Processing Open Text Input in a Scripted Communication Scenario

Author:
F.P.M. Heemskerk

Supervisor:
J. T. Jeuring

Supervisor:
Raja Lala

Abstract

In this paper we describe a method which uses only local data available in a scripted scenario to match a player input to a player statement option. A scripted scenario is a sequence of steps in a conversation with a virtual character. At each step a virtual character talks to the player and provides one or more statement options. A virtual character reacts to a statement option with an emotional reaction and the conversation continues to a next step until an end step is reached.

In our method we use the emotional reaction of a virtual character to create a scenario specific corpus. We create a scenario specific corpus by calculating and combining emotion vectors for a stem of a word in a response and dividing it by the total amount of occurrences of a stem of a word in a scenario, creating a scenario specific word vector. An emotion vector is a unit vector with a value for at the index of an reactional emotion. The words in our method are stemmed with an algorithmic stemmer. By utilizing a scenario specific corpus, we calculate a statement vector by combining and averaging word vectors for each word in a player input statement for a step in a scenario. We also calculate a statement vector for each player statement option for that step. We compare each player statement option vector with a player input statement by calculating a cosine similarity score.The highest cosine similarity score of a player statement option vector and a player input statement is considered as an possible match.

The considered player statement option is then compared with the Differential Angle method. The Differential Angle method not only compares a player statement option with a player input statement, but also compares an emotion vector of a player statement option with a calculated vector of an option. If the angle between an emotion vector and a calculated vector is too large, the emotion vector is used for comparison. The Differential Angle method also considers if the player statement option is a possible match or that the best player statement option is not good enough to be a match and that there is no match for a player statement option.

In comparison to ReaderBench, our method performs better if there is an option that matches a player input statement.

Contents

Chapter 1

Introduction

Educational institutions and vocational programs train students in social and communication skills. People often acquire communication skills through role play to enhance inter-collegial communication [29][33]. Role play is often done with a colleague or peer in practice, or with a virtual character.

Nørgaard et al. [29] investigated the effect of communication skills training on inter-collegial communication and found a significant increase in inter-professional communication effectiveness after an in-house training. The participants were asked to fill a form on communication skills, before the training, directly after the training and 6 months after. Ruyak et al. [33], who used simulation-based learning to practice communication with both client and colleagues, confirm this result.

In a digital learning environment for training communication skills, a student converses with a virtual character, while the learning environment assesses the performance of a student on predefined learning goals for a performed conversation. Examples of such digital environments are Communicate! [21], Visual SceneMaker [17] and Tactical Questioning [16].

Terminology

First we describe the terms used for this paper.

Scenario
A scenario is a sequence of subjects that are traversed until an end node is reached in a specific context.
Subject
A subject is a directed graph of statements within a scenario, usually about a particular theme.
Statement
A statement is a piece of text, either for a virtual character or a user
Author
An author is the creator of a scenario
Authoring tool
An authoring tool is software used by a author to create and adjust a scenario

Research Problem Description

Communicate! is a digital training environment for practising communication skills. The system provides structured scenarios created by a domain expert. A user converses with a virtual character via statements. Each statement is followed by an emotional reaction from the virtual character and the system scores a chosen statement based on predefined values by the author. Communicate! determines a score at the end of a scenario and, if defined by an author, an advice for a player on how to improve his/her communication skills.

The score given by Communicate! at the end of a scenario, played by a player, is determined based on the parameters defined by an author in a scenario. An example of a parameter is tact, the ability and sensitivity to deal with difficult situations in a conversation. Communicate! might judge a player on her chosen statement options and score her on her level of tact. An example of a scenario can be found in Appendix A

A set of information defined by an author is the emotion property of a virtual character. For each statement option an emotional reaction is set for a virtual character. A virtual character can, for example, react happily, surprised or angry. We will use these two sets of information (user-defined parameters and the emotion property) to categorise a statement.

