8. *Three-party* quantum secure direct communication based on GHZ states / Jin X.-R., Ji X., Zhang Y.-Q. et al // Physics Letters A. – 2006. – V. 354, № 1-2. – P. 67 – 70.

9. *Василиу Е.В., Василиу Л.Н.* Пинг – понг протокол с трех– и четырехкубитными состояниями Гринбергера – Хорна – Цайлингера // Труды Одесского политехнического университета. – 2008. – Вып. 1(29). – С. 171 – 176.

10. *Василиу Е.В.* Безопасность пинг – понг протокола квантовой связи для передачи текстовых сообщений // Наукові праці ОНАЗ ім. О.С. Попова. – 2007. – № 2. – С. 36 – 44.

11. *Василиу Е.В.* Анализ безопасности пинг – понг протокола с квантовым плотным кодированием // Наукові праці ОНАЗ ім. О.С. Попова. – 2007. – № 1. – С. 32 – 38.

12. *Василиу Е.В.* Анализ атаки на пинг – понг протокол с триплетами Гринбергера – Хорна - Цайлингера // Наукові праці ОНАЗ ім. О.С. Попова. – 2008. – № 1. – С. 15 – 24.

13. *Василиу Е.В., Василиу Л.Н.* Оценка количества информации, попадающей к злоумышленнику, для трех вариантов пинг-понг протокола квантовой безопасной связи // Материали за IV международна научна практична конференция «Научно пространство на Европа – 2008», 15 – 30 апреля 2008 г. – София, «Бял ГРАД-БГ» ООД. – Т. 29. – С. 34 – 40.

14. *Overbey, J., Traves, W., Wojdylo, J.* On the keyspace of the Hill cipher // Cryptologia.– 2005. – V. 29, № 1. – P. 59 – 72.

15. *Levine J., Nahikian H.M.* On the Construction of Involutory Matrices // American Mathematical Monthly. – 1962. – V. 69, № 4. – P. 267 – 272.

16. *Фергюсон Н., Шнайер Б.* Практическая криптография: Пер. с англ. – М.: Изд. дом "Вильямс", 2005. – 424 с.

17. *Experimental* demonstration of a hyper – entangled ten – qubit Schrodinger cat state / Gao W.-B, Lu C.-Y., Yao X.-C. et al // [Электронный ресурс] http://arxiv.org/abs/0809.4277.

Поступила 22.12.2008

УДК 519.676:681.51 **Блавацкая Н.Н.**

ОПТИМИЗАЦИЯ АДАПТИВНОСТИ МЕТОДОВ СЖАТИЯ ИНФОРМАЦИИ

Наилучших на сегодняшний день результатов по степени сжатия информации позволяют достичь методы вероятностного моделирования информационного источника. Преимущество методов данной группы особенно заметно при сжатии текстов, где они являются признанными лидерами.

Однако в последние несколько лет лидирующее положение методов группы моделирования источника с учетом контекстуальных зависимостей (ММКЗ) несколько пошатнулось по причине изобретения блочно-сортирующих преобразований BW94 и ST [1,2,3], которые тоже фактически используют контекстуальные зависимости, но без построения вероятностной модели источника, а также появления новых разновидностей словарных методов сжатия (таких как LZP, оптимальное LZ –кодирование). Упомянутые методы позволяют достичь сравнимой с контекстуальными методами степени сжатия при заметно лучшем быстродействии.

Рассмотрим процесс сжатия информации методом группы ММКЗ. Он разбивается на 2 этапа – моделирования и кодирование. Структурная схема таких методов приведена на рис.1.

Моделировщик строит статистическую модель входного потока. Модель источника в схеме ММКЗ строится следующим образом.

Рис.1. Структурная схема методов сжатия группы ММКЗ

Пусть информационный источник имеет алфавит æ и генерирует сообщение A, представляющее собой последовательность символов алфавита:

$$A=(\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots \alpha_{N)}, \forall i : \alpha_i=a_m \in æ. \quad (1)$$

Контекстом A_i^l порядка l символа α_i назовем последовательность l символов сообщения, непосредственно предшествующих символов α_i:

$$A_i^l = \langle \alpha_{i-l}, \alpha_{i-l+1}, \ldots, \alpha_{i-1} \rangle. \quad (2)$$

Введем понятие равенства контекстов следующим образом:

$$A_{i_1}^{l_1} = A_{i_2}^{l_2} \Leftrightarrow (l_1 = l_2 = l) \& (\forall k : 1 \le k \le l \Rightarrow \alpha_{i_1-k} = \alpha_{i_2-k}) \quad (3)$$

Причем если $\alpha_i = a_m \in æ$, то будем говорить, что символ a_m появился в контексте A_i^l.

