

УДК 004.932.72'1

Н. В. Коваленко

АВТОМАТИЧЕСКОЕ АННОТИРОВАНИЕ ВИДЕО НА ОСНОВЕ ОНТОЛОГИЙ И БАЙЕСОВСКОЙ СЕТИ

Аннотация. В данной статье предложена система автоматического аннотирования видеoinформации, построенная на основе онтологии предметной области, содержащей словарь, используемый для аннотирования; а также на основе Байесовской сети, используемой для распознавания событий и сценариев в видеофрагментах. Тестирование системы было проведено на наборе тестовых видеопоследовательностей CAVIAR.

Ключевые слова: аннотирование видео, онтологии, байесовская сеть, распознавание событий, семантический поиск, обработка видео, сегментация движения

N. Kovalenko

ONTOLOGY AND BAYESIAN NETWORK BASED AUTOMATIC VIDEO ANNOTATION

Abstract. In this paper we present an automatic video annotation system, based on a domain ontology, which contains the lexicon used for the annotation process. A Bayesian Network is also used to recognize events and scenarios in the video fragments. An experimental testing was conducted using the CAVIAR test video dataset.

Keywords: video annotation, domain ontology, Bayesian Network, event recognition, semantic search, video processing, movement segmentation

М. В. Коваленко

АВТОМАТИЧНЕ АНОТУВАННЯ ВІДЕО НА БАЗІ ОНТОЛОГІЙ ТА БАЙЄСІВСЬКОЇ МЕРЕЖІ

Анотація. У даній статті запропонована система автоматичного анотування відеоматеріалу, побудована на основі онтології предметної області, що містить словник, що використовується для анотування; а також на основі Байєсівської мережі, що використовується для розпізнавання подій і сценаріїв у відеофрагментах. Попереднє тестування системи було проведено на наборі тестових відеопослідовностей.

Ключові слова: анотування відео, онтології, Байєсівська мережа, розпізнавання подій, семантичний пошук, обробка відео, сегментування руху

Введение

В последнее время наблюдается значительный рост интереса к информационным системам и технологиям, основанным на обработке видеoinформации. К таким системам относятся, прежде всего, системы видеонаблюдения, имеющие самый широкий спектр применений: от предотвращения преступлений до заботы о людях преклонного возраста.

Особый интерес представляют системы аннотирования видео, т.е. краткого семантического описания происходящего на видео. Следует отметить, что эффективность работы таких систем во многом зависит от качества решения задачи анализа и распознавания действий и поведения человека.

Однако из-за сложности такого анализа в большинстве существующих работ [1] рассматривается только распознавание простых действий человека или его деятельности на однородном фоне. Несмотря на многочис-

ленные исследования, которые проводятся во всем мире в этом направлении, автоматическое аннотирование видеoinформации по-прежнему остаётся достаточно сложной для решения задачей.

Анализ и постановка проблемы

Основные проблемы при автоматизированном аннотировании видеoinформации связаны с двумя группами факторов. К первой относится необходимость обеспечения логической постоянности аннотаций в терминах используемого словаря и способа его использования; необходимость учета предметной области и прагматической цели анализа. Ко второй – слабая формализация информации, а также низкая достоверность выделения значимых объектов из-за влияния освещенности, ракурса, масштаба и др.

Среди наиболее эффективных подходов, направленных на решения задачи аннотирования видеoinформации, с учетом первой группы следует выделить онтологический, а второй – вероятностный.

© Коваленко Н.В., 2015

Онтологии считаются подходящим инструментом для преодоления семантического барьера (semantic gap) при анализе мультимедийной информации и перспективным средством для поиска и извлечения интересующего контента с учетом большого количества обрабатываемого видео [2]. Вероятностные модели, например, на основе Байесовской сети позволяют учесть вероятностный характер и неопределенность, извлекаемых из видеопотока данных [2].

В рамках данной работы предлагается подход, сочетающий преимущества онтологии для описания предметной области и вероятностной Байесовской сети (БС) для учета неопределенности данных.

Онтология предварительно строится вручную и содержит иерархически структурированный словарь семантических видеоконцептов и их основных свойств. Полученной онтологии ставится в соответствие Байесовская сеть, что позволяет учесть особенности движений человека: их вероятностный характер и продолжительность во времени [2].

Однако при реализации такого подхода возникает проблема определения рациональной структуры БС.