Natural Language Processing methods often use a sentiment lexicon to attribute emotions to a statement. A sentiment lexicon is a collection of words with their respective emotional and contextual values. An example of a sentiment lexicon is created by Hamilton et al. [19]. They extracted domain specific sentimental values from the website `reddit.com`. Each domain represents a subreddit on the website. Unfortunately, a general sentiment lexicon for smaller languages, such as Dutch, is not freely available at the moment.

Communicate! uses multiple statements as options to choose from to traverse a conversation with a virtual character. This can be seen as a sequence of multiple choice questions; given a context, what is the most appropriate response from the defined statements? According to Krathwohls' Cognitive Process Taxonomy [2], multiple choice questions test cognitive levels 1 to 5; Remembering, Understanding, Applying, Analysing and Evaluating. The only level not tested is Creating, cognitive level 6. To test the level of Creating, open text input seems more suitable. Open text input allows a user to create a statement to communicate with a virtual character. In our approach we use information from a scenario to create a scenario specific corpus. Our research question is: Can we use a scenario specific corpus to match an open text input to one of the player statement options and how does it compare to other available open-source natural language processing methods?

Chapter 2

Related work

In this chapter we discuss several frameworks for text analysis and text categorization, and how they relate to the requirement for our system. Here we will discuss ReaderBench and Natural Language Toolkit (NLTK).

Reader Bench

ReaderBench is a multi-lingual framework for text analysis based on text mining techniques [11]. It supports a range of analysis methods to deconstruct a text, including identification of reading strategies and text complexity assessments.

One of the methods is semantic analysis of text. ReaderBench can categorize a text on different textual levels (such as word, sentence and text level) into different emotions. The set of emotions used by ReaderBench is the same for all documents. The emotions used by ReaderBench are indicated with values ranging between 0 and 1. A value of 1 for an emotion for a document translates to that emotion being predominant. Since the emotions are always the same set, it can be translated to a vector with each dimension having a value between 0 and 1. This method can determine if a given input text is semantically similar to a player statement option in a step of a scenario.

ReaderBench uses the Stanford NLP [26] toolkit for calculating the semantic values of a document. It uses the semantic values of a large dataset from online websites.

Natural Language Toolkit

Natural Language Toolkit (NLTK) is an open source framework for natural language processing written in Python [25]. It provides symbolic and statistical processing modules for linguistic analysis. It was developed in conjunction with a course for the University of Pennsylvania.

Example modules are tokenisation and tagging of natural language. It also provides synsets for the English language. A synset is a tuple of a lemma (a word) and a set of synonyms.

It unfortunately does not contain a semantic analysis module, but provides methods to add a semantic analysis tool.

Due to the lack of a semantic analysis module, we will not use NLTK for our method and we will not compare our method with NLTK.

Chapter 3

Design Space

In this section we discuss different methods for text analysis and categorization. For our method we are looking for methods to use information in a scenario in selecting a correct player statement option. We will be looking at chatbots, intelligent agents and text categorization.

Chatbots

A possible way of communicating with a virtual character is via a chatbot. According to Bayan Abu Shawar and Eric Atwell [1] the definition of a chatbot is a machine conversation system that interacts with human users via natural conversational language. Nowadays chatbots have found their way into homes and allow communication with and manipulation of the environment. Examples are Siri, Google Assistant and Amazon Echo. Augello et al. describe an overview of several open-source chatbot technologies and their limitations [5].

A.L.I.C.E.(Artificial Linguistic Internet Computer Entity) [27] uses a pattern matching algorithm and a rule based template mechanism to determine a response on user input. A.L.I.C.E. allows the use of wildcards in template rules to allow the chatbot technology to converse on topic areas it has no knowledge of or rules for.

Chatterbot [10] uses machine learning to determine a response to user input. The bot will initially respond with a predefined response learned from a corpus and a set of predefined statement-response pairs. Chatterbot compares a statement from a user with statements from the learning set and finds the statement with minimal textual distance. The chatbot then replies with the corresponding response. After each statement Chatterbot updates its statement-response corpus by increasing or decreasing the likelihood of making the same response to the given statement, depending on a users given score. A negative score from a user results in decreasing the possibility of responding with the same statement, a positive score results in increased likelihood.