Построение марковской модели источника методами ММКЗ и ее использование для сжатия ими для других целей происходит путем последовательной обработки символов сообщения.

Модель, хранит встречавшиеся ранее различные контексты (все или лишь некоторую их часть – «наиболее важные» с точки зрения контекстной разновидности метода). Для каждого из них хранится M счетчиков, содержащих количество появлений $C(A_i^l, a_m)$ каждого из символов алфавита $æ$ в данном контексте, то есть количество таких контекстов $A_{i'}^{l'}$, что $A_{i'}^{l'} = A_i^l$ и в контексте $A_{i'}^{l'}$ появился символ a_m.

Для каждого следующего поступающего на вход символа α_i моделирующих дает оценку $\hat{p}_i(a_m)$ распределение вероятностей для него:

$$\hat{p}_i(a_m) = \hat{p}\{\alpha_i = a_m\} = \frac{C(A_i^l, a_m)}{\sum_{j=1}^{M} C(A_i^l, a_j)}. \quad (4)$$

На основании полученного из входного потока очередного символа $\alpha_i = a_{m^*}$ и выданной моделировщиком оценки $\hat{p}_i(a_{m^*})$ кодировщик, в соответствии с сообщением [4]

$$C(a_m) = -\log_2 p(\{\alpha_i = a_m\}),$$

оптимально кодирует этот символ кодом длиной $C(\alpha_i) = -\log_2 \hat{p}_i(a_m \cdot)$ бит при помощи арифметического кодирования.

После этого осуществляется процедура, называемая обновлением текущей модели символом a_m в контексте A_i^l. А именно, увеличиваются на единицу значения счетчиков модели (всех или некоторых – зависит от конкретной разновидности метода) вида $K(A_{i'}^{l'}, a_m)$ таких, что $\exists l' : (1 \le l' \le i) \& (A_{i'}^{l'} = A_i^{l'})$.

Эти действия повторяются для каждого символа входного потока.

Асимптотическая оптимальность кодирования методами марковского моделирования источника доказана, но на практике сходимость методов ММКЗ оказывается довольно медленной (хотя и более быстрой, чем у словарных методов группы LZ).

В рамках описанного общего метода ММКЗ к построению марковской модели источника остается несколько принципиальных проблем, от способа решения которых зависит скорость сходимости и эффективности получаемой модели, то есть точность даваемых его оценок $\hat{p}$, что, в свою очередь, определяет качество сжатия. Среди упомянутых проблем следует отметить выбор оптимального порядка модели источника, оценивание вероятности символа ухода, эффективная адаптация модели к меняющимся характеристикам источника. Этим проблемам посвящено множество исследований [5, 6, 7].

На основании анализа имеющихся разновидностей схемы построения ММКЗ – моделей источника, с целью повышения эффективности сжатия бинарной и смешанной информации методами ММКЗ предлагается изменить схему моделирования источника и применить мультимодельный подход с конкурирующими ММКЗ-моделями. Соответствующая схема алгоритма сжатия представления на рис. 2.

