# Classification of Multi-Media Content (Video's on YouTube) Using Tags and Focal Points

**By:**

Ankur Satyendrakumar Sharma [shar0306@umn.edu]
(Department of Computer Science and Engineering)

Mohamed Elidrisi [elidr002@umn.edu]
(Department of Computer Science and Engineering)

## Abstract:

Tags are the new valuable source information in the web. The multi-media content has been heavily tagged by the owners and the viewers. Thus, Tags represent a social classification of the content and at the same time it also adds to its semantics. We intend to base our classification on the information carried in tags. This relieves us from the pain of explicitly performing the analysis and summarization of multimedia, which potentially could be a costly affair. However, the basic shortcoming in the usage of tags lies in the looseness of it's the representation. As tags are created by human beings they represent the human interpretation of the multimedia content, the personal bias of the human being comes into the context, which in turn acts as noise in this scenario. Thus, it becomes difficult to identify tags that are relevant to the multimedia content. Tags not only act as a mechanism of describing a multi-media object, but as means of communication to the other users. But due to the tacit nature of the communication, the concept of focal point (from Thomas Schelling) comes into play. The project is an attempt to classify the videos present on YouTube on the basis of Tags and Focal Points.

# Introduction

With the emergence of Web 2.0, the Multimedia content has shown an exponential growth, both, in terms of quantity and views. For instance, millions of videos, photographs and office presentation are uploaded and viewed over the web daily. This has led to an urge for better categorization and maintenance of the content.

Textual data, like plain old web pages, provide with an opportunity to generate certain level of formal description that can be exploited for generating low level semantics and classification. This is not the case with multimedia content like videos.

*"The main challenge is to narrow the semantic gap between the low-level content descriptions that can be computed automatically by a machine and the richness and subjectivity of semantics in high-level human interpretations of audiovisual media." [2]*

### The Problem with Multi-Media Content

Unlike text data, Multi-Media Content (MMC) is way too much rich and expressive; however this poses a fundamental problem towards large scale semantic analysis of the MMC. In real time environment, like web, it is not only economical unviable, but also technically difficult to carry out analysis of MMC and generate sane results in reasonable amount of time. For example, it would have been impossible to service a user search request at YouTube (a storehouse of at least 4 million videos), if traditional video analysis was to be used.

### The Solution to the Problem

Tags have emerged as very powerful mechanism of information re-discovery from users perspective and a service providers (like YouTube) perspective. Over the web not only the textual information, but also the MMC information has been tagged. Users constantly use tags as means of finding and re-finding the resources they are interested in as well as the resources they want to be viewed by other users. A user utilizes tags from two perspectives:

a) In order to recall some information in some later point of time. Thus making the information recall process less time consuming. For example in case of YouTube, users tend to exploit YouTube as a video storing repository and usually store videos that are personal to them. In such cases, they would like to use tags that give them high recall rate, while a very low or no discovery on the part of other users.
b) In order to share the information with other users, with whom the user may not have any relationship, i.e. the user wants the information submitted by her to be discovered by other users with maximal ease. In case of YouTube, users also upload videos which they want to be viewed by maximum number of users; such videos can be personal video blogs, recent news item or even product advertisements. In these scenarios, they want the discovery of the video by other users to be as easy as possible and thus use tags that concisely describe a video.

Table-1 briefly outlines the concept of Tag.

| Table 1 - Profile Of a "Tag" | |
|---|---|
| **Definition:** | |
| ***Loose*** | Tags are the keywords or labels that assigned to an object. |
| ***Strong*** | Tags are textual meta-data, i.e. text data about data, where the actual data can be a multi-media object or a text object. |
| **Properties:** | • Assignment of Tags to an object is highly user centric.<br>• They are loosely structured, i.e. normally they are represented in the form of *words* separated by some delimiter like white spaces or comma.<br>• The world of tags is inherently *flat*. There is no hierarchy in the relationship between tags that describe an object. |
| **Benefit:** | • Ideally tags represent the principal semantic component of an object.<br>• The presence of linguistics in Tagging provides stability to otherwise complex environment [3].<br>• Tags are extensively used and play an important role in information retrieval on large tag based databases like YouTube. |
| **Issues:** | • They are loosely structured.<br>• People tend to use same tags for different documents.<br>• People also to tend to use semantically un-correlated tags for the same document. |

## The Classification Problem

As the multi-media content over the web is growing quantitatively, classification of the content is become more and more important for various reasons:

- Content based advertising has played a key role for revenue generation over the web, thus with the emergence of the rich multi-media content it is logical to exploit it for monetization. Therefore in order to enable this mechanism the multi-media content needs to be classified.
- Classification can play a critical role in enhancing the user experience. For example, when people upload videos on YouTube, they normally know what the relevant tags of the video are, but find it difficult to decide to which category the video belongs, thus a suggestion mechanism would really ease out the entire process.

## Problem Domain

The multi-media content domain that the project has addressed is "Online Videos" at YouTube™ (http://www.youtube.com). YouTube has a very large user base, as it serves more than 100M requests a day. Thus, there is no shortage of any kind of data on YouTube. Not only this, YouTube also provides developer API's that researchers and web developers can use to query the YouTube database from various perspective. The project queried the YouTube database from the perspective of retrieving videos and their corresponding tags and categories.

The videos on YouTube are tagged and categorized by the user who uploaded the video, such a user is referred to as the author/owner of the video. Only the author of the video can add or change tags and categories on YouTube. Other users at YouTube, who act as consumers/viewer of video, can add only comments to the video. They cannot change the category or the tags of the video, while at the same time they can rate the video.

## Problem Definition

Given a Video $V$ and set of associated tags $T=\{t1,t2,t3,\ldots,tm\}$, where $m\geq 1$, determine the category $c$, where $c$ belongs to $C=\{c1,c2,\ldots,cn\}$ $n\geq 2$. $C$ represents the category space, i.e. set of all the possible categories that a given video could belong.

