Video Classification Using Consensus Learning and CrowdSourcing

HELLO!

I am Aditya Kumar Akash

I am here to present BTP Phase 1 work.

You can find me at adityakumarakash@gmail.com

3

"If a picture is worth a thousand words, a video says it all."

4

Contents

5

MOTIVATION

- Using videos, farmers can stay in touch with the latest technology and success stories of other farmers

- *This motivates to maintain a video library where relevant video suggestion could be made*

- We need a video classification which is domain specific. So we seek to ***crowdsourcing using Game*** as an option for video collection as well as obtaining tags.

- To achieve classification using the game data, we use meta-learners like Consensus Learning

6

Related work

People have tried video classification before -

- Active Learning using Associative Markov Network - Computationally expensive for us

- Video Mule - Based on consensus Learning - But only maintains consensus on Textual and audio-video metadata

We have tried using consensus learning on crowdsourced data of users tagging videos.

7

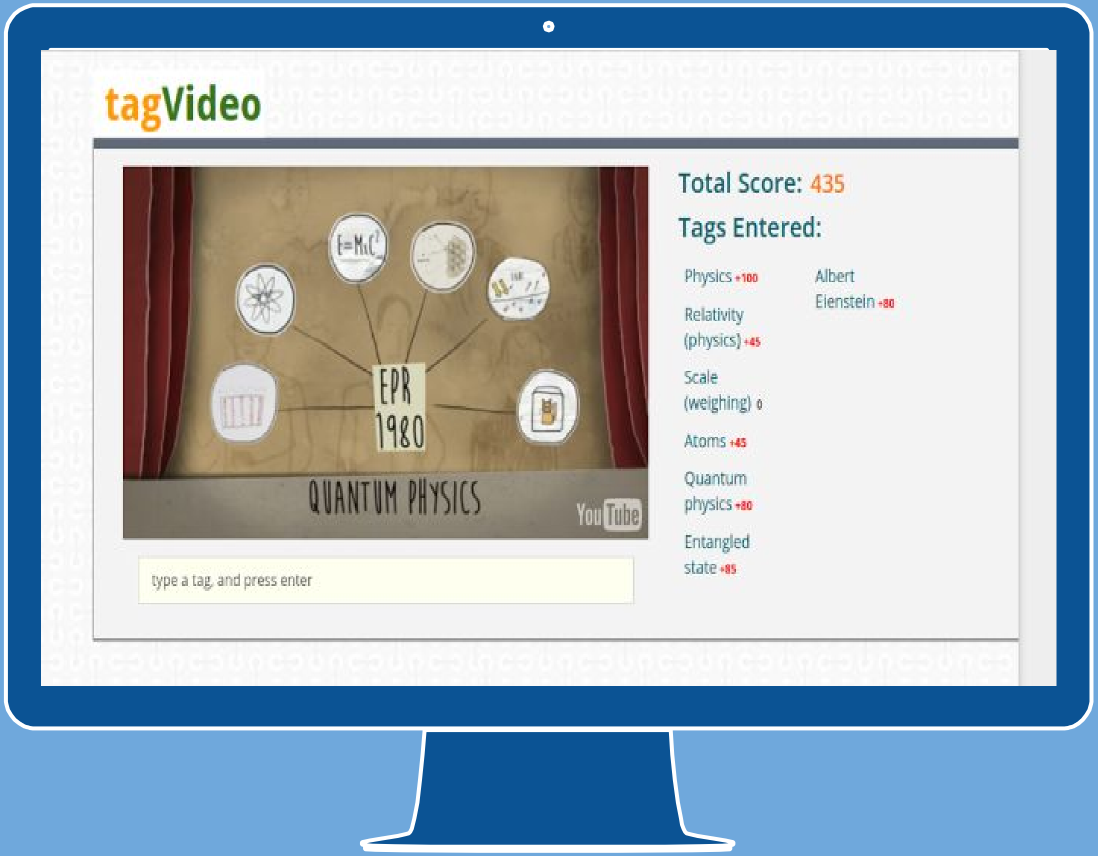
TAG VIDEO GAME

Let’s see the tag video game.