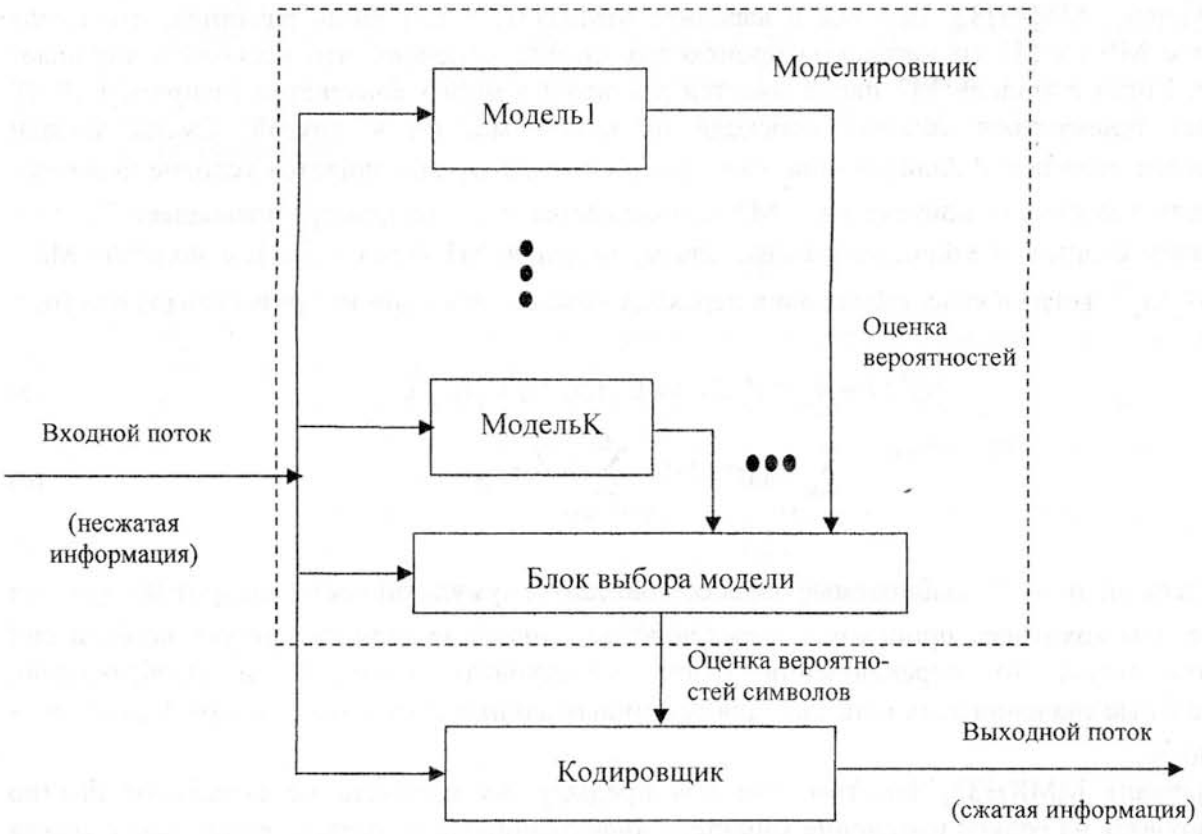

Рис. 2. Структурная схема предлагаемых методов сжатия ММКНЗ

Как видим, отличие от классической схемы (рис. 1) здесь в том, что моделировщик создает несколько моделей источника (каждая из моделей создается одинаковым алгоритмом – одним из известных уже методов группы ММКЗ, и все модели обычно обновляются каждым из символов входного потока), а встроенный в него интеллектуальный блок выборы управляет всеми моделями и выбирает для использования кодировщиком оценки той из них, которая, по его мнению, наиболее адекватно отражает текущие свойства информационного источника. Кроме того, блок управления может по мере необходимости реинициализировать (вернуть к первоначальному «пустому» состоянию, с тем чтобы начать ее построение заново) некоторую модель, а также принять решение не обновлять некоторые модели текущим символом. Очевидно, что при достаточно разумном алгоритме блока выбора такая схема позволит заметно повысить адаптивность схемы сжатия.

Поскольку предлагаемый вариант моделировщика содержит несколько вариантов практической реализации предложенной схемы ММКНЗ.

Рассмотрим несколько вариантов практической реализации предложенной схемы ММКНЗ.

Вариант ММКНЗ1 Используются 2 модели (М1 и М2) с равным максимальным числом запоминаемых контекстов J (будем называть это число размером модели). В начале сжатия начинается построение только модели М1. Она же и используется в это время для сжатия. Когда количество контекстов в М1 достигнет $J/2$, начинается построение модели М2, но для сжатия по-прежнему используется М1. Когда модель М1 заполняется (количество контекстов в ней достигает максимально возможного - J), модели меняются местами – М1 реинициализируется, а для сжатия начинает использоваться М2 (к этому моменту она содержит не менее $J/2$ контекстов). Когда М2 заполняется, они снова меняются местами и т.д. Данная схема позволяет повысить адаптивность алгоритма за счет того, что распределение вероятностей символов в прошлом не используется для оценивания вероятностей в будущем.