Constraints:

- A video $V$ can only belong one and only one category of $C$.
- Set $T$ should be at least a singleton set.
- Set $C$ should have at least two elements;

## Focal Points

In order to solve the above problem, the project tries to use the concept of focal point along with pattern classification methodology. The intention was to answer the question whether the performance i.e. classification capability, of a classification algorithm be improved (indirectly) if the underlying data incorporates some of the focal point features.

| Table 2 - Profile Of "Focal Points" | |
|---|---|
| **Definition:** | A focal point, also called Schelling Point [A concept introduced by Thomas Schelling in his book *The Strategy of Conflict*], is a solution that people will tend to use in absence of communication, because it seems natural, obvious and relevant to them.[6]<br><br>Thomas Shelling, in the book "*The Strategy of Conflict*", describes "focal point[s] for each person's expectation of what the other expects him to expect to be expected to do"[5] |
| **Example:** | A classic example is the solution most people choose when asked to divide \$100 into two piles, of any sizes; they should attempt only to match the expected choice of some other, unseen player. Usually, people create two piles of \$50 each, and that is what Schelling dubbed a focal point.[4] |

The fact that focal points (please refer to table 2) come into play when there is lack of communication between two people, makes them an interesting concept to be included in this project. Current trend has shown that people upload the video so that it can be viewed with the maximum number of audience. As there is a fundamental communication divide between the author of the video and the audience, focal points come into play in the form of tags that the author provides to the video.

# Related Work

As web2.0 is relatively young, the associated data too is relatively raw. The issue of large scale classification of web based multi-media content has remained largely unaddressed. This is may be due to the fact that a stronger emphasis was on the information retrieval, i.e. designing the right kind of query mechanism.

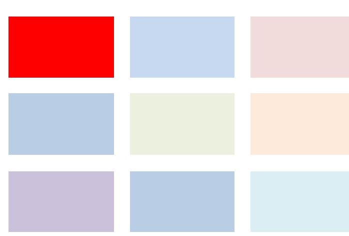
In the context of this project, the most co-related work has been done in [1], which is about clustering of tags. [1] emphasized the generation of appropriate tag cluster/clouds, so that document then be classified and thus transforming the flat tag space into a one that is hierarchical. The categories to which the content belonged was not known before hand, so the process followed by [1] was of clustering rather than classification. [1] used approximately 30,000 tags from 200,000 documents for their clustering mechanism. In order to cluster, [1] generated a co-occurrence matrix, a matrix showing the degree of occurrence of two given tags at the same time in a document across the document space. The concept of co-occurrence matrix has been used and extended in the project.

However, instead of creating a cluster of tags, the project directly classifies videos. This has been possible due to availability of the categorical information. YouTube allows a user to categorize a video into one of the twelve categories only. As the category information is present, the entire project is modeled on supervised learning methodology. Also the project takes into consideration approximately 350,000 tags generated using 600,000 videos.

[4] first suggested the coupling of focal point with machine learning methodologies and showed that how the performance can be improved by incorporating right kind of focal point in the feature space. However, the usage of focal point is shadowed by the fact presented in [4] that with an increase in the complexity of domain, there is no role played by the focal points. The experiments conducted in [4], were from a domain (shape matching game – a game between human beings and computer agent) that was very simple as compared to the current domain of Online Videos of YouTube. So our question whether Focal points can be used for improving the performance of large scale classification is still open.

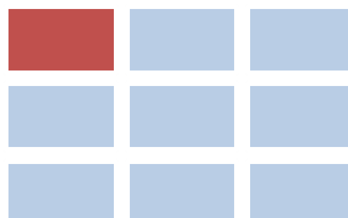
[4] identifies four general kind of focal points that people tend to use:

1. **Centrality:** People tend to give prominence to a choice that is at the center of set of choices. The example quoted in table 2 is a classic example of this scenario.
2. **Extremeness:** People tend to give prominence to the choices that are extreme to other choices. Choosing a red color square a 3x3 grid, where all other squares of light color.



3. **Firstness:** People tend to give prominence to the choices that appear first in the set of choices. The example for this case is the tags for some content. The first tag usually represents the most important idea about the video.
   *Video Title:* Apple WWDC 2006-Windows Vista Copies Mac OS X
   *Video Link:* http://www.youtube.com/watch?v=N-2C2gb6ws8
   *Video tags:* Apple Macworld Keynote Mac Macintosh Steve Jobs Bill Gates Windows OSX Computer ipod imac ibook power macbook pro vista

   The video predominantly deals with Apple computers; under this assumption the author of this video gave more priority to the tags like apple and MacWorld that have high co-relation with the video by placing them first!
4. **Singularity:** People tend to give prominence to the choices that are unique as compared to the other set of choices. Choosing a red colored square on a 3x3 grid, where all other squares are of blue color.