Получение структуры графа этих сетей из имеющегося набора данных является NP-сложной задачей, время решения которой при росте количества данных даже наиболее эффективными методами, возрастает экспоненциально. Рассмотрим методы определения структуры Байесовской сети.

Определение структуры БС

Определение структуры БС состоит из двух этапов: первоначальное ее определение и обучение (получение рациональной структуры).

При определении структуры БС основываются на выделении переменных и извлечении зависимостей и независимостей в предметной области, используя знания экспертов, что не гарантирует построения рациональной структуры Байесовской сети.

Альтернативным подходом является непосредственное использование экспериментальных данных, что предоставляет определенные преимущества, особенно при крупных предметных областях или недоступности надежного экспертного знания.

Обучение БС обычно осуществляется с использованием методов, основанных на ограничениях или методах поиска и оценивания [3]. Первые используют статистические тесты, такие как хи-квадрат и взаимная информация, для поиска отношений условной независимости между переменными и использует эти отношения и основанные на причинных связях правила для построения Байесовских сетей.

Методы поиска и оценивания состоят из двух элементов: процедуры поиска структуры сети и оценивания каждой найденной структуры при помощи метрики или оценочной функции. Для структур небольшого размера возможно использование подхода поиска грубой силы (brute-force), с рассмотрением всех возможных комбинаций [4]. Поиск также может проводиться, начиная со случайной или заранее выбранной точки, с рассмотрением всех соседних структур путем добавления, удаления или обращения одной арки на каждой итерации. К таким методам относится K2 – алгоритм жадного поиска, который начинается с пустой сети, но с начальным упорядочиванием узлов. Байесовская сеть составляется итеративно, путём добавления направленных дуг к текущему узлу от того родительского узла, добавление которого максимально увеличивает оценку результирующей структуры графа. Процесс оканчивается в том случае, если ни одно из возможных добавлений не увеличивает оценку.

Наиболее эффективные методы обучения структуры БС основаны на динамическом программировании [5], и требуют затрат времени и памяти пропорциональных $n \cdot 2^n$, где n – число переменных. Такая сложность не позволяет использовать эти методы уже при нескольких десятках переменных, из-за вычислительной сложности.

Модифицированный метод определения структуры Байесовской сети

В данной работе предлагается модифицированный метод определения структуры Байесовской сети, основанный на методе ветвей и границ [6] с использованием структурных ограничений для уменьшения пространства поиска, и онтологии предметной области как набора переменных и размеченных видеоданных.

Алгоритм ветвей и границ хорошо подходит для построения структуры БС, при сравнительно большом количестве переменных, а также гарантирует нахождение оптимального решения при остановке своей работы.

Предлагаемый метод состоит из трех основных этапов.

1. Производится построение онтологии предметной области, что осуществляется экспертом вручную. Онтология содержит иерархически структурированный набор видео-концептов, служащих для семантического описания содержимого видео. В данной работе онтология для предметной области распознавания событий и человеческих действий сформирована на основе тестового набора данных CAVIAR [7], содержащего набор простых сценариев, с участием отдельных людей и группы лиц.

2. Формируются структурные ограничения, которые используются для уменьшения пространства возможных структур Байесовской сети. Ограничения учитываются при построении онтологии предметной области в виде онтологических правил. Ограничения позволяют уменьшить пространство поиска и представляют собой указания, какие связи должны или не должны быть включены в БС. Предлагается следующий набор правил для формирования структурных ограничений:

– $indegree(X_j, k, op)$ – узел X_j должен иметь менее чем (если $op = lt$) или ровно (если $op = eq$) k родителей, здесь $op \in \{lt, eq\}$ и k – целое число;

– $arc(X_i, X_j)$ указывает, что узел X_i должен быть родителем узла X_j ;

– операторы *или* (\square) и *не* (\neg) используются для формирования правил. Оператор *и* (\square) не используется явно, так как предполагается, что каждое ограничение находится в дизъюнктивной нормальной форме.

3. Обучение (поиск рациональной структуры) Байесовской сети с использованием алгоритма ветвей и границ. Под рациональной структурой понимается такая структура, которая может быть представлена ориентированным ациклическим графом G , и которая имеет наивысший показатель оценочного критерия $s_D(G)$. В данной работе в качестве оценочной функции используется метрика Байеса-Дирихле [8]:

$$s_D(G) = \log(p(G) \cdot \int p(D|G, \theta) \cdot p(\theta|G) d\theta), (1)$$

где $\theta = (\theta_{ijk})_{\forall ijk}$ – вектор параметров, таких что $\theta_{ijk} = p(x_{ijk} | \pi_i)$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, r_{\Pi}\}$, $k \in \{1, \dots, r_i\}$ и $\pi_{ij} \in \Omega_{\Pi}$.

