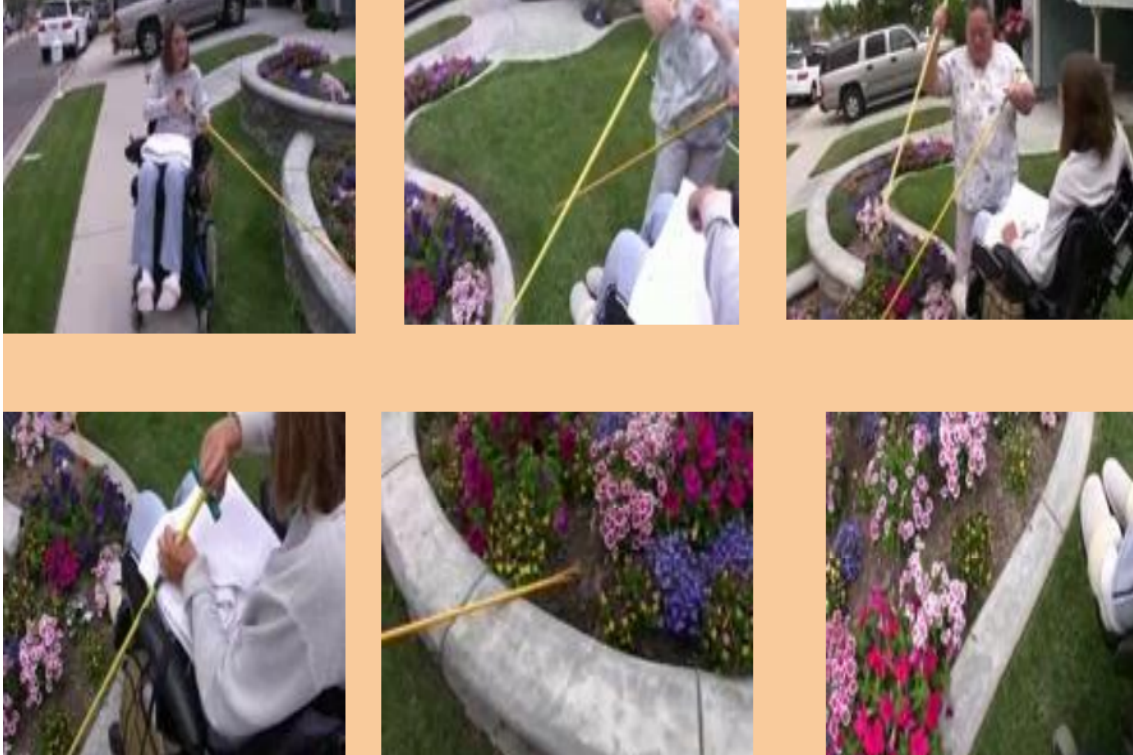


Figure 3.4: Can you think of some tags for these?

1. Get images frames after a fixed interval. This interval will be shorter for short videos, longer for long videos. This steps exploits the temporality of videos and the fact that in video successive frames are expected to be very similar.
2. Starting with first image, keep on adding subsequent images if they exceed certain threshold of pixel difference.
3. If the set obtained above is within our budget, return else compare every image to every other image and rank them in order of them being diverse from the remaining images. We will then extract top images under our budget as the final Image set.

Frames are extracted from the given video using ffmpeg [5] with $\text{fps} = 1/\sqrt{s}$, where s is the length of video in minutes.

Command:

`ffmpeg -i video-name -f image2 -s 160x120 -vf fps = 1/sqrt(s) ./Images/imagename-%02d.jpeg'`
 Here 160X120 is the image dimension. Keeping it low gives quick computation.

Pixel comparison method:

1. Convert image from rgb to gray. $[rgb2gray : 0.2989 * R + 0.5870 * G + 0.1140 * B]$
2. Threshold image to put all values greater than 125 to 255 to 1 and all values below 125 to 0 and create a binary image.
3. Boolean comparison of final binary images pixel by pixel.
4. If two images differed by more than a threshold(our case 28%) we kept it. Otherwise we rejected the other.

This comparison was quick and fast enough. On an average it took about 4-5 seconds to compare if the initial set of frames(from ffmpeg) was already present.