# Our Approach

The overall approach can be summarized by the following diagram:

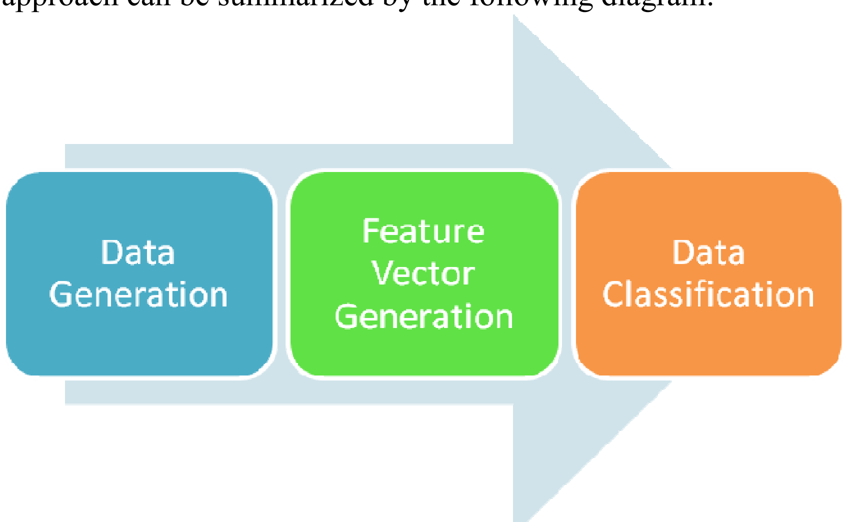



**Figure 1**

## Data Requirement

For the purpose of the effective classification we were required to collect the following fields for a video:

1) *Video's Author:* The name of the author of the video. This is may help understand the nature an author tags the videos, given a set of video submitted by the author.
2) *Video Tags:* The set of tags that are given by the author to the video.

3) *Video Category:* The category of the video as suggested by the author. Again, YouTube has levied a constraint on the categorization of videos by providing the author with 12 categories to choose.
4) *Video Link:* The actual hyperlink to the video for later access to it.

## Data Generation

YouTube is a fully owned subsidiary of Google. Out of various developer tools that Google provides, it also provides API's for accessing the YouTube database. A developer only has the read access to the database and the query mechanisms though are limited and predefined, are good enough to give access to all kinds of relevant data.

Out of various available querying mechanisms, we choose to query the YouTube database on the basis of tags, i.e. given a tag we tried to find out what are the associated videos with it. This way the we not only got the videos as a part of our search result, but also the set of associated tags with them, which were later used to query the YouTube database for more videos. Thus, over all the use of tags and the process of data generation can be summarized as a breadth first search process. The algorithm is as follows:

Algorithm:

**i. While (***there are un-queried tags in the database***)**
**ii. Generate list of un-quereid tags call it** *taglist*
**iii. For each** *tag* **in** *taglist*
**iv. If** *tag* **is an English word**
**v. Query YouTube for videos**
**vi. Push new tags from videos in the database**
**vii. Else Do nothing**
**viii. For End**
**ix. While End**

In order to limit the scope of the project, only those tags that had English alphabets (also excluding numbers) and were valid English words, were used for querying purpose. While all those tags that had English alphabets, but were not valid English words were retained in the database for the future analysis.

## Data Statistics

*Miscellaneous*

| | |
|---|---|
| Number of videos in the dataset | 813,119 |
| Number of unique Tags in the dataset | 423,797 |
| Number of Tags that are valid English words | 49,248 |
| Number of Categories | 12 |
| Number of authors | 438,183 |

*Data Distribution Over Categories:*

| *Category* | *Number of videos* |
|---|---|
| Games | 33,066 |
| News | 43,335 |
| Comedy | 80,150 |

| | |
|---|---|
| Film | 57,200 |
| Travel | 22,081 |
| Autos | 21,698 |
| Howto | 41,165 |
| People | 91,669 |
| Sports | 50,798 |
| Music | 218,092 |
| Animals | 21,784 |
| Entertainment | 132,081 |

*Tag Analysis:*

| | |
|---|---|
| Number of Tags | 423,797 |
| Average Length of a Tag | 7.729736170855386 |
| Maximum length of a Tag | 26 |
| Minimum length of a Tag | 1 |

*Video Analysis:*

| | |
|---|---|
| Average Number of Tags Per Video | 9.056568595740599 |
| Maximum Number of Tags Per Video | 70 |
| Minimum Number of Tags Per Video | 1 |

## Feature Vector Generation

The tags that are present in a video cannot be directly used as feature vectors. Also, the number of tags per video varies from one video to other. Thus, there was need of consistent feature space that could be applied to all the videos in the data set.

***Feature Space***

YouTube identifies 12 categories for a video. Our preliminary analysis showed that there was an inherent hierarchy in this categorization, shown in the figure 2. In order to avoid the data over-lap that would have occurred due to this hierarchy, we decided to take only those categories that were leaf-nodes of the tree. Thus, we have 10 categories for the classification instead of 12.

In order to identify feature space dimensions, we decided to evaluate distance of each video from each of the 10 category, disregarding the actual category provided by the user. Thus, for each video there will ten such evaluations, one for each category, giving rise to a 10-dimensional feature space.

The distance measurement of a video from a category was based on two factors:

1) *Semantic Distance*: i.e. how is the video closely associated to category on semantic grounds.
2) *Symbolic Distance*: i.e. considering tags as mere symbols rather than words with some meanings, a video was evaluated for the fact that how much similar it is, symbolically, to some category.

