# What are the Limits to Time Series Based Recognition of Semantic Concepts?

Peng Wang[1](✉), Lifeng Sun[1], Shiqiang Yang[1], and Alan F. Smeaton[2]

[1] National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, China
{pwang,sunlf,yangshq}@tsinghua.edu.cn

[2] Insight Centre for Data Analytics, Dublin City University, Glasnevin, Dublin 9, Ireland
alan.smeaton@dcu.ie

**Abstract.** Most concept recognition in visual multimedia is based on relatively simple concepts, things which are present in the image or video. These usually correspond to objects which can be identified in images or individual frames. Yet there is also a need to recognise semantic concepts which have a temporal aspect corresponding to activities or complex events. These require some form of time series for recognition and also require some individual concepts to be detected so as to utilise their time-varying features, such as co-occurrence and re-occurrence patterns. While results are reported in the literature of using concept detections which are relatively specific and static, there are research questions which remain unanswered. What concept detection accuracies are satisfactory for time series recognition? Can recognition methods perform equally well across various concept detection performances? What affecting factors need to be taken into account when building concept-based high-level event/activity recognitions? In this paper, we conducted experiments to investigate these questions. Results show that though improving concept detection accuracies can enhance the recognition of time series based concepts, they do not need to be very accurate in order to characterize the dynamic evolution of time series if appropriate methods are used. Experimental results also point out the importance of concept selection for time series recognition, which is usually ignored in the current literature.

**Keywords:** Concept detection · Time series · Activity recognition · Attribute dynamics · Event classification

P. Wang—Work was part-funded by 973 Program under Grant No. 2011CB302206, National Natural Science Foundation of China under Grant No. 61272231, 61472204, 61502264, Beijing Key Laboratory of Networked Multimedia.

A.F. Smeaton—Supported by Science Foundation Ireland under grant number SFI/12/RC/2289.


Q. Tian et al. (Eds.): MMM 2016, Part II, LNCS 9517, pp. 277–289, 2016.
DOI: 10.1007/978-3-319-27674-8_25

## 1 Introduction and Background

The proliferation of online videos and personal media has created multimedia data which require effective indexing and recognition techniques to support flexible retrieval and management. The development of automatic concept based indexing of multimedia has shown the importance of concepts in supporting the understanding of such media. Such concept labels might include the occurrences of scenes, objects, persons, etc. Though various efforts have been tried such as providing large annotated corpora for training, improving the discriminative algorithms, utilising external ontological knowledge, post-processing the indexing results for enhancement, etc., the detection of concepts is still far from being perfect.

While low-level feature-based methods have been shown to be ill-suited for multimedia semantic indexing due to the lack of semantics for user interpretation, high dimensionality, etc., high-level concept attributes are widely employed in the analysis of complex semantics corresponding to things like events and activities. Since such semantic structures can be represented as typical time series, the recognition of events or activities usually involves two components, initially concept detection followed by dynamics-based recognition. This means that initial concept detection results are used as input for modeling the evolution of time-based semantics such as events or activities. This is usually carried out by representing the time series as a sequence of units such as video clips or image frames. After concatenating the results of concept detectors on each unit, time series can then be represented by a temporally-ordered sequence of vectors, as shown in Fig. 1.