8

tagVideo Game

User watches a given Video. He provides tags for the video which fetch him Game points.



Initial Game made by Ankit and Puja.

9

Scoring of Tags

The score for a tag is based on its relevance to video.

- Initially, the score is based on the relevance score returned by IBM Watson Concept Insights
- As number of taggers increase, a *TF-IDF component of the scoring function* starts gaining more weight

$$w_{watson} = \frac{1}{e^{c_0*(N_v-1)}}$$

N_v is number of taggers

$$tf = \frac{f_v^t}{c_v}$$ and $$idf = 1 + \log\frac{N}{N_t}$$ gives $$tf\_Idf = tf * idf$$

Final Score function looks as -

$$score = s_{watson} * w_{watson} * idf * c_1 + (1 - w_{watson}) * \frac{tf\_Idf}{max_t^v(tf\_Idf)} * c_2 + c_3$$

10

Properties of Scoring Function

We expect following from our scoring function -

- Reflect the consensus of majority
- Not allow users to game the system
- Should give low scores to junk tags in long run
- Tags given for one video should be given high score for a related video

Existing scoring ensures the above except the last point . But contains some problems -

- In term frequency, *division by c_v is not needed.* It get cancelled out in another term.
- To normalize the tf-Idf score, we divide by max tf-Idf over all tags. This causes *influence of Idf_2 on score of tag_1*. This means tagging one video effecting tagging different tags for other video which is *not desirable.*

11

Improving Scoring Function

Introducing changes -

- Changing term frequency $$tf = \frac{f_v^t}{max_v(f_v^t)}$$

- Introducing new normalization $$idf_{max} = 1 + \log N$$

New Scoring Function -

$$score = s_{watson} * w_{watson} * \frac{idf}{idf_{max}} * c_1 + (1 - w_{watson}) * \frac{tf\_Idf}{idf_{max}} * c_2 + c_3$$

The issues with previous function addressed.

12

New Scoring in Action

Two videos with different scoring mechanism are used,
Users tag the video using two tags - relevant and irrelevant,
Table 2.2 with old scoring, Table 2.3 with new scoring

Tag	S1	S2	S3	S4	S5	S6	S7	S8	S9
Iron Man	80	100	100	100	-	-	100	100	-
Metal	0	10	-	-	40	90	-	-	100

Table 2.2: Variation of Tag Score for Video - Will you be iron man?

Tag	S1	S2	S3	S4	S5	S6	S7	S8	S9
Food	70	75	90	100	-	-	95	95	-
Shoe	0	0	-	-	20	60	-	-	75

Table 2.3: Tag Score for Video - Science of sweetness

13

Conclusions -

New scoring shows better effect of change in idf.

New scoring allows for gradual increase in score for junk tags as compared to old scoring which caused spiked increase.

14

Meta Learning

"We do not learn from experience...

we learn from reflecting on experience"

- John Dewey

15

Meta Learners

Assumptions -

1. Users are base learners. They tagging videos is similar to classification
2. Tags are picked from a global set of fixed tags, say concepts of wikipedia

Two popular choice of meta learners-

- Ensemble based - e.g Adaboost
- Consensus based

Adaboost would need manual tagging of videos to train it, It is works on supervised training set. So we focus on consensus learning.

Consensus Based learners

Consensus based learners use predictions from raw data, and not the raw data itself. This makes them most suitable for our use of multilabel classification.

We use “Multilabel consensus maximization for ranking” - MLCM-r from [1].