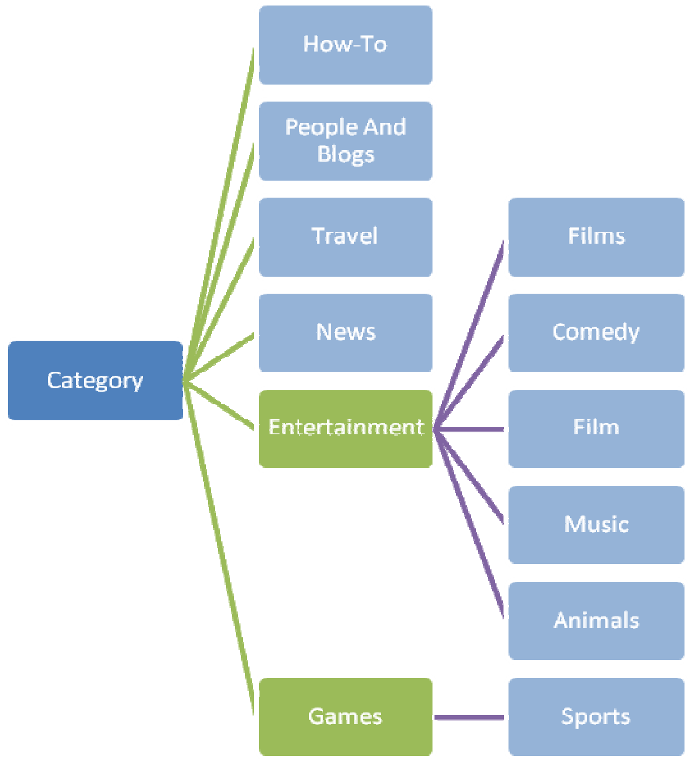



**Figure 2 – Categorization Hierarchy**

***Semantic Distance Generation***

For every category under consideration a thesaurus tree was generated. A thesaurus tree is tree like structure with parent nodes representing words and the child nodes representing the synonyms of the corresponding parent nodes. It is a tree-like and not an exact tree because the children nodes may point to some nodes at higher level that are not their parent nodes. This is due to the fact that in the synonym set of a word there could be word that has already appeared and is at a higher level. Figure-3 shows a very simple thesaurus tree.

Such a tree could extend to infinity, covering the entire dictionary, but for the sake of computational simplicity, the tree was constructed till level 5 only. Once such a tree is constructed every word in the tree is assigned a weight according to the following factors:

a) *Weight Due To Level*: A word is weighted on the basis of level it is present. The higher the level more the weight the word. Here Level 1 is the highest level. It is logical to assign more weights to the words that are higher level, as they represent the semantic context of that category more crisply than other words that are lower to them
b) *Weight Due To Cross Linking:* As words may link each other across the same level as well from lower level to the upper level. A word is weighted according to the cross-linking it has received (the in-degree). More the in-degree of a word more the weight it will be assigned. This has been stemmed from the fact that words that have more in-degree generate strong semantic context for the category.

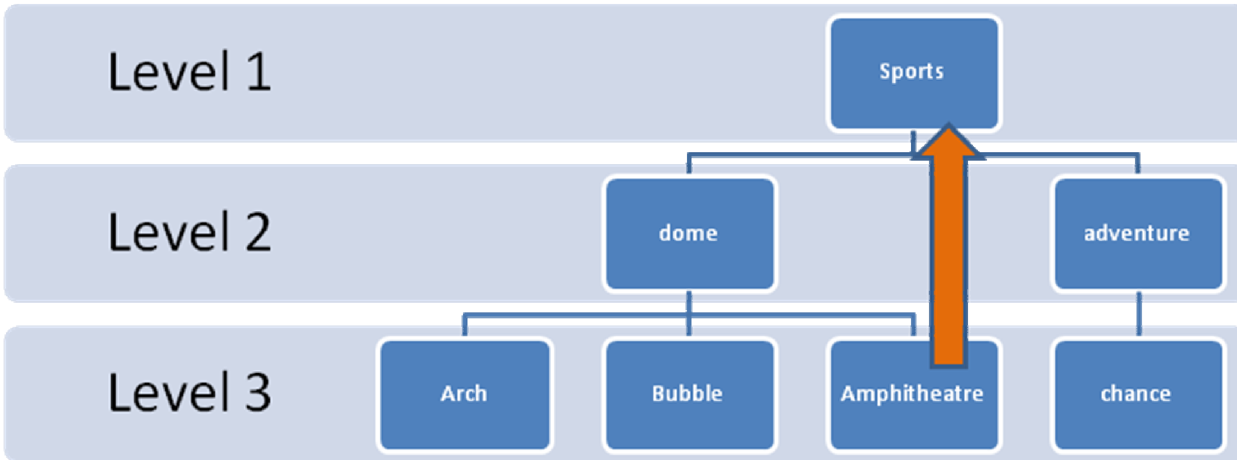



**Figure 3 – Figure showing a very simple thesaurus tree for the category sports. The important point to be noted here is that the word "*Amphitheatre*" points to a word at a higher level.**

Once such tree is constructed for each of the 10 categories, a video's semantic distance is calculated by summing up the semantic distance of each of the tags of the video from the 10 categories. The semantic distance of tag is calculated by looking up weight of that tag in each of 10 thesaurus trees. This is, however, based on the assumption that the tag is an English word. In cases, when the word is not an English word, it was out-of-scope of the project to determine its semantic sense, thus such non-English tags were assigned a default weight (in this case 1). Also, for those tags that were English words, but for those there was no entry in the hierarchy tree [this may be due to the fact that the tree was generated till level 5 only], default weights were assigned to them.

The semantic distance vector a video V is shown below, where Se1 to Se10 are the semantic distance values for the video for each of the 10 categories:

$$V = [\ Se1, Se2, Se3, Se4, Se5, Se6, Se7, Se8, Se9, Se10]$$

A thesaurus tree with nodes till level 5, on an average, had 80,000 words. *It is to be noted that even though the category information is used in generation of thesaurus trees, they are ultimately used as lookup databases by video for feature vector generation.*

***Symbolic Distance:***

The concept of co-occurrence matrix as suggested in [1] is used here. However, [1] generates the co-occurrence matrix for the entire document space, in light of the current project the co-occurrence matrix is generated for each category space.

*A co-occurrence matrix is a matrix that shows the degree of occurrence of two given tags together in a video, across the considered space (category space).* A typical co-occurrence matrix is shown in table 3. The co-occurrence matrix is a symmetric matrix and is sparser than what is shown in the table.