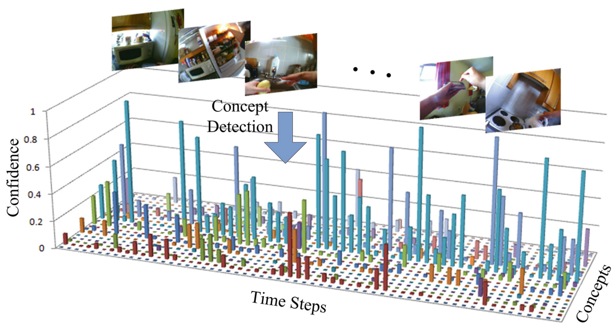


**Fig. 1.** The dynamics of concept attributes quantified by confidences returned by concept detections.

Attribute-based event and activity detection has attracted much research attention. For example, [1] presented an approach to learn a visually grounded storyline model of videos directly from weakly labeled data. In [3], a rule-based method is proposed to generate textual descriptions of video content based on concept classification results. [3] also found that although state-of-the-art concept detections are far from perfect, they still provide useful clues for event classification. In [4], a multimedia event recounting method is proposed based on detected concepts in order to build discriminative event models using a SVM. Similar work is carried out in [5] aiming at video event classification using semantic concept attributes of different categories like action, scene, object, etc. [6] employed an intermediate representation of semantic model vectors trained from SVMs as a basis for detecting complex events, and revealed that this representation outperforms – and is complementary to – other low-level visual descriptors for event modeling. [7] showed that concept-based temporal representations are promising in more complex event recognition. Other efforts as presented in the TRECVid event detection tasks [8,9] also demonstrated promising results for concept-based event detection. Similar work is also carried out using concept detections to characterize everyday activities as reported in [10,11] where activity recognition is built on the basis of concept detection. In [11], detection results are first binarized and then applied to learn temporal dynamics in order to train an activity model. In addition, the correlation of activity and concept detection performances is analyzed in [11].

Though effectiveness is shown using some of the above algorithms in recognizing segments of interest for multimedia time series analytics, there still exists research questions which remain unanswered, due to imperfect concept detection performances. Since these methods use *detection and assemble* schemes which aggregate concept detection results, how concept detection affects the final time series analysis is unclear. Because current research tends to report results based on their own concept detections, whether a proposed event or activity recognition method can adapt to other concept detections is not addressed, such as in cross-domain applications. To overcome these limitations, the following research questions need to be addressed:

- What levels of concept detection accuracy are needed for satisfactory time series analytics? In real-life applications, the pursuit of perfect concept detectors is non-trivial and only a manually annotated groundtruth can be regarded as an Oracle, which is time consuming to obtain. In most cases, however, certain levels of accuracy of concept detection are provided for time-series modeling and classification. In the work we report in this paper we imply that dynamic correlations among imperfect concept detections can still reflect patterns of time series which vary over context.
- How can different recognition methods adapt to varying concept detection accuracies? Most results are reported using individual researchers' own concept detections. It would be of help for researchers to choose methods if the correlations between these methods and the concept detection accuracies, are made clear. More importantly, whether downgraded concept detection accuracies will propagate over time and across models needs to be validated for choosing the right recognition method. As demonstrated later in this paper, the typical methods chosen in our experiments can adapt to varying concept

detection accuracies. This shows why state-of-the-art concept-based event/activity recognitions are feasible and have satisfactory results.
- What factors affect concept-based time series analytics? While most research focuses on temporal modeling of concept occurrence dynamics, a systematic view of concept-based time series analytics involving both performances of concept detection and temporal modeling will provide guidance on this topic. In experiments reported in this paper, we quantify the factors affecting this and we point out that besides time series modeling methods, concept detection and concept selection also need to be considered to improve the performance of the final recognition.

## 2 Experimental Datasets

In our experiments, we take recognition of everyday activity as demonstration of the kind of time series based concept recognition we focus on and we explore the research questions on two datasets, namely lifelogged image streams (Dataset1) and egocentric video collections (Dataset2) respectively.

For lifelog activity recognition, the set of 85 everyday concepts investigated in [12] are used as semantic attributes. Dataset1 includes event samples of 16 activity types collected from 4 SenseCam wearers and consisting of 10,497 SenseCam images [11]. Meanwhile, for egocentric video analysis, we also evaluated various algorithms for recognising activities of daily living (ADL) [2] with 45 underlying semantic concepts. There are a total of 18 activity types and 23,588 frames used in this ADL corpus. To make full use of activity samples in both datasets, we decompose each set of positive samples into 50:50 ratios for training and testing respectively. The two datasets are summarized in Table 1.

**Table 1.** Summary of the lifelog and ADL datasets.

| Datasets | #Types | #Concepts | #Samples | #Frames | Domain |
|---|---|---|---|---|---|
| Dataset1 | 16 | 85 | 500 | 10,497 | Lifelog |
| Dataset2 | 18 | 45 | 624 | 23,588 | ADL |

For both datasets, concept detectors with different accuracy levels were applied by simulating the confidence score outputs from concept detectors as a probabilistic model of two Gaussians. In a state-of-the-art concept detector simulation by Aly *et al.* [13], concept detection performance is controlled by modifying the models' parameters based on manually annotated groundtruth of concept occurrences. These parameters are the mean $\mu_1$ and standard deviation $\sigma_1$ for the positive class, as well as the mean $\mu_0$ and the standard deviation $\sigma_0$ for the negative class. The performance of concept detection can be varied by controlling the intersection of the areas under the two probability density curves by changing the means or the standard deviations of the two classes for a single concept detector.