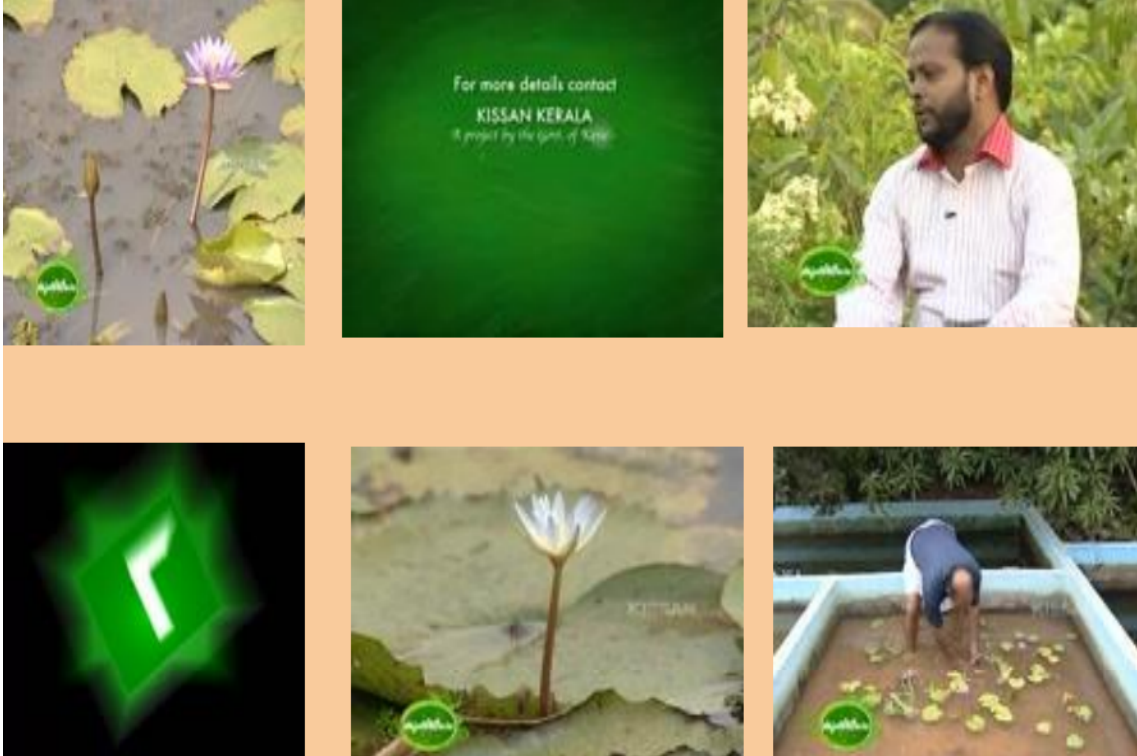


Figure 3.5: What about this?

I have explicitly hidden the title of images to check whether a person is able to guess the process. Spoiler alert! figure 3.3 is a reference image set(top 6) for the video titled 'Antique Corn-Sheller with hit and miss engine'. Similarly video for figure3.4 is titled 'Weeding in wheel-chair' and figure 3.5 is titled 'Cultivation of water lily and lotus on large scale'

These figures represent the top 6 images returned by our algorithm. However when our budget is larger say 12-15 we will have much better clarity about what the video is about. This seemed to have worked pretty well in our favour. Almost in every case we are able to tell what the main-story of video is just by few seconds glance. This can be of great aid to player to himself judge on video quality, his interest in watching/learning from the same or skipping it, and hence saving him a lot of time which he can utilize in tagging other videos. In order to gain high scores he may just go wildly tagging images itself without watching the video at all, which again works on our favor of collecting more and more tags. However we don't just want to gamify, we want our players to learn, so how to provoke him into watching video till the end will be a challenge. Automatic video summarization is an interesting area of research and we

have implemented an initial solution in form of Image summary which covers only one aspect of video.

Audio aspect needs to be worked on, especially in cases where video is audio dominated, like a professor teaching by talking or other speech/music oriented videos. Image summary may not be of much help in those scenarios.

Also for a process oriented video where the crucial process is covered in just a glimpse may or may not be captured by the Image summary.

In future we can think of building a recommender system which suggests videos based on the past history of the user. Such a system could take as input the tags assigned to all the videos viewed by the user in the past and then search for other videos which have similar tags. If videos having the exact same tags are not available then those having semantically similar tags can be suggested.