Здесь Π_i – набор родителей узла X_i , $\Omega_{X_i} = \{x_{i1}, x_{i2}, \dots, x_{ir}\}$ – это пространство состояний узла, а $r_i = |\Omega_{X_i}| \geq 2$ – число (конечное) категорий узла X_i . Наконец, $D = \{D_1, \dots, D_N\}$ – полный набор данных, где $D_u \doteq x_u \in \Omega_x$.

Описание алгоритма обучения БС

Алгоритм ветвей и границ производит поиск среди набора всех возможных графов, которые помещены в очередь с приоритетом. Для каждого графа осуществляется проверка, является ли он ориентированным ациклическим графом. В положительном случае для него вычисляется оценочная функция. В противном случае, граф разбивается на несколько подграфов путем случайного удаления или добавления новых дуг. Перед началом работы алгоритма необходимо инициализировать следующие переменные:

– H – изначально пустая матрица, содержащая для каждой возможной дуги между узлами пометку, должна ли эта дуга существовать, или же еще существование запрещено.

– Q – приоритетная очередь триад (G, H, s) , упорядоченная по s . Каждый граф, содержащийся в очереди, сформирован как каждое возможное сочетание узлов и дуг, с учетом структурных ограничений.

– (G_{best}, s_{best}) – наилучшая на данный момент структура сети и её оценка.

– $(G_{cur}, H_{cur}, s_{cur})$ – триада, содержащее текущий граф, его матрицу H и значение оценочной функции.

Предлагается следующий алгоритм обучения:

1. Поместить первый или последний элемент из очереди Q в $(G_{cur}, H_{cur}, s_{cur})$ для четной или нечетной итерации соответственно. Если $s_{cur} \leq s_{best}$, отбросить текущий элемент и начать цикл заново.

2. Если G_{cur} является ориентированным ациклическим графом, обновить (G_{best}, s_{best}) ,

отбросить текущий элемент и начать цикл заново.

3. Взять цикл из G_{cur} и для каждой дуги из цикла попытаться удалить её или изменить её направление, учитывая структурные ограничения, после чего пересчитать для полученного подграфа значение оценочной функции и добавить текущую триаду $(G_{cur}, H_{cur}, S_{cur})$ в Q .

Алгоритм завершает свою работу, когда очередь оказывается пустой.

В результате работы алгоритма обучения получается рациональная структура Байесовской сети с максимальным показателем оценочной функции. Предложенный алгоритм позволил уменьшить затраты временных и вычислительных ресурсов за счет использования структурных ограничений.

Экспериментальная проверка

Для тестирования предложенного подхода разработана информационная технологии автоматического аннотирования видео. В качестве входных видеоданных использована выборка CAVIAR. Извлекаемые в результате низкоуровневой обработки видео параметры объектов являются входными переменными БС.

Проведенное сравнение эффективности (показатели точности и полноты) разработанной информационной технологии для аннотирования видеофрагментов с другими существующими системами, в частности, системой, использующей только онтологические правила [9], и системой, использующей онтологию и SVM-классификатор [10] показало ее преимущество. Для сравнения использовался набор видеороликов «Проход мимо» из базы CAVIAR. Соответствующие количественные оценки приведены в таблице 1.

1. Сравнительная оценка методов

Метод	Точность	Полнота
Предложенный метод	94,3 %	92,7 %
Первый метод [9]	91,2 %	89,1 %
Второй метод [10]	89,6 %	78,9 %

Заключение

В данной работе разработан метод автоматического аннотирования видеoinформации на основе оригинального подхода, который сочетает использование онтологии пред-

метной области как словаря видео-концептов, содержащего основной набор событий и сценариев, который используется при аннотировании видеопоследовательностей, и вероятностной Байесовской сети, построенной с учетом разработанной онтологии.

Предложенный подход позволяет учесть особенности движений человека: их вероятностный характер и продолжительность во времени. Это позволило повысить достоверность аннотирования движений человека. В работе предложен модифицированный метод получения рациональной структуры БС на основе алгоритма ветвей и границ учета структурных ограничений, что позволило уменьшить пространство поиска и повысить быстродействие.