In addition to the simulation of the confidence distribution for the negative and positive classes as $N(\mu_0, \sigma_0)$ and $N(\mu_1, \sigma_1)$ respectively, we can also obtain the prior probability $P(c)$ for a concept $c$ from the dataset by using manual annotations. Then the sigmoid posterior probability function is needed to fit which has the form of a sigmoid function as [11,13]:

$$P(c|o) = \frac{1}{1 + exp(Ao + B)} \tag{1}$$

where a random confidence score $o$ is drawn from the corresponding normal distribution after parameters $A$ and $B$ are fixed. The posterior probability of the concept is then returned using Eq. (1) for each image frame or video shot.

For each setting of parameters, we executed 20 repeated simulation runs and the averaged concept $AP$ and $MAP$ were both calculated. During the simulation procedure, we fixed the two standard deviations and the mean of the negative class and varied the mean of the positive class $\mu_1$ in the range [0.5...10.0]. Figure 2 shows the improvement in concept $MAP$ for both datasets with increased $\mu_1$, and near-perfect detection performances are achieved when $\mu_1 \geq 4.0$, as segmented by the blue dashed line in Fig. 2.

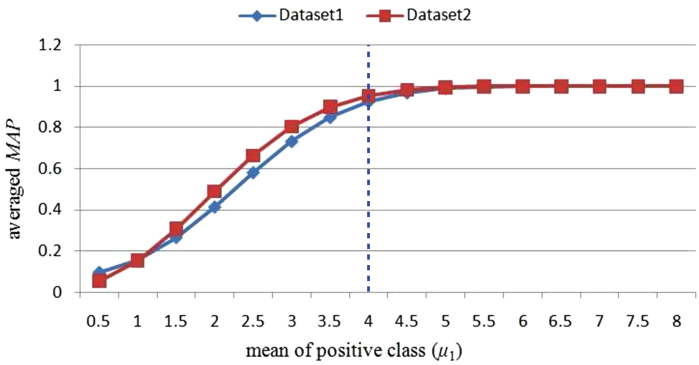


**Fig. 2.** Averaged concept $MAP$ with different $\mu_1$ values for two datasets. (Color figure online)

## 3 Methods

In this experiment, we provide discussions based on the investigation into various time series recognition methods on the above described datasets. To fully exploit the interacting correlations between concept detections and time series recognition, as well as providing investigation into affecting factors on the recognition performance, we validated a selection of methods including temporal and non-temporal features, generative and discriminative models, holistic and pyramid representations, signatures of dynamic systems, etc. The time series recognition methods investigated are summarized in Table 2. Whether they utilise

**Table 2.** Summary of time series recognition methods investigated.

| Methods | Temp? | #Types | #Concepts |
|---|---|---|---|
| Max-Pooling | $\times$ | 10 [14], 15 [19], 25 [7] | 50 [14], 101 [19], 93 [7] |
| Bag-of-Features | $\times$ | 10 [14], 18 [2], 3 [6] | 50 [14], 42 [2], 280 [6] |
| Temporal Pyramids | $\surd$ | 18 [2] | 42 [2] |
| HMM | $\surd$ | 16 [11] | 85 [11] |
| Fisher Vector | $\surd$ | 16 [11], 15 [17] | 85 [11], 60 [17] |
| Dynamic System | $\surd$ | 25 [7], 5 [18] | 93([7], [18]) |

temporal features, the number of event/activity types and the number of concept attributes are all depicted in Table 2.

As we can see from Table 2, the number of categories and concepts used in our datasets (Sect. 2) are both within the prevalent range reported in recent literature. Therefore, the setup of our experiment is close to a realistic implementation and the conclusions should therefore be valid. The details of the different recognition methods we implemented are now outlined.