Table 3 – Sample Co-occurrence Matrix

| | Tag 1 | Tag 2 | Tag 3 | Tag 4 | Tag 5 | Tag 6 | Tag 7 | Tag 8 | Tag 9 |
|---|---|---|---|---|---|---|---|---|---|
| Tag 1 | - | | | | | | | | |
| Tag 2 | 0 | - | | | | | | | |
| Tag 3 | 0 | 1 | - | | | | | | |
| Tag 4 | 3 | 2 | 3 | - | | | | | |
| Tag 5 | 3 | 6 | 0 | 9 | - | | | | |
| Tag 6 | 9 | 0 | 0 | 3 | 5 | - | | | |
| Tag 7 | 3 | 0 | 5 | 7 | 8 | 0 | - | | |
| Tag 8 | 5 | 3 | 0 | 0 | 0 | 8 | 9 | - | |
| Tag 9 | 1 | 3 | 2 | 0 | 6 | 3 | 2 | 7 | - |

Normally for a given category there are on an average 45,000 tags. So a co-occurrence matrix is normally of the size 45000 x 45000, but the matrix has a sparse nature due to the fact that there are on average 8-9 tags per video.

Once such a co-occurrence matrix has been defined for each of the categories, they are used as look up tables for calculating symbolic distance of a video form a given category. For each tag pair that is possible in a tag set of a video, a lookup is made in a co-occurrence matrix of a category for the co-occurrence value. The summation of all these co-occurrence value represents the distance of the video from that category. This process is repeated for all the categories.

The symbolic distance vector a video V is shown below, where Sy1 to Sy10 are the symbolic distance values for the video for each of the 10 categories:

$$V = [\, Sy1, Sy2, Sy3, Sy4, Sy5, Sy6, Sy7, Sy8, Sy9, Sy10]$$

***Focal Points:***

As discussed previously, out of the set of focal points identified by [4], the project uses *firstness* as the focal point feature. Thus, according to this feature the tags that are appear first in the tag set are more important than other tags in the set, as they represent one of the focal point of the author's/owner's tag creation process.

In order to in-corporate this feature, no explicit feature space dimension was created, instead changes were made in the calculation process of symbolic distance and semantic distance. The contribution made by the first tag in a given videos symbolic and semantic distance was given a boost my multiplying the tags symbolic and semantic distance by some multiplier *M(=5 for the implementation)*.

***Feature Vectors:***
Finally, the feature vector for a video is generated by multiplying the symbolic distance with the semantic distance. Thus, the feature vectors for a video looks like the following:

$$V = [\ Se1*Sy1,\ Se2*Sy2,\ Se3*Sy3, \ldots,\ Se10*Sy10\ ]$$

Over all, six types of feature vectors were generated:

1) Feature Vectors containing only Semantic information
2) Feature Vectors containing only Semantic information with focal point boost
3) Feature Vectors containing only Symbolic information
4) Feature Vectors containing only Symbolic information with focal point boost
5) Feature Vectors containing both Symbolic and Semantic information
6) Feature Vectors containing both Symbolic and Semantic information with focal point boost

Performance was evaluated for all the six types.

### Classification

| **Tool Used** | **Weka-3.4.11** |
|---|---|
| **Tool Type** | Off-The-Shelf |
| **Classification Algorithm** | Classification (M5P Trees) Via Regression |

For deciding the algorithm to be used for the classification, several runs were made using different classification algorithms provided by Weka for a small set of data. Algorithms like K-Nearest Neighbor, Bayes Net and Multilayer perceptron along with Classification via Regression were tried, out of which classification via regression consistently generated the lowest test error.

## Results

This section presents results of the runs carried out for different set of feature vectors and the plot of data in low dimensional space.

***Feature Vectors with only Semantic Information with and without focal point boost:***
An attempt was made to classify the videos just on the basis of semantic information. But from the data visualization itself, it was clear that the data at this stage was fundamentally inseparable. The nature of the data remained unchanged even when the focal point boost was used. Following is a plot of two categories plotted in three dimensional space i.e. instead of considering all 10 categories [10 dimensional space] only 3 categories were considered and the plot was generated:

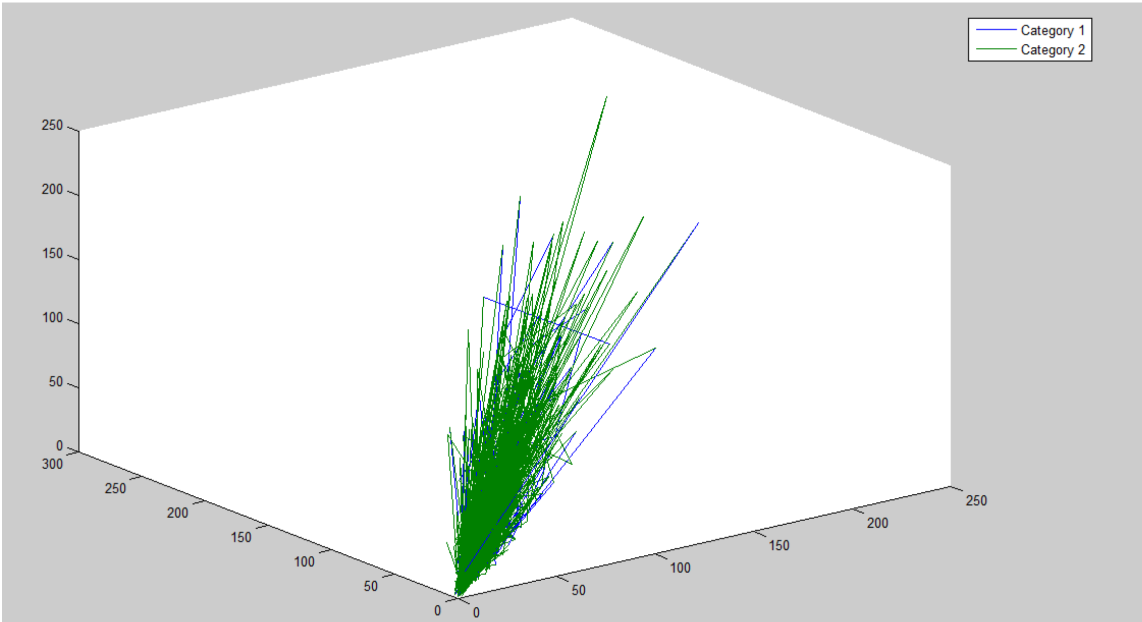



**Figure 4 – Plot of two categories in three dimensional space**

With this kind of nature of data it was clear that the Semantic analysis alone won't suffice.