3.4 Using a fair scoring function :

To bring in the element of competition (and thereby keep up the interest of the user), we need a fair scoring function which rewards the player for correct tags and penalizes her for incorrect tags. This is a tricky proposition because the ground truth is not available in advance. We suggest a collaborative scoring solution which takes into account the tags assigned by other players.

A good scoring function is needed for following reasons -

1. Maintain user interest to give better and relevant tags.
2. Not allow users to tame the game by some global consensus to give irrelevant tags.

For 2, specifically, we are making sure that a player and a video is selected at random., and there is no communication between them.

For 1, if a tag assigned by a user matches with tags assigned by other users then he will receive a high score for this tag. Here again, we need to consider direct match as well as semantic match. For example, if the player assigns a tag Durum Wheat and the set of tags assigned by other users contains Wheat (but not Durum Wheat) then he should still get credit (since Durum Wheat is semantically related to Wheat). Further, if the constructed taxonomy is already aware that Durum Wheat is a /emph type of Wheat then the player could be awarded extra credit for being more specific than the other players. Note that it will often be the case that a given player is the first player to view the video and hence a set of previously assigned tags is not available for scoring (*i.e.*, the current player is the first player to view the video). We propose to build an automatic tagger to overcome this problem. This tagger can be viewed as an bot player in the system.

Discouraging general tags: Tags which are very generic can be spotted by checking its presence on umpteen videos. We would like to discourage that *i.e.*, tags for eg., youtube, google, farmer[in our case] may not be of much help.

All this being said, let us have a mock play. *Let us dive in tagging ride..*

One of Shraddha's friend was telling her about the game 'Video-de-vidya'. She checks it out by signing up. After logging in, she watches a quick instruction tutorial which tells about rules, tips and other ways of scoring in the game. She goes ahead and starts playing the video given to her. She is performing well and also using concept suggestions wisely. Finally after watching entire video, gives every possible tag which she could relate the video to and clicks on submit button and then gives post viewing feedback. *But* she didn't seemed much enthusiastic in playing again.

Plans for further gamifying, bringing in the challenges, awards!!

1. Regular competitions where users are rewarded for entering the most tags in a week/double the scoring on particular day of week.
2. Encourage newbie.[Initial gift points]
3. Attracting and rewarding super taggers.[Leader points]
4. *Attracting new tags:* A player enters a pioneer tag when this tag is not entered for this video before,and afterwards it is matched with tags entered by other players,Increase gift points.[As these points are earned even when the player is not playing] We can change terminologies as per the theme. Ours being agriculture: something like green points, sun points, land points.
5. New videos to be given first to super-taggers. In this way they have the first chance to create novel tags, and we also can rely on tags generated by super taggers with more confidence.

Evaluations of a waisda?[10] a video labelling game in Dutch states that "Over 2,000 people played waisda? and within six months,over 340k tags have been added to over 600 items from the archive.

45.8% players added between one to ten tags. 35.3% added between ten and a hundred tags. 16.2% added between a hundred and a thousand tags. 2.7% added more that thousand tags but together were responsible for adding the largest number of all contributed tags! This indicates the necessity that this kind of project shouldn't only aim for a wide audience,but should also find a way to specifically target these super taggers."

They also quoted that "More than 70% of the traffic on the website was generated through referrals by external websites.The three main referring websites also resulted in the lowest bounce rate,suggesting that visitors that arrive at the website through an external link are more specifically interested in the content and the project than direct visitors." This provided a near practical results for us too,and we can keep these in mind for sustainability and scalability of the game.

Chapter 4

Meta-Learner

How will the bot learn?

Initial Settings: Following can be promising unsupervised features for a given video.

Textual meta-data: which uploader gives while uploading the video like Title, Category, Description, Keywords, Channel in which this has been uploaded etc. We ran an evolving **Associative Markov Network classifier** [18] built by *Ramkrishna Bairi* for Document categorisation on our video setting.

We passed these textual metadata per video as a single document. Lets us brief about working of AMN classifier and then checkout the results. AMN classifier first spots keywords from document, looks for candidate categories in Wikipedia, Initilaize the knowledge graph with those candidate categories and build nodes seeing the strength, association etc from wikipedia concepts. It then propagate the weights and give the potential categories as results and uncertain ones for feedback. And retrains the model and the evolving continues.