**Max-Pooling (MP):** As one of the fusion operations for concept detection results, Max-Pooling [14] has been demonstrated to give better performance compared to other fusions for most complex events. In Max-Pooling, the maximum confidence is chosen from all keyframe images (or video subclips) for each concept to generate a fixed-dimensional vector for an event or activity sample. Since by definition the maximum value cannot characterize a temporal evolution of concepts within a time series, this method can be regarded as non-temporal.

**Bag-of-Features (BoF):** Similar to Max-pooling, Bag of Features is a way of aggregating concept detection results by averaging the confidences over time window. Because Bag-of-Features and Max-Pooling reflect the statistical features within the holistic time series, they both ignore the temporal evolution of concept detection results.

**Temporal Pyramids (TP):** Motivated by the spacial pyramid method, the temporal pyramids proposed in [2] approximate temporal correspondence with a temporally-binned model [15]. In this method, the histogram over the whole time series represents the top level while the next level can be represented by concatenating two histograms of temporally segmented sequences. More fine-grained representations can be formalized in the same manner. By applying the multi-scale pyramid to approximate a coarse-to-fine temporal correspondence, this method generates fixed-length features with temporal embeddings.

**Hidden Markov Model (HMM):** A generative method based on HMMs as used in [11] is employed in this representation. HMMs are first trained for each activity class and we concatenate the log-likelihood representations of per-class posteriors into a vector. Assume there are $l$ hidden states in the HMM and each pair of states have a transition probability $a_{ij} = P(s_i|s_j)$. The parameters of the

HMM can be denoted as $\lambda = (A, B, \pi)$, where $A = \{a_{ij}\}$, $\pi = \{\pi_i\}$ stands for the initial state distribution. $b_j(X_t)$ is the distribution of the concept observation $X_t$ at time step $t$ with respect to state $j$.

**Fisher Vector (FV):** The principle of the Fisher kernel is that similar samples should have similar dependence on the generative model, i.e. the gradients of the parameters [16]. Instead of directly using the output of generative models, such as in the HMM method, using a Fisher kernel tries to generate a feature vector which describes how the parameters of the activity model should be modified in order to adapt to different samples. Based on the above formalization of an HMM, $X$ can be characterized as Fisher scores with regard to the parameters $\lambda$, $U_X = \nabla_\lambda log P(X|\lambda)$. Therefore, the Fisher kernel can be formalized as $K(X_i, X_j) = U_{X_i}^T I_F U_{X_j}^T$, where $I_F = E_X(U_X U_X^T)$ denotes the Fisher information matrix.

**Liner Dynamic System (LDS):** As a natural way of modeling temporal interaction within time series, Liner Dynamic Systems [7] can characterize temporal structure with attributes extracted from within a sliding window. The time series can be arranged in a block Hankel matrix $H$ whose elements in a column have the length of sliding window (denoted as $r$) and successive columns are shifted with one time step. According to [7], singular value decomposition of $H \cdot H^T$ has achieved comparable performance to more complex representations. We constructed the feature using the $k$ largest singular values along with their corresponding vectors.

## 4 Results

For recognition based on HMM log-likelihood and on Fisher Vector, generative models were obtained with two-state ergodic HMMs to model the sequence of concept occurrences. Because the confidence vector $X_t$ has continuous values, we employed Gaussian emission distributions $b_j(X_t) = \mathcal{N}(X_t, \mu_i, \sigma_i)$ and $B = \{\mu_i, \sigma_i\}$. Parameters $\mu_i$ and $\sigma_i$ are the mean and covariance matrix of the Gaussian distribution in state $i$ respectively. This setting was applied both in HMM-based and Fisher Vector-based time series recognition. To alleviate the sub-optimal problem of Fisher kernels induced by (nearly) zero gradient representations of a generative model, we employed a model parameter learning as proposed in [20], to train the model so that samples of the same class will have more similar gradients than other classes. The Fisher kernel was then embedded in the SVMs for activity classification. To simplify the computation, we approximated $I_F$ by the identity matrix in the implementation.