***Feature Vectors with only Symbolic Information:***

An attempt was made to have a classification based on just symbolic information. Three categories and their corresponding category space were taken into consideration.
Following is the plot of data:

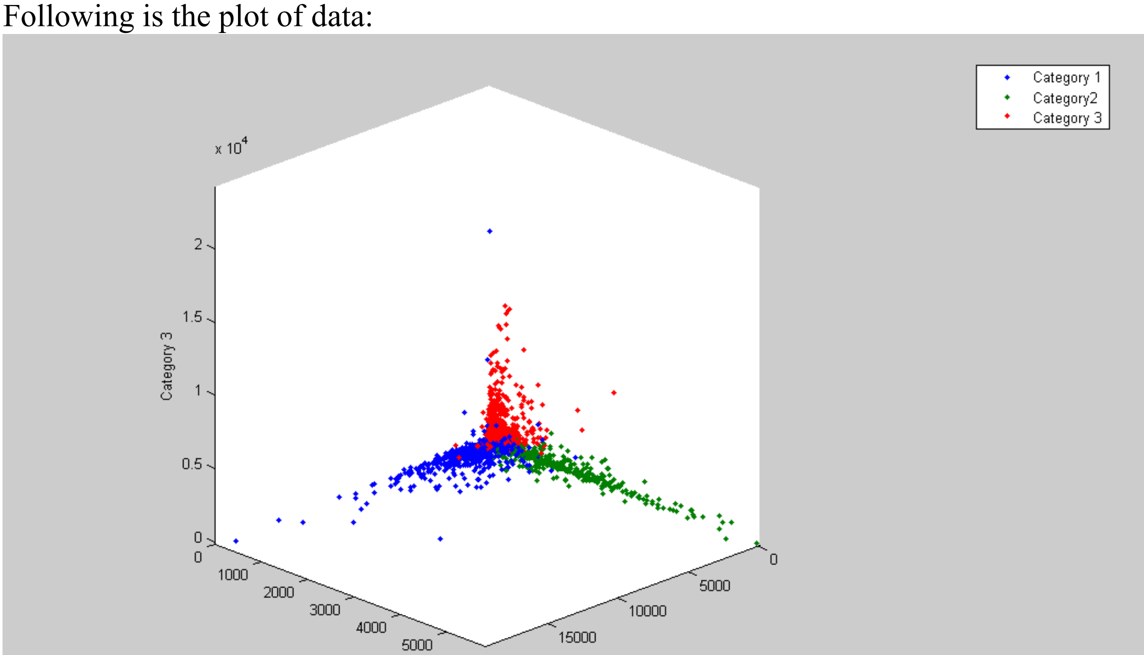



**Figure 5 – Plot of three category in their respective category space**

It can be seen that the data in this case is clearly separable.
Following is the classification performance:

| | |
|---|---|
| Algorithm Used | Classification Via Regression |
| Categories Involved | 3 |

| | |
|---|---|
| Vectors For Each Category | 1000 (Total = 3000 Vectors) |
| Training Data | 66% of Total |
| Test Performance | 90.576% correctly classified |

***Feature Vectors with only Symbolic Information with focal point boost***:
An attempt was made to have a classification based on symbolic information along with focal point boost. Three categories and their corresponding category space were taken into consideration. Following is the plot of data:

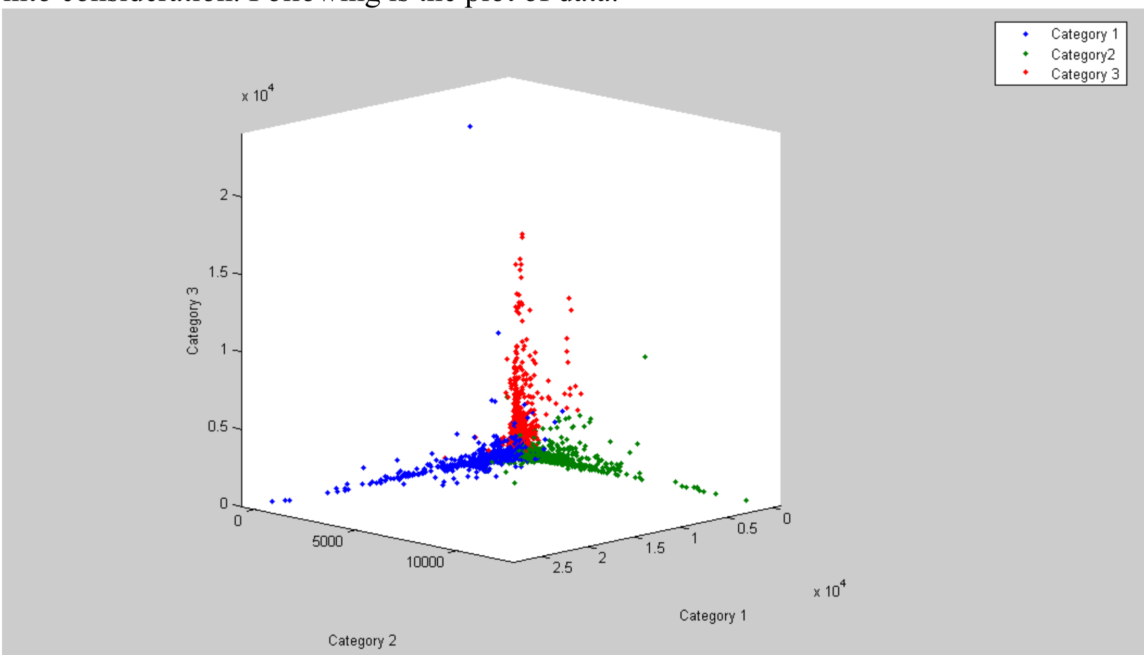



**Figure 6 – Plot of 3 categories in their category with focal point boost**

Again, it can be seen that the data in this case is clearly separable.
Following is the classification performance:

| | |
|---|---|
| Algorithm Used | Classification Via Regression |
| Categories Involved | 3 |
| Vectors For Each Category | 1000 (Total = 3000 Vectors) |
| Training Data | 66% of Total |
| Test Performance | 91.3725% correctly classified |