User:	Hello, how are you?
Chatbot:	I'm fine, how are you?
User:	Good, thanks for asking.
Chatbot:	You're welcome.

Table 3.1: Chat session 1 - Positive user rating

User:	Hello, how are you?
Chatbot:	A bird landed on the porch today.
User:	Was it a pretty bird?
Chatbot:	It was raining outside.

Table 3.2: Chat session 2 - Negative user rating

ChatScript [35], like A.L.I.C.E., uses rules to determine which response is best for a user input. The main difference between A.L.I.C.E. and ChatScript is the inclusion of Natural Language Processing. ChatScript uses the stems of the words in a given statement and uses those stems to determine which statement is best to reply with. Furthermore, ChatScript also features a topic control module; a system to make response rules topic specific. The system first checks for rules within the current topic and replies with a defined response. If a confidence value (the likelihood of being correct) is too low within the current topic, ChatScript will search in other topics for a rule with a higher confidence value.

Rule based methods do not work with our system, since rules are created by hand. An author of a scenario in Communicate! is most likely not a programmer and is unlikely able to create rules which fit the system. Another reason for not using rule based methods is the separation of scenario creating and simulation. An author is only responsible for the creation of a scenario and does not have to work with the simulation of the scenario.

A restriction in these three chatbot technologies is the lack of context management. The chatbots can respond in any context, but make no difference between topics. As long as the response from a chatbots' database scores sufficiently high, the chatbot can use that response. This is undesirable for our system since the goal is to teach and to test a user deterministically in a specific context. A.L.I.C.E., Chatterbot and ChatScript respond to a statement without knowing the current context. Although ChatScipt has a topic control module, it lacks the ability to stay within a certain topic or context. A difference between our desire for our system and chatbots is, the ability to converse confined. Chatbots converse unconfined, a chatbot conversation can go on until a user determines the conversation is over, whereas a scenario has a beginning and an end.

Machine learning is often used to support chatbots. The chatbots base their responses on previously learned responses and determine their future replies on feedback values from users. After a session with a chatbot, a user rates the chatbot on its credibility with a grade. The chatbot processes the user ratings to determine if the responses used in the chat session were convincing and appropriate. Consider chat sessions 1 and 2 (Table 3.1 & 3.2). The user in the first chat session determines the conversation positive, the responses of the chatbot were appropriate with the statements from the user. In chat session 2 the responses of the chatbot were not appropriate responses to the statements of the user and therefore the user rated the conversation negatively. The ratings of the user affect the probability of using the same replies again for the same statements. The usage probability of negatively rated responses will be lowered and for positively rated it is increased. A result will be that the appropriate responses will be used more often for a statement given.

Our system will not use machine learning, since we do not posses a large dataset to train the system with. Furthermore, if an author adjusts a parameter value or changes a emotion of a virtual character, the trained data has to be recreated.

Intelligent agents

A different approach for interacting with a virtual character is via an intelligent agent. According to Wooldridge and Jennings [36], an intelligent agent is a piece of software with the following properties: autonomy, social ability, reactivity and pro-activeness. An agent works without the interference of a human or other system and can communicate with its surroundings both reactively and pro-actively. An intelligent agent in our setting could interact with a player via Communicate! and will also respond to the statements given. A difference between a chatbot and an intelligent agent is the tracking of state. While a chatbot is stateless and is only reacting to the input of a user, an agent keeps track of the state of a conversation and can react pro-actively.

An example of using intelligent agents instead of chatbots is in a library. According to Rubin and Chen [32], intelligent agents can enhance the experience of library visitors. They conclude that agents used in libraries need to know what kind of information a visitor wants, for example a visitor looks for educational conversation or requires assistance with looking for a book.

Augello et al. [4] created a system similar for social learning called SALVE, Social Agents for Learning in Virtual Environments. Using rule-based languages and pattern matching, they created a system where conversations are defined as rules. A rule contains a question, a pattern and a possible response. The pattern in a rule corresponds with a response of a user. A agent will respond to a statement of a user based on provided responses.