In implementing the Hankel matrix $H$ for dynamic system characterization, the length of the sliding window $r$ reflects the temporal influence range which can be regarded as one parameter. Besides $r$, the number $k$ of largest singular values along with their corresponding vectors determines the final dynamic system signature from singular value decomposition of $H \cdot H^T$, and was accepted as the other parameter. Similar to work in [7], we examined performances in the assignment range $r \in \{2, 4, 8\}$ and $k \in \{1, 2, 4\}$. The final time series recognition accuracies were compared and we chose the best performances at $\{r = 2, k = 1\}$ for Dataset1 and $\{r = 4, k = 1\}$ for Dataset2 in the evaluation. As for the temporal pyramids method, two levels of feature histograms were extracted and concatenated to construct the final time series representation. This was chosen empirically without further optimization since two level pyramids have shown to be effective in our datasets which are also employed in [2].

After fixed-length features were extracted using the methods listed in Table 2, the same discriminative classifier SVM was employed to classify activities, with the same way of parameter optimization for fair comparison. The resulting accuracies on two datasets are shown in Figs. 3 and 4, across various concept detection accuracies controlled by the simulation parameter $\mu_1$.

As shown by these two figures, the resulting accuracy curves for different recognition methods have very similar shapes on both datasets, implying similar correlation with concept detections. That is, while improving concept detection accuracies (increasing $\mu_1$), the recognition accuracies of time series are enhanced for all methods, in both datasets. This enhancement is especially significant when the original concept detection accuracies are low, say, at $\mu_1 \leq 2$ in Fig. 3 on the left side of the blue dashed line. When concept detections are accurate enough, say on the right side of the blue dashed line in Fig. 3, time series recognitions converge with relatively stable performances.

In Figs. 3 and 4, the performances of time series recognitions differ across concept detection accuracies. For example, FV, BoF and TP achieve better recognition performances than the others. The advantage of FV is obvious when concept detections have very poor performances at $\mu_1 \leq 1$. This implies that

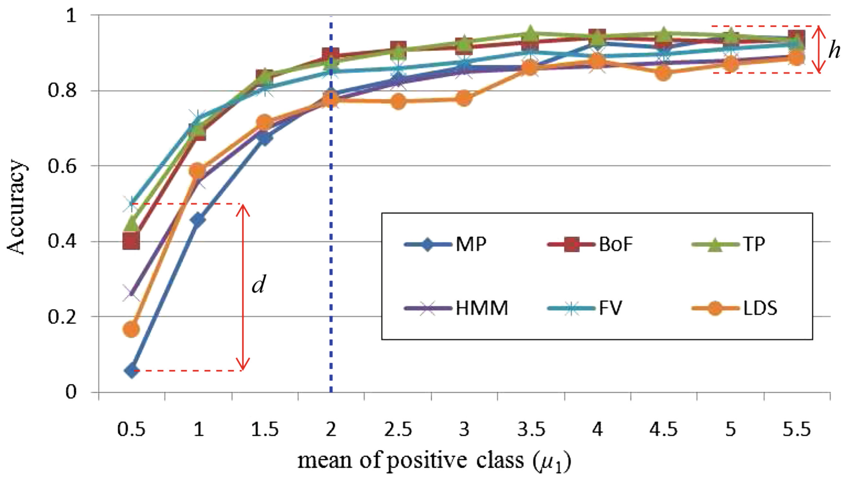


**Fig. 3.** Averaged recognition accuracy on lifelog dataset with different $\mu_1$ values. Accuracy increases significantly when concept detections are low (left of blue dashed line) and converges when concept detections are high (right of blue dashed line) (Color figure online).