So as compared to the previous case, there is an increase in the classification performance.

***Feature Vectors with Symbolic and Semantic Information***:
An attempt was made to have a classification based on symbolic and semantic information without focal point boost. Three categories and their corresponding category space were taken into consideration. Following is the plot of data:

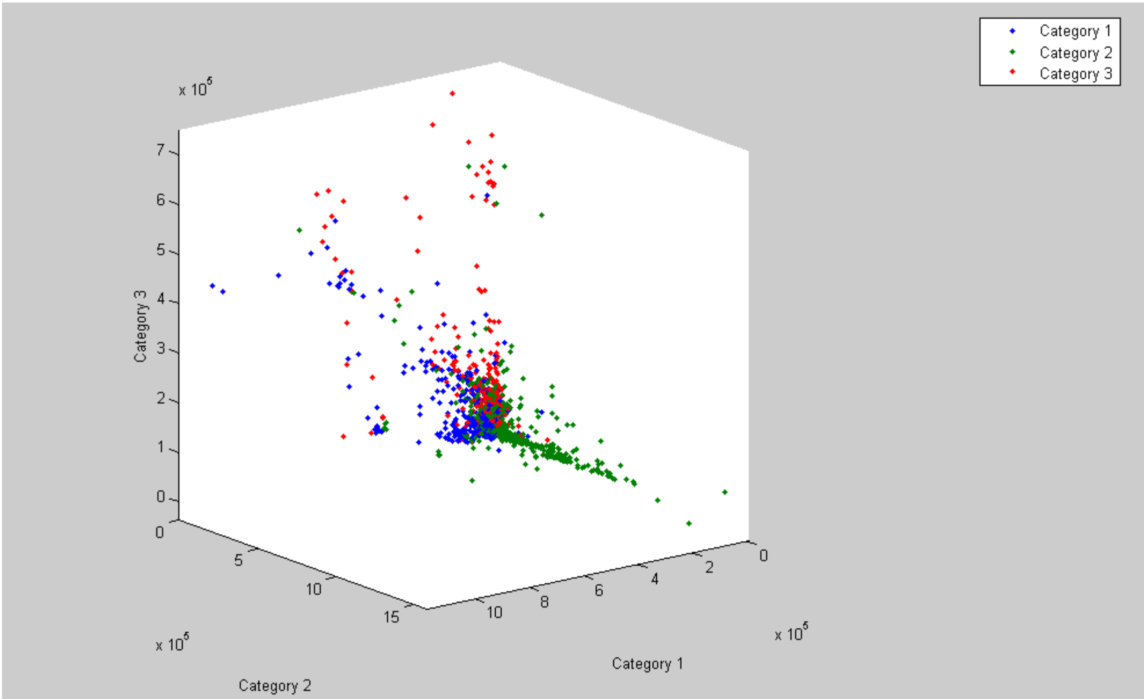



**Figure 7 – Plot of three categories with symbolic and semantic information**

Strangely, data loses its degree of separability in this case and because of this there is deterioration in the performance of the classifier.
Following is the classification performance:

| | |
|---|---|
| Algorithm Used | Classification Via Regression |
| Categories Involved | 3 |
| Vectors For Each Category | 1000 (Total = 3000 Vectors) |
| Training Data | 66% of Total |
| Test Performance | 69.8039% correctly classified |

***Feature Vectors with Symbolic and Semantic Information along with Focal point boost****:*

An attempt was made to have a classification based on symbolic and semantic information along with focal point boost. Three categories and their corresponding category space were taken into consideration. Following is the plot of data:

