Tag Video Weekly Update

Aditya Kumar Akash

IITB

adityakumarakash@gmail.com

September 1, 2015

Overview

Old Scoring

$tf = \frac{f_v^t}{c_v}$, $t = tag, v = video$, f_v^t =freq of t, c_v =distinct t for v

$idf = 1 + \log \frac{N}{N_t}$, N =Total Videos, $N_t = v$ where t is present

$tfIdf = tf * idf$, $max_v(tfIdf)$ =max $tfIdf$ for video v

weight $w_{watson} = \frac{1}{e^{c_0 * (videoTaggers - 1)}}$ given to watson score

score $= s_{watson} * w_{watson} * idf * c_1 + (1 - w_{watson}) * \frac{tfIdf}{max_v(tfIdf)} * c_2 + c_3$

Comments on Old Scoring

- $tf = \frac{f_v^t}{c_v}$, division by count of distinct tags not necessary
- Presence of $\frac{tfIdf}{max_v(tfIdf)}$ is just to get score in 0-1
- $\frac{tfIdf}{max_v(tfIdf)} = \frac{tf_1 * idf_1}{tf_2 * idf_2}$, 2 coresponds to max
- idf_2 can be reduced by tagging other videos with tag 2, which causes steep score increase for some junk tag 1
- In long run $\frac{tfIdf}{max_v(tfIdf)}$ decreases for junk tags

Suggestion on Scoring

- $tf = \frac{f_v^t}{f_v}$, division by max frequent tag, Handles decrease in tf of junk tag over long run
- idf_2 removed and normalization of idf by division by $1 + logN$
- Above tf and normalization ensure scores between 0-1 along with property that junk tag scores decreasing with time
- Implemented this scoring in Game

New Scoring

$tf = \frac{f_v^t}{max(f_v)}$, $t = tag, v = video$, f_v^t =freq of t

$idf = 1 + \log \frac{N}{N_t}$, N =Total Videos, $N_t = v$ where t is present

$tfIdf = tf * idf$, $Idf_{max} = 1 + \log N$

weight $w_{watson} = \frac{1}{e^{c_0 * (videoTaggers - 1)}}$ given to watson score

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

Experiment with old scoring

After **S4** 9 more videos where tagged *Iron Man*, then **S5** onwards proceeded to re-tag this video

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 : Tag Score for Video - Will you be iron man?

'-' denotes this tag was not given

Experiment with new scoring

After **S4** 9 more videos where tagged *Food*, then **S5** onwards proceeded to re-tag this video

