An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Aug. 8, 2016

Lecture #7B: Universal Source Coding-II: Lempel-Ziv Welch Algorithm (LZW)

Outline of the lecture

- LZW encoding

Outline of the lecture

- LZW encoding
- LZW decoding

LZW Encoding

- It was published by Terry Welch in 1984 as an improved implementation of the LZ-78 algorithm published by Lempel and Ziv in 1978.

LZW Encoding

- It was published by Terry Welch in 1984 as an improved implementation of the LZ-78 algorithm published by Lempel and Ziv in 1978.
- This procedure creates a dictionary that contains a list of previously encountered phrases along with associated codewords.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

LZW Encoding

- It was published by Terry Welch in 1984 as an improved implementation of the LZ-78 algorithm published by Lempel and Ziv in 1978.
- This procedure creates a dictionary that contains a list of previously encountered phrases along with associated codewords.
- In this procedure, the input is incrementally parsed into phrases where each new phrase is a concatenation of an old phrase that has occurred in the past and an innovation symbol.

LZW Encoding

- It was published by Terry Welch in 1984 as an improved implementation of the LZ-78 algorithm published by Lempel and Ziv in 1978.
- This procedure creates a dictionary that contains a list of previously encountered phrases along with associated codewords.
- In this procedure, the input is incrementally parsed into phrases where each new phrase is a concatenation of an old phrase that has occurred in the past and an innovation symbol.
- Same dictionary is created at the encoder as well as decoder with list of phrases represented by an integer index.

LZW Encoding

- It was published by Terry Welch in 1984 as an improved implementation of the LZ-78 algorithm published by Lempel and Ziv in 1978.
- This procedure creates a dictionary that contains a list of previously encountered phrases along with associated codewords.
- In this procedure, the input is incrementally parsed into phrases where each new phrase is a concatenation of an old phrase that has occurred in the past and an innovation symbol.
- Same dictionary is created at the encoder as well as decoder with list of phrases represented by an integer index.
- Initially the dictionary is initialized to all length-one phrases (q of them).

LZW Encoding

- It was published by Terry Welch in 1984 as an improved implementation of the LZ-78 algorithm published by Lempel and Ziv in 1978.
- This procedure creates a dictionary that contains a list of previously encountered phrases along with associated codewords.
- In this procedure, the input is incrementally parsed into phrases where each new phrase is a concatenation of an old phrase that has occurred in the past and an innovation symbol.
- Same dictionary is created at the encoder as well as decoder with list of phrases represented by an integer index.
- Initially the dictionary is initialized to all length-one phrases (q of them).
- Codeword is a pair representing $(C(w) \cdot w_k)$ where $C(w)$ is the code (index in the dictionary) of the prefix of the new phrase and w_k is the innovation symbol.

LZW Encoding

- When the dictionary is full, different methods can be used to reset parts of the dictionary

LZW Encoding

- When the dictionary is full, different methods can be used to reset parts of the dictionary
  - Reset half of the phrases in the dictionary that contain oldest unused phrases (excluding the alphabet letters.) Use these entries for new phrases.

LZW Encoding

- When the dictionary is full, different methods can be used to reset parts of the dictionary
  - Reset half of the phrases in the dictionary that contain oldest unused phrases (excluding the alphabet letters.) Use these entries for new phrases.
  - Do not reset the dictionary. Continue coding without adding any new phrases.

LZW Encoding

- When the dictionary is full, different methods can be used to reset parts of the dictionary
  - Reset half of the phrases in the dictionary that contain oldest unused phrases (excluding the alphabet letters.) Use these entries for new phrases.
  - Do not reset the dictionary. Continue coding without adding any new phrases.
  - Reset entries representing shorter phrases that are prefix of larger entries already in the dictionary.

LZW Encoding

- When the dictionary is full, different methods can be used to reset parts of the dictionary
  - Reset half of the phrases in the dictionary that contain oldest unused phrases (excluding the alphabet letters.) Use these entries for new phrases.
  - Do not reset the dictionary. Continue coding without adding any new phrases.
  - Reset entries representing shorter phrases that are prefix of larger entries already in the dictionary.
- $\lceil \log |D_n| \rceil$ bits can be used to represent codewords where $|D_n|$ is the size of the dictionary at time n.