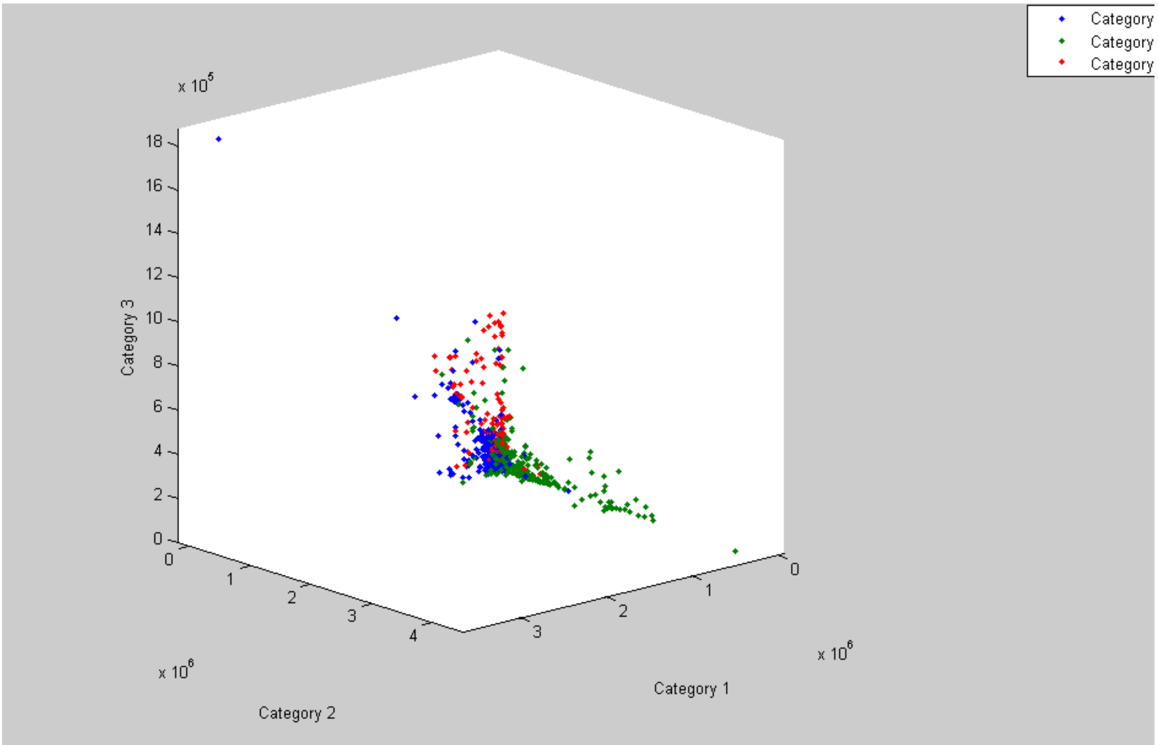



**Figure 8– Plot of three categories with symbolic and semantic information**

Again, the data loses its degree of separability in this case and because of this there is deterioration in the performance of the classifier. However, because of the inclusion of the focal point boosting the performance is better than the previous case.
Following is the classification performance:

| | |
|---|---|
| Algorithm Used | Classification Via Regression |
| Categories Involved | 3 |
| Vectors For Each Category | 1000 (Total = 3000 Vectors) |
| Training Data | 66% of Total |
| Test Performance | 72.6461% correctly classified |

***Other Runs: RUN-1***
Following is the run information:

| | |
|---|---|
| Algorithm Used | Classification Via Regression |
| Categories Involved | 3 |
| Category Space (Dimensionality of Space) | 4 |
| Vectors For Each Category | 5000 (Total = 15000 Vectors) |
| Training Data | 66% of Total |

Following are the results for the run for different types of feature vectors:

| *Run Type* | *Test Performance* |
|---|---|
| Semantic + Symbolic Information | 76.116% correctly classified |
| Semantic + Symbolic + Focal Point Information | 76.8627% correctly classified |

| | |
|---|---|
| Symbolic Information | 86.5686% correctly classified |
| Symbolic + Focal Point Information | 89.1275% correctly classified |

***Other Runs: RUN-2***

Following is the run information:

| | |
|---|---|
| Algorithm Used | Classification Via Regression |
| Categories Involved | 10 |
| Category Space (Dimensionality of Space) | 10 |
| Vectors For Each Category | 15000 (Total = 150,000 Vectors) |
| Training Data | 66% of Total |

This run includes all the 10 categories and their respective category space (10 dimensional space).

Following are the results for the run for different types of feature generations:

| *Run Type* | *Test Performance* |
|---|---|
| Symbolic Information | 62.8902% correctly classified |
| Symbolic + Focal Point Information | 64.4902% correctly classified |

Again here too, though the run performance is less, the involvement of focal point boost has increased the performance.

## Conclusions and Future Work

Following are our conclusion and learning's from the project:

- It was quite surprising to us, but the inclusion of semantic information for the classification purpose deteriorated the performance of the classifier. The deterioration in the performance may be due to the some weakness in the semantic model. Future work can be extended in this direction of creating a more powerful semantic model, because we still firmly believe that semantic aspect of tags should not be neglected.
- It got proved that focal points do increase the performance of the classifier, though not drastically. The future work can be made in the direction of identifying new type of focal points. It is only our surmise at this stage that the similarity measure between the tags and title of a video can act as a focal point.
- Co-occurrence matrix is a very strong model that can be used for the purpose of classification of large scale complex system that rely on tag like structure for their representation.
- The future work, of course, can also extend in the direction of improving the classification quality. For large scale classification involving 150,000 videos the best achieved accuracy was around 65% and improving this can be a good direction to work on.

# References

[1] G.Begelman, P. Keller, and F.Smadja. Automated Tag Clustering: Improving search and exploration in the tag space. In Proc. of the Collaborative Web Tagging Workshop at WWW'06, 2006.

[2] Special Session on: Neural Networks, Semantic Web Technologies And Multimedia Analysis. 16th International Conference on Artificial Neural Networks, Athens, Greece. 10-14 September 2006.

[3] Clark A., Magic Words: How Language Augments Human Computation in P. Carruthers and J.Boucher (Eds) LANGUAGE AND THOUGHT: INTERDISCIPLINARY THEMES (Cambridge University Press: Cambridge, 1998) P.162-183

[4] Inon Zuckerman, Sarit Kraus, and Jeffrey S. Rosenschein. Using Focal Point Learning to Improve Tactic Coordination in Human-Machine Interactions, IJCAI 2007, Hyderabad, India, January 2007, pages 1563-1568.

[5] Schelling T C. The strategy of conflict. Cambridge, MA: Harvard University Press. (1960 1963) 1980. 309 p. [Center for international Affairs. Harvard University. Cambridge. MA]

[6] http://en.wikipedia.org/wiki/Schelling_point

[7] Hammond, T., Hannay, T., Lund, B., and Scott, J., Social Bookmarking Tools (I): A General Review, D-Lib Magazine 11(4), April 2005.

[8] Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata, UIC Technical Report, 2004.

[9] Janssen, Maarten C.W. Focal Points. New Palgrave Dictionary of Economics and the Law. London, 1998.