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

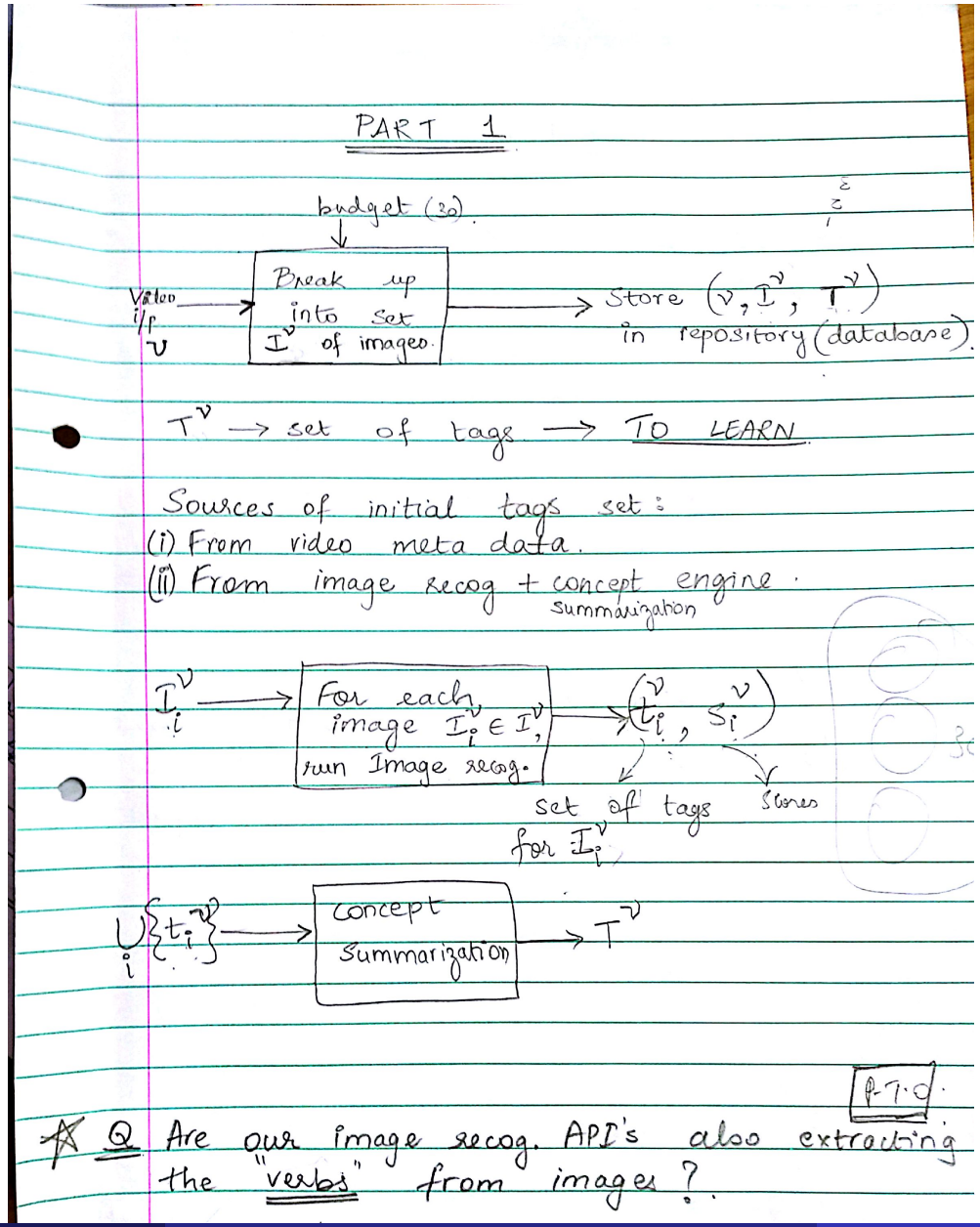
Table : Tag Score for Video - Science of sweetness