Following are the results with no feedback:

1. *Title :* Watch the Winged Weeder in action!:

Description: Marty Klipper demos and describes some of the myriad uses of the Winged Weeder garden tool. For most new Winged Weeder owners, this specialized tool becomes a favourite tool within a year, and the exclusive tool within two years.

Keywords: ["Winged", "Weeder", "garden", "weeding", "weeds", "Marty", "Klipper", "soil", "furrow", "flower"]

Facets: Klipper, Glipper

2. *Title:* Agriculture of Jute:

Description: Agriculture of Jute

Keywords: ["Agriculture", "of", "Jute"]

Facets: Jute, Rope, Fiber crop, Kenaf, Fiber, Adamjee Jute Mills, Jute trade, Corchorus, Hessian (cloth), Twine, Bast fibre, Natural fiber, Jute cultivation, Jute genome.

This motivates us for a good start. However there was high topic drift like in case 1 : Klipper was a person's title not the Clipboard manager which through Wikipedia has been wrongly inferred. We also found that whenever there was some proper noun, It delve deeper along those lines and ignoring other potential categories. Hence, this motivated us to look into Wikipedia for knowledge expansion, but mining domain-specific content and accounting for the kind of relations between nodes from the unstructured Wikipedia remains a challenge.

Image meta-data: From the set of images extracted earlier in section 3.3, we feed it to some Image recognition API's like Alchemy[2], Watson[11], JustVisual(paid) [7] and get to know what the image is about and in turn, what the video is about. As the figures shows, Alchemy proves to be a decent candidate in general settings. JustVisual is good but is a paid service. It tells very correctly what the image contains in scientific or very specific terms. Watson seems to give random and very abstract results.

However, adapting these tools to the domain and task at hand poses several challenges. Firstly, these tools are trained to recognize specific objects (typically, of the order of 10000 objects). As a result, they may not provide a good coverage for objects and entities which are important in the domain of interest. Secondly, most of these tools focus on *object detection* and not on *process detection*. We are interested in recognizing the process "corn threshing" for instance from figure 3.3 which cannot be done using these off-the-shelf tools as it is. Hence, there may be some research on domain specific Image tagging further, and that understanding the process as such does not seems to be trivial. Note that there may also be more errors creeping in the system when images are blur, fade-in/out or not of good quality. A check on these need to be done.



Image	Alchemy	JustVisual	Watson
	clock 0.598688 gauge 0.28905	CafePress ABARTH Wall Clock CafePress Maltese Portrait Wall Clock CafePress Fire Chief Wall Clock CafePress Waltham Railroad Pocket Watch 2 Large Wall Clock	Indoors 68% Meat Eater 67% Room 66% Object 65% Vertebrate 65%
	banana 0.5 vegetable 0.401312 fruit 0.354344	plantains(Musa balbisiana)	Flower 77% Human 67% Food 67% Group of People 62% Indoors 60% Activity 56% ClubSport 55% Nature Scene 55%

Figure 4.1: Image Tagging results

Audio transcription again, can also be of help similar to image tagger. We hope that we will get good transcription of audios from google speech-to-text systems. However we won't be able to generalize it to regional languages although. So, major bottlenecks would be lesser language support and videos with less or no human-speech., which from our video repository of farming videos, seems to be the case. Mostly it contains machinery/musical sounds.

Our Future Plan in Meta-Learner: Till now we saw how text and image feature could help

us know about the video tags. We observed that there is no learning about users of video categories being done. So we would like to design a meta learner which would make use of the data available in form of (video,tag set) pairs to learn video categorization.

We would use the above modalities as static base learners like image classifiers, audio/metadata based classifiers. Our players will also act as learning models, based on which our final learner will evolve. This problem is closely related to multi-label classification problem in unsupervised settings. The final aim of learner would be to classify the video based on feature set into a class consisting of different labels or to reach a consensus based on labels given by users.