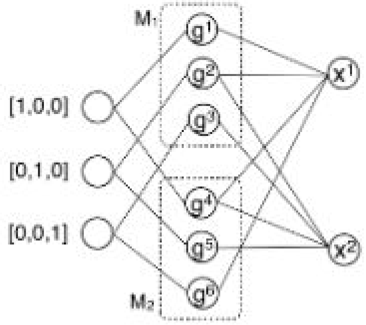
Involves bipartite graph with nodes as -

- *Instance nodes, videos are instances*
- *Group nodes* for each label from each classifiers, *tags are the labels*
- *Connection matrix*, representing *prediction of a label for an instance by a classifier*

17

MLCM-r

Example graph for 2 instances, 3 labels, 2 classifiers



Following probability distributions with each node -

- *__u_i with instance nodes__* - u_{il} : probability of relevance of lth label for ith instance
- *__q_j with group nodes__* - q_{jl} : probability of seeing label l when in node g_j, relates to how closely the labels are in opinion of the classifier to which g_j belong

18

MLCM-r optimization problem

Consensus maximized by solution of following -

$$\min_{U,Q} \sum_{i=1}^{n} \sum_{j=1}^{v} a_{ij} \|\boldsymbol{u}_i - \boldsymbol{q}_j\|^2 + \alpha \sum_{j=1}^{v} \|\boldsymbol{q}_j - \boldsymbol{b}_j\|^2$$

$$u_{ik} \geq 0, \sum_{k=1}^{l} u_{ik} = 1, i = 1, ..., n$$

$$q_{jk} \geq 0, \sum_{k=1}^{l} q_{jk} = 1, j = 1, ..., v$$

1. First term means two connected nodes should have similar probability distribution
2. Second term ensures groups do not deviate much from its initial distribution

19

Solution to MLCM-r

The solution is obtained by block coordinate descent-

$$\boldsymbol{q}_j^t = \frac{\sum_{i=1}^{n} a_{ij}\boldsymbol{u}_i^{t-1} + \alpha \boldsymbol{b}_j}{\sum_{i=1}^{n} a_{ij} + \alpha}$$

$$\boldsymbol{q}_j^t = \frac{\sum_{j=1}^{v} a_{ij}\boldsymbol{q}_j^t}{\sum_{j=1}^{v} a_{ij}}$$

Solution gives u_i which is used to infer the most likely label.

20

Experiments and Observation

Video Id	Tag-Score by TF-IDF	Tag-Score by MLCM-r	Correlation τ
guh7i7tHeZk	• ScienCe,1	• ScienCe,0.49 • Technology,0.26 • Play & Win,0.26	0.67
mtg9p6A6xnY	• Element 13,0.54 • Periodic Table,0.46	• Element 13,0.41 • Periodic Table,0.36 • Chemistry,0.23	1.0
w2Qk-jz_tWc	• Animals,1	• Animals,0.38 • Irrigation,0.15 • Crops,0.15 • Food,0.15 • Farming,0.15	0.4

21

Results

TF-IDF

It is only capable of ordering tags which are actually assigned to the video.

MLCM-r

It provides new tags which are related to previous tags. It is capable of capturing label correlations

The correlation between TF-IDF and MLCM-r Rankings was 0.26, which indicated positive correlation.

22

Consensus Approach Everywhere

We would not be limited by consensus only at level of meta learning, and that forms the core of the project.

Consensus at all levels where it is applicable would be used -

- Consensus amongst people to decide tags
- Consensus with the Knowledge graph in deciding relations between the tags and concepts
- Consensus with the tags obtained from textual metadata and audio video processing

23

Future Work

- MLCM-r does not allow online updates on addition of videos, users or users tagging videos
- *No domain knowledge utilized. Knowledge graph is currently being built by teammates. Need to incorporate it in MLCM-r*
- Would try to separate tags from concepts, to give users freedom of expression in tagging. Would modify the model to use this.
- Would like to improve upon the complexity of algorithm to handle large number of users and videos
- We would like to add trust factor for the users based on how good tagging set is given by them

24

References

[1] Sihong Xie, Xiangnan Kong, Jing Gao, Wei Fan, and Philip S. Yu. Multilabel
consensus classification. In ICDM,2013.

[2] Jing Gao, Feng Liang, Wei Fan, Yizhou Sun, and Jiawei Han. Graph-based consen-
sus maximization among multiple supervised and unsupervised models. In NIPS,
2009

[3] Robert E. Schapire. Explaining Adaboost. pp 37-52, Book Title: Empirical Infer-
ence

THANKS!

Any questions?

You can find me at

- adityakumarakash@gmail.com