'-' denotes this tag was not given

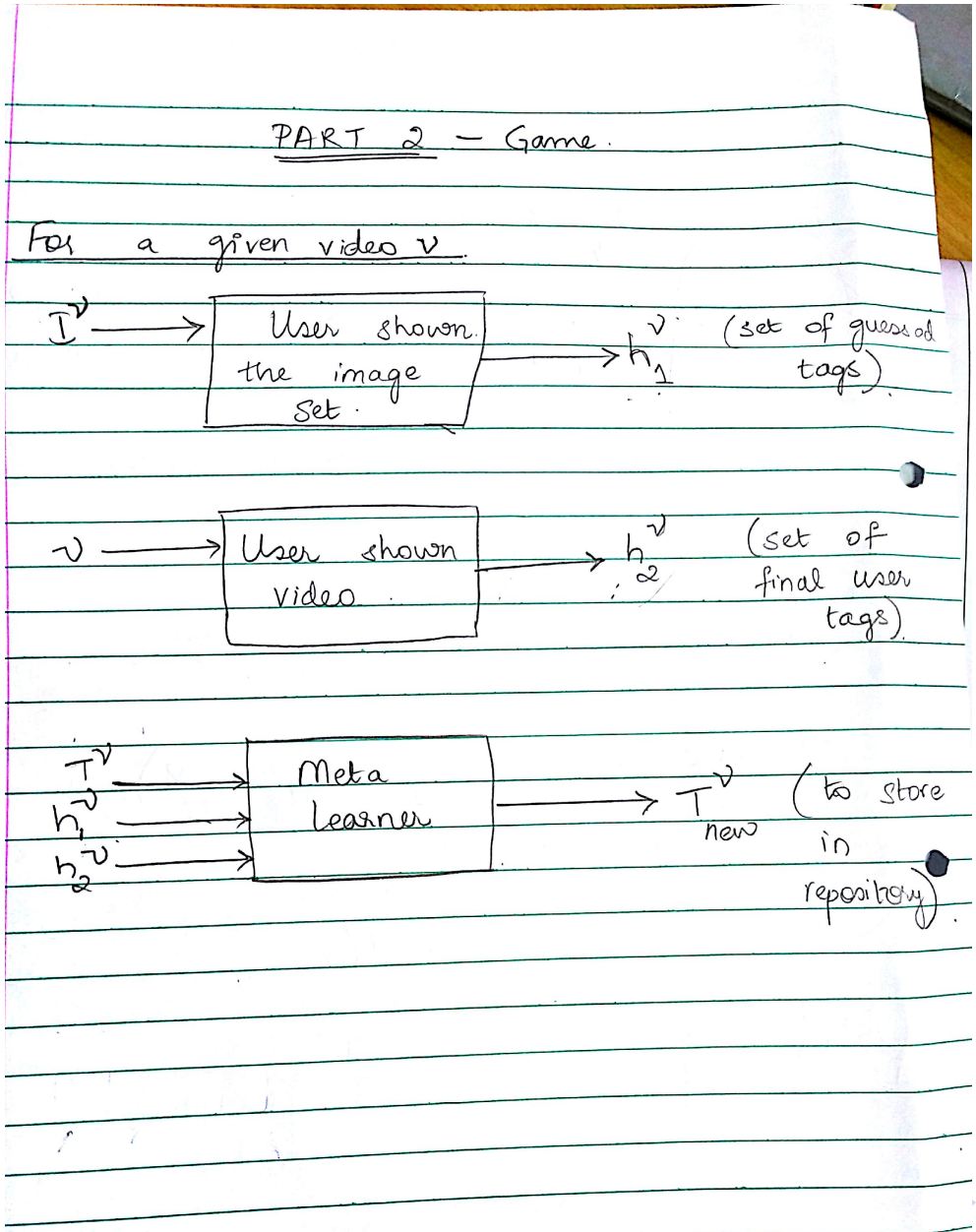
Remarks from experiment

- The old scoring does not reflect clearly idf factor affecting the score
- Old scoring allows unrelated tags have spiked increase in score
- New scoring handles these cases well

Last Meeting Notes



Last Meeting Notes



Meta Learner

- Problem of multi-label, multi-class classification
- **Adaboost.MH** suitable for the task of multi-label, multi-class classification
- *Adaboost.LC* also looks promising, A Boosting Algorithm for Label Covering in Multilabel Problems, http://research.microsoft.com/pubs/78221/AmitDeSi07.pdf
- Other learners - Classifier Chains for Multi-label Classification

Adaboost

Algorithm 1. *AdaBoost [8]*

1. *Initialize the observation weights* $w_i = 1/n$, $i = 1, 2, \ldots, n$.
2. *For* $m = 1$ *to* M*:*
   (a) *Fit a classifier* $T^{(m)}(\boldsymbol{x})$ *to the training data using weights* w_i.
   (b) *Compute*

$$err^{(m)} = \sum_{i=1}^{n} w_i \mathbb{I}\left(c_i \neq T^{(m)}(\boldsymbol{x}_i)\right) / \sum_{i=1}^{n} w_i.$$

   (c) *Compute*

$$\alpha^{(m)} = \log \frac{1 - err^{(m)}}{err^{(m)}}.$$

   (d) *Set*

$$w_i \leftarrow w_i \cdot \exp\left(\alpha^{(m)} \cdot \mathbb{I}\left(c_i \neq T^{(m)}(\boldsymbol{x}_i)\right)\right),$$

   for $i = 1, 2, \ldots, n$.
   (e) *Re-normalize* w_i.
3. *Output*

$$C(\boldsymbol{x}) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \cdot \mathbb{I}(T^{(m)}(\boldsymbol{x}) = k).$$

Problems

- Not clear how to use (video, tagSet) generated when user tags the video in game
- Need to decide when to stop using (video, tagSet) for a given video to avoid over fitting
- Decision of when to include (video, tagSet) to be made
- Deciding on how to score the tags
- Junk tags gets introduced due to game
- ??

Probable Solutions

- Thinking of each user as a weak learner helps - Need to think on how to update learner weights
- Scoring could be based on similarity measure on predicted tagSet using model and tagset given
- Similarity could be weighted sum of similarity and relevance of tags
- Scoring of entire tagSet would help in deciding whether to take (video, tagSet)
- Pre-filter junk tags based on offline dictionary of such tags
- ??

References


Ji Zhu , Hui Zou , Saharon Rosset and Trevor Hastie
Statistics and Its Interface Volume 2 (2009) 349360

The End