As described in Section *Chatbots*, the authoring tool in Communicate! is created from the perspective of a social skill teacher (who is most likely a non-programmer) with ease-of-use expressive constructs. Therefore, a rule based system will not be implemented.

Information categorization

An aspect of Natural Language Processing (NLP) is information categorization. Latent Semantic Analysis (LSA) [13] is a technique to analyse documents. LSA creates vectors from documents by creating a frequency vector. A frequency vector is a vector with for each word in a document its frequency. Each vector has a direction in the document vector space, which is used to determine if a specific term is associated with the document. To ensure that a document vector does not get too large, only the most prominent words in a document are used in the vector. The vector is then normalized by dividing it by the length of the vector.

LSA requires full documents to create a document specific frequency vector. A statement option in a scenario in Communicate! contains a small amount of words and is insufficient to use with LSA.

A neural network can be used to classify and categorize text. Yu et al. [38] have created a neural network to categorize text using LSA and support vector machines. A support vector machine uses positive and negative training vectors to create a multidimensional plane [20]. These vectors can represent text (such as a bag-of-words [39]) or image information. This plane is used to categorize a new input vector in either the positive plane or the negative plane. A vector is in the positive plane if the vector is a positive match compared to the training data. Yu et al. used this technique to categorize a text with the latent semantic analysis information. Using a neural network the plane is optimized to represent the data better. A neural network is not feasible for our problem. A neural network requires user to converse with a scenario and we do not have a large group of users. Furthermore if an author changes a parameter value, the test data is not usable any more.

Vectorization of information can be used for our problem. Information vectors can be analysed, such as with similarity calculations or distance calculations.

Chapter 4

Multiple choice to open text

To offer the possibility to enter open text input in Communicate! we first have to determine which method is best for matching an open text input to a statement. We want to answer the research question: Which of the following methods is most accurate in matching open text input to predefined statement options within a scenario? A method is accurate if it matches an open text input to a predetermined best player statement option, including a non-matching option. A non-matching option is selected if none of the predefined player statement options can be considered a best match with an open text input. The best option is determined by the author of an open text input.

In our approach we use information of a scripted scenario to match an open text input to one of the possible player statement options. An advantage of keeping the structure of a scenario is that experts can still create scenarios without understanding complex storage structures, such as XML.

To use player input text in Communicate! we take the following steps; first we calculate a vector from an input text and the player statement options. Then we calculate either the distance or the angle between an input text vector and the player statement options. The option with the smallest angle or distance with the open text input is chosen as a possible option.

The methods we use for Scenario Specific Corpus Matching (SSCM) for determining a best match for a text input from a student are: Euclidian distance and Cosine similarity. We will also test which information (parameter versus emotion) works best for determining a best match.

We chose Euclidian distance and Cosine similarity for testing the similarity of vectors. Euclidian distance calculates the distance between two vectors and Cosine similarity calculates the angle between vectors. Both vector measures are used in multiple systems to calculate vector similarities [6, 8, 14, 37].

We considered to use LSA for our method, however LSA requires large quantities of text to create a vector. The quantity of words in a statement in a scenario is limited. The difference between two statement options can be minimal, which is not observed by LSA since it looks at the most prominent words.

We will test these methods against two other textual matching algorithms: ReaderBench and FuzzyString.

ReaderBench is a multi-lingual framework for analysing text, relying on advanced Natural Language Processing techniques [12]. The system has the ability to analyse text on sentiment on document, sentence and word level [18]. The analysis is based on several sentiment models based on words, making it a context free analysis. Matching with ReaderBench is based on the sentiment of statements and an open text input.

The FuzzyString matching algorithm [3] calculates the number of string operations that need to be performed on an open text input to match with a statement. The lower the number of operations, the higher the score of a statement. This method is fast and accurate for similar sentences, but lacks accuracy for sentences with similar semantics with a different syntax.

First we describe how we use Natural Language Processing to filter and substitute words in a statement to prepare for usage in matching methods. Then we describe the methods to determine which of the statement option given by Communicate! matches best with the statement of a player. In the last section we describe validation methods to determine the quality of our method in matching open text to a statement option.