In our problem we have following adaptations and concerns :

1. The boosting learners require training data with true labels. We need to generate initial training data to initialize the learner.
2. We would treat each of our players as base or 'weak' learners and model them using the inputs we receive from them. Then use AdaBoost [20] like algorithm over them. The data available would be sparse so need some other online algorithm like LPBoost[14].
3. However the tag set we obtain from users are noisy and do not represent true label set. So we have to design a robust consensus algorithm which will predict near to accurate truth labels.
4. Also, we have to handle online updates.
5. We have to take into account the initial hierarchy structure among the classes and frame the problem accordingly.

Chapter 5

Conclusion and Future Work

Annotations of video files can be used to search and index video databases, provide data for system evaluation, and generate training data for machine learning. Unfortunately, the cost of obtaining a comprehensive set of annotations manually is high. One way to lower the cost of labeling is to create games with a purpose that people will voluntarily play, producing useful metadata as a by-product.

Assuming most taggers are fairly inexperienced, We will after project completion make a tutorial video depicting following points strongly through game session:

1. Temporal aspect of the moving images tagged within the game.
2. Also make them aware of the fact they can add multiple tags over the course of the whole video., and that they can still add a tag after the subject has disappeared from the image because they weren't done typing.
3. We will also bring transparency in scoring mechanism., so that player doesn't feel lost/cheated.

In this report, we described various sub-problems that need to be addressed to build a system for semi-automatically tagging videos.

Future Research:

1. Design a fair scoring mechanism based on the knowledge graph built.
2. Evolving Knowledge Graph.. on how to add novel concepts in the graph.

References

- [1] AGROVOC Multilingual agricultural thesaurus. <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>.
- [2] Alchemy Image tagging API. <http://www.alchemyapi.com/>.
- [3] Commonsense Reasoning in and over Natural Language. <http://alumni.media.mit.edu/~hugo/publications/papers/KES2004-csr-nl.pdf>.
- [4] Farmer Portal: "One Stop Shop For Farmers". <http://farmer.gov.in/>.
- [5] FFMPEG. <https://www.ffmpeg.org/>.
- [6] Hindi Wordnet. <http://www.cfilt.iitb.ac.in/wordnet/webhwn/index.php>.
- [7] Just Visual Image tagging. <http://www.justvisual.com/>.
- [8] Lokavidya Android Application. <https://play.google.com/store/apps/details?id=com.iitb.mobileict.lokavidya>.
- [9] Presentation on Classification of Videos using semi-automated tagging techniques. <https://docs.google.com/presentation/d/12SqQJ9-CQt11kXIEn3dec98V7iw6u0tHAsSwSq7BzG4/edit?usp=sharing>.
- [10] Waisda?Video Labeling Game:Evaluation Report. <http://research.imagesforthefuture.org/index.php/waisda-video-labeling-game-evaluation-report/>.
- [11] Watson Image tagging API. <https://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/>.
- [12] wikihow. <http://www.wikihow.com/Main-Page>.
- [13] Wiktionary, the free dictionary. https://en.wiktionary.org/wiki/Wiktionary:Main_Page.
- [14] Thomas Pock Christian Leistner Amir Saffari, Martin Godec. Online Multi-Class LPBoost. In *Conference of Computer Vision and Pattern Recognition (CVPR)*, pages 3570 – 3577, San Francisco, CA, 2010.
- [15] Ian H. Witten David Milne, Olena Medelyan. Mining Domain-Specific Thesauri from Wikipedia: A Case Study. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, WI '06*, pages 442–448, Washington, DC, USA, 2006.
- [16] Steven Bird Edward Loper. NLTK: the Natural Language Toolkit. ETMTNLP '02, pages 63–70, Stroudsburg, PA, USA., 2002.
- [17] George A. Miller. WordNet: a lexical database for English. In *Communications of the ACM*, pages 39–41, New York,NY,USA, 1995.

- [18] Vikas Shindwani Ramakrishna B Bairi, Ganesh Ramakrishnan. Personalized Classifiers: Evolving a Classifier from a Large Reference Knowledge Graph. IDEAS 14, pages 132–141, Porro, Portugal., 2014.
- [19] Sachindra Joshi Subhabrata Mukherjee. Sentiment Aggregation using ConceptNet Ontology. In *International Joint Conference on Natural Language Processing*, pages 570–578, Nagoya, Japan, 2013.
- [20] Robert E. Schapire Yoav Freund. A Short Introduction to Boosting. In *Journal of Japanese Society for Artificial Intelligence*, pages 771–780, Japan, 1999.