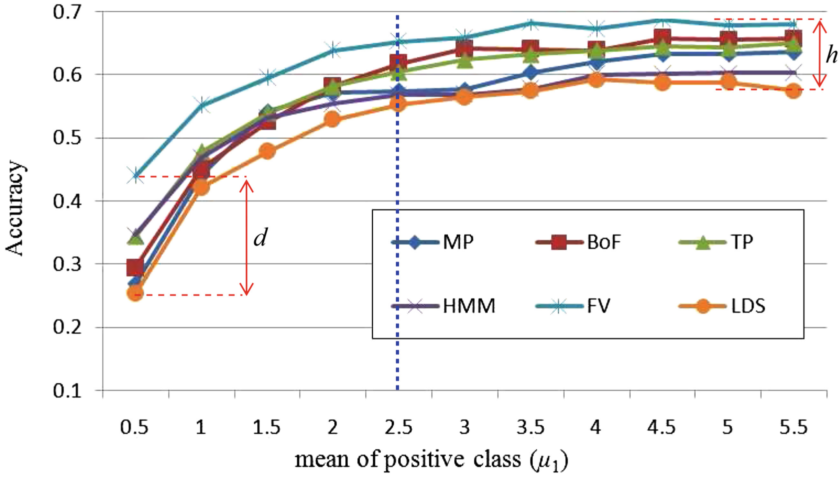


**Fig. 4.** Averaged recognition accuracy on ADL dataset with different $\mu_1$ values. Accuracy increases significantly when concept detections are low (left of blue dashed line) and converges when concept detections are high enough (right of blue dashed line) (Color figure online).

FV can better adapt to less accurate concept annotations. However, when concept annotation burdens are relieved at high $\mu_1$ values, FV is outperformed by BoF and TP. A similar tendency for performance variations at different concept detection levels is also common for other recognition methods in Figs. 3 and 4. From this we suggest that the evaluation of current concept-based time series recognitions on single concept detection performance is limited. More comprehensive assessments on various concept annotations are required to demonstrate the robustness and adaption capabilities of recognition methods.

## 5 Discussion

As previously presented in Sect. 2, concept detection accuracies converge around $\mu_1 = 4.0$ and near-perfect detection performances are achieved after $\mu_1 \geq 4.0$. It is interesting to note that this critical value is lower for recognition accuracy curves in Figs. 3 and 4, in which the best-performing value occurs around $\mu_1 = 2.0$ and $\mu_1 = 2.5$ respectively, as denoted by blue dashed lines. The earlier convergence in Figs. 3 and 4 is good news for time series recognition based on concept detections which are still not accurate. This also manifests why current concept-based event/activity recognition can outperform using low-level descriptors even though state-of-the-art concept detections are far from perfect. The deviation of the dashed line from $\mu_1 = 4$ (Fig. 2) to $\mu_1 = 2$ (Fig. 3) and $\mu_1 = 2.5$ (Fig. 4) implies that an improvement in concept detection accuracies after $\mu_1 = 2$ is of less value in enhancing the time series based recognition due to the adaptive capabilities of recognition methods in the presence of erroneous concept detections.

According to the above experiments, the affecting factors on concept-based time series recognition can be summarized as:

- **Recognition Methods.** As we can see from Figs. 3 and 4, recognition methods play the dominant role in obtaining different performances, as depicted by two distances of $d$ and $h$ for low and high concept detection accuracies respectively. In both figures, $d > h$ holds which means recognition performances differ more significantly at lower concept detections. Therefore, how to better classify high-level events/activities streams based on noisy semantic attributes needs to be addressed. This is especially important for cases where we then build upon the detected concepts such as using them to infer activities, complex events or behaviours. For such applications, there may be further challenges because of the diverse range of usable concepts, or the noisy nature of the multimedia data.
- **Concept Detection.** Accuracies of all recognition methods climb while improving concept detection accuracies (increasing $\mu_1$) in Figs. 3 and 4. Despite recent progress, automatic concept detectors are still far from perfect. The climbing rate is especially significant when original concept detections are not good enough, say, on the left side of the blue dashed lines in Figs. 3 and 4. Given the fact that current state-of-the-art concept detectors are still far from perfect despite recent progress, improving original concept detections for enhanced time series recognition represent another possible research area.
- **Concept Selection.** The benefits of appropriate concept selection can also be demonstrated in our experiments. To eliminate the effects of the above two factors of recognition methods and concept detection, we can focus on the extreme performances on two datasets, which are 0.95 and 0.69 in Figs. 3 and 4 respectively. In addition to the inherent nature of two datasets, the sophisticated selections of more appropriate concepts also lead to a difference in performances. For the lifelog dataset as shown in Fig. 3, topic-related semantic concept selections are investigated in [12] to choose concepts in terms of user experiments and semantic networks.

As shown in Table 1, our lifelog dataset uses more concepts than the ADL dataset (85 versus 45) to characterize activity time series. To further validate the effect of concept selections on the finial recognition performances, we carried out further experiments by randomly selecting $n$ concepts ($n \leq 85$) from the lifelog dataset and performing time series recognition based on the selected concept subsets. Results for $n = \{20, 40, 60, 85\}$ are shown in Figs. 5 and 6 for two levels of concept detection accuracies ($\mu_1 = \{2.0, 5.5\}$).