Natural Language Processing

When a student enters text in Communicate, we want to determine which of the predefined scenario options at this point matches best with the student text. According to Liddy [23] and Feldman [15] linguistic analysis consists of the following levels: phonology, morphology, lexical, syntactic, semantic, discourse and pragmatic.

The phonological level describes the pronunciation of words. This linguistic level is unimportant for our information retrieval systems, since a statement in Communicate! is stored as a constant string of characters without audio information. Phonetics is important for voice recognition and speech-to-text systems.

Morphology describes the individual meaning of word components (morphemes). Examples of morphemes are stems (such as *parent* for parenting, parenthood, etc.), suffixes (mov*ing*, valid*ation*) and prefixes (*un*pleasent, *pre*render).

The lexical linguistic level represents the meaning of a word. Some words have multiple meanings (homonyms, such as fly; it represents both a name of an animal as well as a verb of a way of transportation) and some words have the same meaning, but different spelling (synonyms, such as beginning and start)

Syntax is the grammatical structure of a sentence. The syntax covers an orderly representation of word structures in a correct order. A sentence can be structured in different ways, each having a different meaning.

Semantics describes a meaning of a sentence. The semantics of a sentence is determined by the order of a sentence (syntax) as well as the context in which a sentence is produced.

Linguistic analysis on discourse level examines text and its structure to extract additional meanings. For example, a research paper contains an abstract with important findings and conclusions, which can usually be found in the first paragraph. Discourse level analysis extracts a specific role of a statement given the type of document the statement is in, and its position within the document.

The pragmatic level analysis depends on knowledge outside of an analysed document. The knowledge is not present in a document and is retrieved from outside the document. This level of linguistic analysis is difficult to realise since the knowledge retrieved from outside a body of text can contain references to other knowledge bodies. If an information retrieval system follows these references and goes deeper into its references, the amount of information increases significantly as well as the time to process all the information. Another problem is the usefulness of the retrieved information. A retrieved body of information may not be useful for the analysis of the original text.

The linguistic levels that SSCM processes are the semantic and discourse levels of words. To process these levels, we must first manipulate the statements from a scenario to fit our process. The rules for text rewriting, as suggested by Lie et al. [24], are:

- Deletion of irrelevant words, such as "the", "a(n)" and "to"
- Substitution of words by standard synonyms and expressions.
- Transformation of complex sentence forms into simple normal forms
- Changing the order of phrases into a predefined normal order.

For our system, we will only consider the first two rules of text rewriting: deletion and substitution. By deleting irrelevant words (such as articles), we remove noise from a statement. These words are irrelevant for the meanings of words in a statement on both semantic and discourse level.

Techniques

In this section we will discuss the techniques to find a best predefined statement option from an input text, if a best match is available. First we explain the vectorization of words. Then we discuss the matching techniques for finding a best match, as well as the possibility of finding no correct match.

Vectorization

To bind information values to a word, we create a vector for each unique word in a scenario. We store an emotion property in a vector with dimensions equal to the amount of possible emotions in a scenario and user-defined parameters in a vector with a dimension for each parameter. We refer to an emotion property or a user-defined parameter as an information value.

For each word in a statement, we store the information value in a temporary vector T. For example, if the emotion of a virtual character changes to "happy", we add a value of 1 to $T[happy]$. Vector V is then created from vector T by dividing it by the number of occurrences for that specific word.

$$V_{word} = \frac{T}{\#occurrences} \tag{4.1}$$

Vector V represent a scenario specific vector with the average parameter change, for a word for the defined information set (emotion or parameter).