Список использованной литературы

- Zhu F., Shao L., and Lin M., (2013), Multi-view Action Recognition Using Local Similarity Random Forest and Sensor Fusion, *Pattern Recognition Letters*, Vol. 34, pp. 20 – 24.
- Коваленко Н. В. Семантико-вероятностная сеть для определения контекста видеопотока в системах видеонаблюдения [Текст] / Н. В. Коваленко, С. Г. Антощук // Восточно-Европейский журнал передовых технологий. – Харьков :– 2013. – Вып. 4/9 (64). – С. 15 – 18.
- Yehezkel R., and Lerner B., (2009), Bayesian Network Structure Learning by Recursive Autonomy Identification, *Journal of Machine Learning Research*, Vol. 10, pp. 1527 – 1570.
- Fenz S., Min T. A., and Hudec M., (2009), Ontology-Based Generation of Bayesian Networks, *International Conference on Complex, Intelligent and Software Intensive Systems*, pp. 712 – 717.
- Parviainen P., and Koivisto M., (2009), Exact Structure Discovery in Bayesian Networks with Less Space, *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, pp. 436 – 443.
- De Campos C.P., Zeng Z., and Ji Q., (2009), Structure Learning of Bayesian Networks Using Constraint, *Proceedings of the 26th International Conference on Machine Learning*, pp. 113 – 120.

7. CAVIAR. Context Aware Vision using Image-based Active Recognition, (2015), url: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR>.

8. Cooper G.F., and Herskovits E.A., (1992), Bayesian Method for the Induction of Probabilistic Networks from Data, *Machine Learning*, Vol. 9, pp. 309 – 347.

9. Kazi Tani M.Y., Lablack A., Ghomari A., and Bilasco I.M., (2014), Events Detection Using a Video-surveillance Ontology and a Rule-based Approach, *ECCV*, pp. 299 – 308.

10. Khurana K., and Chandak M.B., (2013), Video Annotation Methodology Based on Ontology for Transportation Domain, *International Journal of Advanced Research in Computer Science, and Software Engineering*, Vol. 3, Iss. 6, pp. 540 – 548.

25th Conference on Uncertainty in Artificial Intelligence, pp. 436 – 443.

6. De Campos C.P., Zeng Z., and Ji Q., (2009), Structure Learning of Bayesian Networks Using Constraint, *Proceedings of the 26th International Conference on Machine Learning*, pp. 113 – 120.

7. CAVIAR. Context aware Vision Using Image-based Active Recognition, (2015), url: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR>.

8. Cooper G.F., and Herskovits E.A., (1992), Bayesian Method for the Induction of Probabilistic Networks from Data, *Machine Learning*, Vol. 9, pp. 309 – 347.

9. Kazi-Tani Y., Lablack A., Ghomari A., and Bilasco M., (2014), Events Detection Using a Video-surveillance Ontology and a Rule-based Approach, *ECCV*, pp. 299 – 308.

10. Khurana K., and Chandak M.B., (2013), Video Annotation Methodology Based on Ontology for Transportation Domain, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, Iss. 6, pp. 540 – 548.

Получено 02.04.2015

References

1. Zhu F., Shao L., and Lin M., (2013), Multi-view Action Recognition Using Local Similarity Random Forest and Sensor Fusion, *Pattern Recognition Letters*, Vol. 34, pp. 20 – 24.

2. Kovalenko N.V., and Antoshchuk S.G. Semantiko-Veroyatnostnaya Set' dlya Opre-deleniya Konteksta Videopotoka v Sistemah Videonabl'udeniya [Semantic-Probabilistic Network for Video Context Recognition in Video Systems], (2013), Kharkov, Ukraine, *Eastern-European Journal of Enterprise Technologies*, Vol. 4, No. 4 (64), pp. 15 – 18.

3. Yehezkel R., and Lerner B., (2009), Bayesian Network Structure Learning by Recursive Autonomy Identification, *Journal of Machine Learning Research*, Vol. 10, pp. 1527 – 1570.

4. Fenz S., Min T.A., and Hudec M., (2009), Ontology-Based Generation of Bayesian Networks, *International Conference on Complex, Intelligent and Software Intensive Systems*, pp. 712 – 717.

5. Parviainen P., and Koivisto M., (2009), Exact Structure Discovery in Bayesian Networks with Less Space, *Proceedings of the*



Коваленко
Никита Владиславович,
аспирант каф. инфор-
мационных систем
Одесского нац. поли-
техн. ун-та,
м/т: +3(050)4951713.
E-mail:
kov.nikit@gmail.com