Taking three recognition methods for example, Figs. 5 and 6 show the same trends in accuracies when increasing the number of selected concepts. When more appropriate concepts are utilised to characterize the dynamic evolution of time series based concepts, their recognition performances are enhanced accordingly. This is more obvious when the original concept detections are less satisfactory in Fig. 5 (@$\mu_1 = 2.0$). In this case, when concept annotations are more noisy, introducing more concepts can counteract such noise and enhance recognition performances. For example, increasing the number of selected concepts from 20

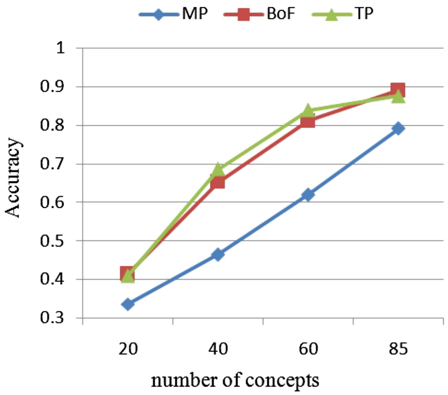


**Fig. 5.** Performance comparison for randomly selected concepts at lower concept detection accuracies ($\mu_1 = 2.0$).

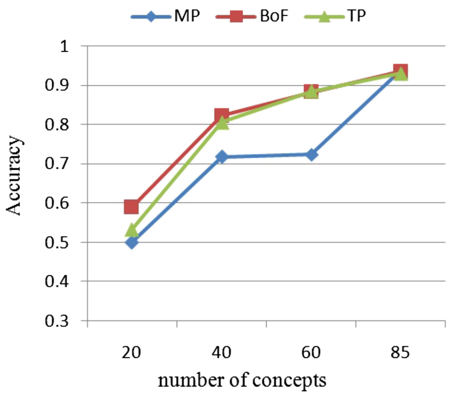


**Fig. 6.** Performance comparison for randomly selected concepts at higher concept detection accuracies ($\mu_1 = 5.5$).

to 40 leads to nearly 0.3 accuracy improvement in Fig. 5, compared to 0.2 in Fig. 6. These results support our hypothesis that concept selection is another factor affecting recognition performance of time series based concepts.

From Figs. 5 and 6, we also note that the performance enhancement is less significant when the number of selected concepts are high, say, when $n \geq 60$. While increasing values of $n$, the slopes of performance curves of BoF and TP decline gradually in both figures. This implies that introducing more redundant concepts is of less value to characterize the time series since these concepts are not independent of each other. This can be captured through various types of correlations among concepts including co-occurrence patterns and ontological relationships. In other words, the semantic space spanned by concepts in the lexicon can be projected to a more compact space with lower ranks since concepts are not independent. This characteristic can also be utilised to enhance concept detection accuracies [22], classify time series [21], etc.

## 6 Conclusions

Though acceptable results can be achieved using state-of-the-art concept detection methods for narrow domains and for concepts for which there exists enough annotated training data, concept detectors are still far from perfect especially for those related to activities, events or behaviour which have a temporal aspect and how time series based recognition methods for such concepts interact with noisy semantic attributes is unclear. To validate the adaption capability of event/activity recognitions to concept detections, we carried out experiments on two datasets with various concept detection accuracies using typical time series based recognition methods. Results show that concept-based time series recognition is feasible when built on the premise of noisy underlying concept detections. In the experiment, we also explored the nature of the affecting factors on time series analytics. Besides recognition methods which are the focus of current research, concept detection and concept selection are also pointed out to have direct influence on analytics performances. This work can provide an analysis framework and guidance for time series recognition based on concept attributes.