Вариант ММКНЗ2. Все как в варианте ММКНЗ1, с той лишь разницей, что смена модели с М1 на М2 (и наоборот) происходит интеллектуально, что несколько улучшает сжатие. Когда в модели М2 накапливается достаточно много контекстов (например, $J/4$), начинает проверяться условие перехода от одной модели к другой. Смена модели происходит если после кодирования очередного символа α_i выполняется условие перехода, либо если количество контекстов в М2 приближается к J (например, превышает $7J/8$). Обозначим стоимость кодирования символа α_i моделью М1 через $C_1(\alpha_i)$, а моделью М2 – через $C_2(\alpha_i)$. Тогда в качестве условия перехода можно взять один из критериев (5) или (6) :

$$\forall i' : i - n_c < i' \le i \Rightarrow C_2(\alpha_{i'}) \le C_1(\alpha_{i'}), \tag{5}$$

$$\sum_{i'=i-n_S+1}^{i} C_2(\alpha_{i'}) < \sum_{i'=i-n_S+1}^{i} C_1(\alpha_{i'}). \tag{6}$$

Здесь n_c и n_S - выбираемые заранее константы кумулятивности алгоритма: чем они меньше, тем «охотнее» происходит переключение с одной модели на другую, но если они слишком малы, то переключения будут происходить случайно и неоправданно. Оптимальные значения этих констант для различных данных составляют n_c - от 4 до 15, n_S - от 10 до 50.

Вариант ММКНЗ3. Заметим, что оба предыдущих варианта не позволяют быстро отреагировать на редкое изменение характера информации (например, с технического текста на литературный или с бинарных данных на текстовые). Чтобы устранить этот недостаток,

помимо двух основных моделей размером J , введем еще дополнительные две модели существенно меньшего размера, $J/10$. Обозначим их М3 и М4. Эти модели (как и основные) развиваются, меняя друг друга по правилам, описанным в варианте ММКН32, используя условия (5) и (6). Однако, в отличие от ММКН32, постоянно происходит выбор наиболее оптимальной в данный момент для кодирования модели из М1-М4. Выбор номера k модели происходит по критерию, аналогичному (6): выбирается такое k , что

$$\sum_{i'=i-n_S+1}^{i} C_k(\alpha_{i'}) = \min_{k'=1}^{4} \sum_{i'=i-n_S+1}^{i} C_{k'}(\alpha_{i'}). \qquad (7)$$

Смысл использования дополнительных моделей М3 и М4 в том, что они обеспечивают быструю адаптацию моделировщика (поскольку размер моделей мал, а следовательно они часто реинициализируются поэтому хранят информацию только о локальных свойствах источника) при изменении характера данных, не ухудшал его характеристик на однотипной информации.

Применение мультимодельного подхода к моделированию источника информации, разумеется, не ограничивается приведенными тремя схемами. Возможно построение и иных схем, отличающихся числом и относительным размером моделей, а также логикой работы блока выбора модели.

Отметим, что во всех предложенных вариантах схемы ММКН3 блок выбора основывает свои выводы на предистории, поэтому точно такие же выводы этот блок может сделать и при декодировании сжатых данных, т.е. нет надобности кодировать в поток сжатой информации какие-либо служебные символы для управления блоком выбора при декодировании.

Мультимодельный подход к моделированию источника информации, на основании которого разработаны три схемы моделирования с использованием нескольких конкурирующих ММК3-моделей и блока интеллектуального выбора модели для оценивания вероятностей символов (ММКН3-схема). Данные схемы позволяют заметно повысить адаптивность ММК3-моделей к изменяющимся данным и тем самым оптимизировать сжатие смешанной информации.

Список литературы

1. Винер Н. Кибернетика: Пер. с англ. – М.: Советское радио, 1958. – 288 с.
2. Колмогоров А.Н. К логическим основам теории информации и теории вероятностей // Проблемы передачи информации. – 1969. Т.5, №3. – С. 3-7.
3. Abramson N. Information Theory and coding. – New York: McGraw-Hill. – 1963.
4. Cleary J.G., Teahan W.J. Some experiments on the zero frequency problem // Proceedings DCC'95. – IEEE Computer Society Press. – 1995.- P. 36-42.
5. Cleary J.G., Teahan W.J., Witten I.H. Unbounded length contexts for PPM // Proceedings DCC'95. – IEEE Computer Society Press. – 1995. – P. 52-61.
6. Moffat A. Implementing the PPM data compression scheme // IEEE Transactions on Communication. – 1990. – Vol. 38, №11. – P. 1917-1921.

Поступила 10.12.2008