LZW Encoding

- When the dictionary is full, different methods can be used to reset parts of the dictionary
  - Reset half of the phrases in the dictionary that contain oldest unused phrases (excluding the alphabet letters.) Use these entries for new phrases.
  - Do not reset the dictionary. Continue coding without adding any new phrases.
  - Reset entries representing shorter phrases that are prefix of larger entries already in the dictionary.
- $\lceil \log |D_n| \rceil$ bits can be used to represent codewords where $|D_n|$ is the size of the dictionary at time n.
- Alternatively, Elias coding of positive numbers can be used to represent codewords where smaller indices can be used for recent phrases (reverse order of dictionary).

LZW Encoding

- Problem: Use LZW algorithm to encode a 24-bit sequence,

$$W_1^{24} = \{001011011000011011011000\}$$

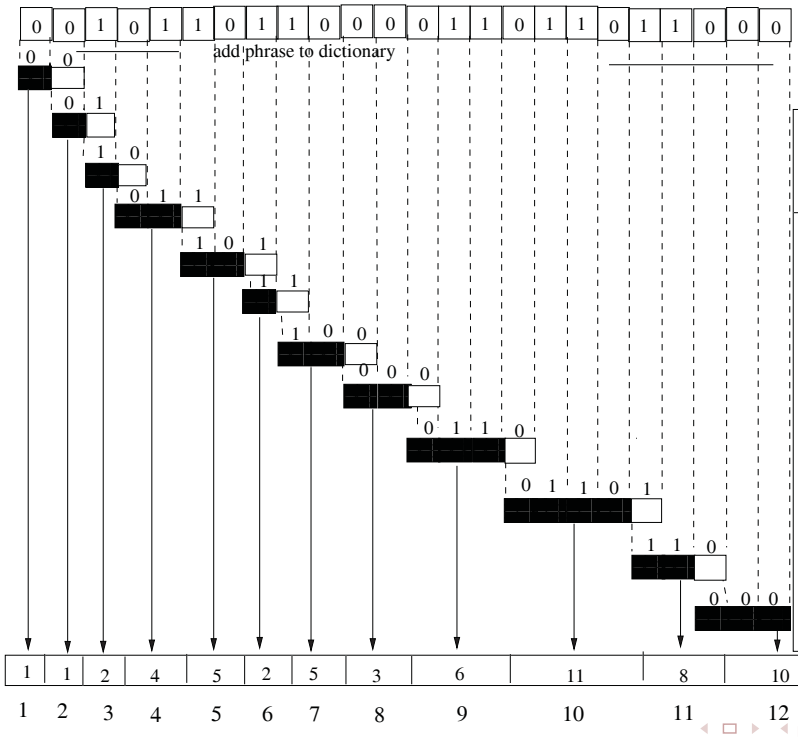
LZW Encoding

- Problem: Use LZW algorithm to encode a 24-bit sequence,

$$W_1^{24} = \{001011011000011011011000\}$$

- Solution: Encoding procedure is shown in a table (next frame).

LZW Encoding



Dictionary

Phrase	C	alt
0	1	1
1	2	2
0 0	3	1.0
0 1	4	1.1
1 0	5	2.0
0 1 1	6	4.1
1 0 1	7	5.1
1 1	8	2.1
1 0 0	9	5.0
0 0 0	10	3.0
0 1 1 0	11	6.0
0 1 1 0 1	12	11.1
1 1 0	13	8.0

LZW Decoding

- LZW decoding is done using alternative representation of the codewords in an iterative fashion.

LZW Decoding

- LZW decoding is done using alternative representation of the codewords in an iterative fashion.
- In alternative representation, codewords are represented as concatenation of another codeword and an innovation symbol.

LZW Decoding

- LZW decoding is done using alternative representation of the codewords in an iterative fashion.
- In alternative representation, codewords are represented as concatenation of another codeword and an innovation symbol.
- Decoding procedure includes timely generation of the dictionary from the received codeword sequence.

LZW Decoding

- LZW decoding is done using alternative representation of the codewords in an iterative fashion.
- In alternative representation, codewords are represented as concatenation of another codeword and an innovation symbol.
- Decoding procedure includes timely generation of the dictionary from the received codeword sequence.
- At the beginning of decoding, dictionary is initialized with codewords corresponding to the source alphabet.