## References

1. Gupta, A., Srinivasan, P., Shi, J., Davis, L.: Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos. In: CVPR 2009, pp. 2012–2019, June 2009
2. Pirsiavash, H., Ramanan, D.: Detecting activities of daily living in first-person camera views. In: CVPR 2012, pp. 2847–2854, June 2012
3. Tan, C.C., Jiang, Y.-G., Ngo, C.-W.: Towards textually describing complex video contents with audio-visual concept classifiers. In: ACM Multimedia 2011, pp. 655–658 (2011)
4. Yu, Q., Liu, J., Cheng, H., Divakaran, A., Sawhney, H.: Multimedia event recounting with concept based representation. In: ACM Multimedia 2012, pp. 1073–1076 (2012)
5. Liu, J., Yu, Q., Javed, O., Ali, S., Tamrakar, A., Divakaran, A., Cheng, H., Sawhney, H.: Video event recognition using concept attributes. In: IEEE Winter Conference on Applications of Computer Vision, pp. 339–346 (2013)
6. Merler, M., Huang, B., Xie, L., Hua, G., Natsev, A.: Semantic model vectors for complex video event recognition. IEEE Trans. Multimedia **14**(1), 88–101 (2012)
7. Bhattacharya, S., Kalayeh, M., Sukthankar, R., Shah, M.: Recognition of complex events exploiting temporal dynamics between underlying concepts. In: CVPR 2014, June 2014
8. Hill, M., Hua, G., Huang, B., Merler, M., Natsev, A., Smith, J.R., Xie, L., Ouyang, H., Zhou, M.: IBM research TRECVid-2010 video copy detection and multimedia event detection system (2010)
9. Cheng, H., Liu, J., Ali, S., Javed, O., Yu, Q., Tamrakar, A., Divakaran, A., Sawhney, H.S., Manmatha, R., Allan, J., Hauptmann, A., Shah, M., Bhattacharya, S., Dehghan, A., Friedland, G., Elizalde, B.M., Darrell, T., Witbrock, M., Curtis, J.: SRI-Sarnoff AURORA system at TRECVid 2012: Multimedia event detection and recounting. In: NIST TRECVID Workshop, December 2012
10. Doherty, A.R., Caprani, N., O'Conaire, C., Kalnikaite, V., Gurrin, C., O'Connor, N.E., Smeaton, A.F.: Passively recognising human activities through lifelogging. Comput. Hum. Behav. **27**(5), 1948–1958 (2011)
11. Wang, P., Smeaton, A.F.: Using visual lifelogs to automatically characterise everyday activities. Inf. Sci. **230**, 147–161 (2013)
12. Wang, P., Smeaton, A.: Semantics-based selection of everyday concepts in visual lifelogging. Int. J. Multimedia Inf. Retrieval **1**(2), 87–101 (2012)
13. Aly, R., Hiemstra, D., de Jong, F., Apers, P.: Simulating the future of concept-based video retrieval under improved detector performance. Multimedia Tools Appl. **60**, 1–29 (2011)
14. Guo, J., Scott, D., Hopfgartner, F., Gurrin, C.: Detecting complex events in user-generated video using concept classifiers. In: CBMI 2012, pp. 1–6, June 2012

15. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: CVPR 2008, pp. 1–8, June 2008
16. Jaakkola, T.S., Haussler, D.: Exploiting generative models in discriminative classifiers. In: NIPS 1999, pp. 487–493 (1999)
17. Sun, C., Nevatia, R.: ACTIVE: activity concept transitions in video event classification. In: ICCV 2013, pp. 913–920, December 2013
18. Li, W., Yu, Q., Sawhney, H., Vasconcelos, N.: Recognizing activities via bag of words for attribute dynamics. In: CVPR 2013, pp. 2587–2594, June 2013
19. Liu, J., Yu, Q., Javed, O., Ali, S., Tamrakar, A., Divakaran, A., Cheng, H., Sawhney, H.: Video event recognition using concept attributes. In: WACV 2013, pp. 339–346 (2013)
20. van der Maaten, L.: Learning discriminative fisher kernels. In: ICML 2011, pp. 217–224 (2011)
21. Assari, S., Zamir, A., Shah, M.: Video classification using semantic concept co-occurrences. In: CVPR 2014, pp. 2529–2536, June 2014
22. Wang, P., Smeaton, A.F., Gurrin, C.: Factorizing time-aware multi-way tensors for enhancing semantic wearable sensing. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015, Part I. LNCS, vol. 8935, pp. 571–582. Springer, Heidelberg (2015)