LZW Decoding

- LZW decoding is done using alternative representation of the codewords in an iterative fashion.
- In alternative representation, codewords are represented as concatenation of another codeword and an innovation symbol.
- Decoding procedure includes timely generation of the dictionary from the received codeword sequence.
- At the beginning of decoding, dictionary is initialized with codewords corresponding to the source alphabet.
- A phrase is added to the decoder dictionary after each new codeword is received except the first codeword.

LZW Decoding

- LZW decoding is done using alternative representation of the codewords in an iterative fashion.
- In alternative representation, codewords are represented as concatenation of another codeword and an innovation symbol.
- Decoding procedure includes timely generation of the dictionary from the received codeword sequence.
- At the beginning of decoding, dictionary is initialized with codewords corresponding to the source alphabet.
- A phrase is added to the decoder dictionary after each new codeword is received except the first codeword.
- Every received codeword represents a prefix of a new entry into the dictionary . The innovation symbol for this entry is determined from the first symbol of the next decoded codeword.

LZW Decoding

- Problem: Use LZW algorithm to decode the received codeword sequence,

$$W_1^{12} = \{1, 1, 2, 4, 5, 2, 5, 3, 6, 11, 8, 10\}$$

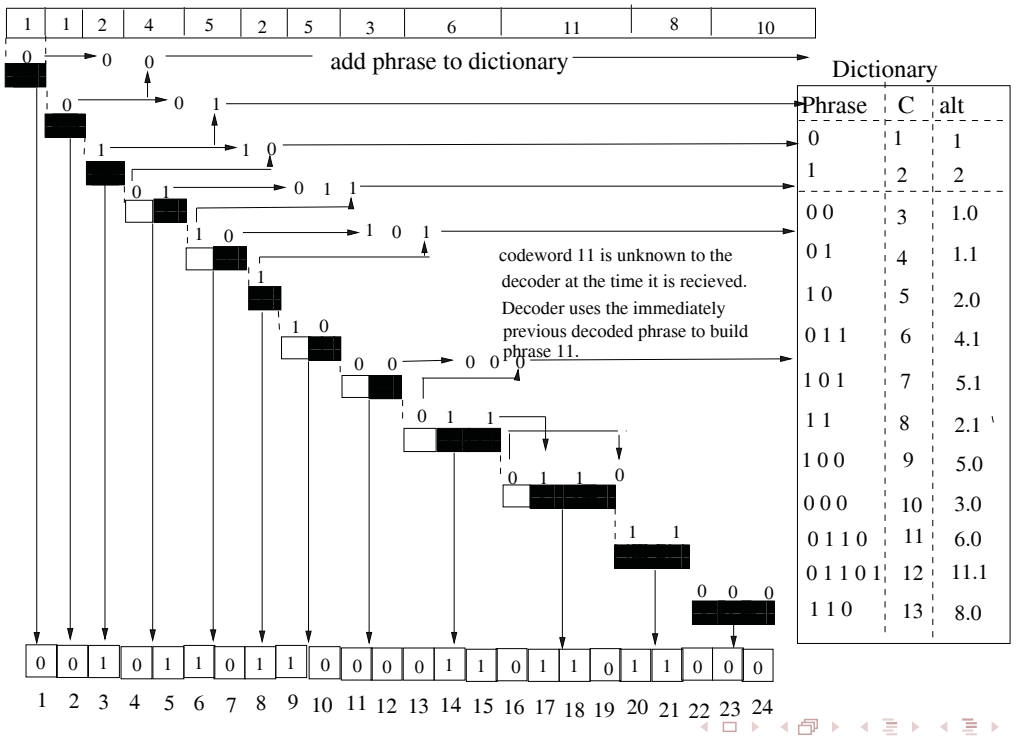
LZW Decoding

- Problem: Use LZW algorithm to decode the received codeword sequence,

$$W_1^{12} = \{1, 1, 2, 4, 5, 2, 5, 3, 6, 11, 8, 10\}$$

- Solution: Decoding procedure is shown in a table (next frame).

LZW Decoding



Dictionary

Phrase	C	alt
0	1	1
1	2	2
0 0	3	1.0
0 1	4	1.1
1 0	5	2.0
0 1 1	6	4.1
1 0 1	7	5.1
1 1	8	2.1
1 0 0	9	5.0
0 0 0	10	3.0
0 1 1 0	11	6.0
0 1 1 0 1	12	11.1
1 1 0	13	8.0