A vector for a statement is also created. For such a vector for a statement, for example for a text input, the average of all word vectors is calculated.

$$V_{statement} = \frac{\sum V_{word}}{\#words} \tag{4.2}$$

$$\begin{aligned} V_{werken} &= \{0.25, 0, 0, 0, 0, 0, 0, 0, 0, 0.5, 0.25, 0, 0, 0\} \\ V_{nooit} &= \{1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\} \end{aligned}$$

Figure 4.1: A vector representation of the word "werken" and "nooit"

For an example scenario (see Appendix B) the created word vector of the words "werken" and "nooit" are presented in Fig. 4.1. It show that the reactions on a statement with the word "werken" are "angry", "little sad" and primarily "sadness". The reactions for the word "nooit" in the same scenario for a statement is "angry".

$$\begin{aligned} V_{input} =& \{0.40581, 0.02978, 0.04559, 0.06359, 0.04537, \\ & 0.07729, 0.09678, 0.13377, 0.01218, 0.01067, \\ & 0.01029, 0.04736, 0.01509, 0.00340\} \\ V_{option2} =& \{1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\} \\ V_{option2} =& \{0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\} \end{aligned}$$

Figure 4.2: Two vector representations for an input text and a player option statement.

In Fig. 4.2 two vectors are presented: one from an input text and one from a predefined player option statement. The input text is: "jij bent er nooit eens een keer" and the player options are: "jij bent er nooit op tijd en werkt slordig" and "hoi je hebt net de daily standup gemist kan gebeuren natuurlijk maar merk wel dat je wat vaker moeite hebt om hier op tijd te zijn toch" The word "een" is ignored as part of the text rewriting rules.

The vector representations of words are stored in a scenario specific corpus, as well as the vectors of the player statement options.

Similarity calculations

In this section we discuss the similarity methods we use in our approach: euclidean distance and cosine similarity.

Euclidean distance

One of the methods we consider to determine which statement option is described by the text input, is Euclidean distance. We expect that the distance between a text input from a player and a best statement option from a scenario is minimal. The position of the two vectors are determined by the average parameter values and the difference between the value is minimal if the values are close to each other.

Euclidian distance D is calculated with Pythagoras' theorem [22] between two vectors V_{input} & V_{option}, each representing a statement. Note that the lengths of V_{input} and V_{option} is the same, since both contain the same set of information of a scenario.

$$D = \sqrt{\sum (V_{option\ i} - V_{input\ i})^2} \tag{4.4}$$

The euclidean distance of the example vectors for a input and player statement options as shown in Fig. 4.2 is $D(V_{input}, V_{option1}) \approx 0.60507$
and $D(V_{input}, V_{option2}) \approx 1.07251$. According to the euclidean distance, the first player statement option is considered best.

Cosine similarity

Another method to test similarity between a statement option and a text input based on their vector representation is cosine similarity. Instead of calculating the distance between two vectors, the angle is calculated.

$$S = \cos\phi = \frac{\sum V_{option\ i} V_{input\ i}}{\sqrt{\sum V_{option\ i}}\sqrt{\sum V_{input\ i}}} \tag{4.5}$$

Using the cosine similarity, we test if a statement option and a text input are within a similar region. The higher the similarity, the higher the chance the statement option is plausible to use.

The cosine similarity of the example vectors for a input and the player statement options as shown in Fig. 4.2 is $S(V_{input}, V_{option1}) \approx 0.88590$ and $S(V_{input}, V_{option2}) \approx 0.06501$. The calculation shows that the first player statement option is considered best.

Validation

To answer our research question, two validation tests will be used; Some Option Matches (SOM) and No Option Matches (NOM). The accuracy of a matching algorithm is determined by the number of correct option selections in a test.

Test data

To test our method and compare it to other methods, a dataset and 'golden-standard' has been generated. At University Utrecht we conducted an experiment with final year bachelor computer science students. The students work in a team project to develop a software product for a real customer. In the second semester of college year 2017 - 2018, a total of eighty two students were assigned in eight teams. Seventy eight students gave consent to use their data for research. The age range of the students was between twenty and twenty eight years.

In the first round of guided-sessions with each team, students played the scenario with multiple choice answers in Communicate!.

We conducted a 2nd round of guided sessions to gather open-text input from students. We gathered data for the matching methods. We adapted Communicate! to gather data in the 2nd round: a student gets an open-text input box in which she writes her response instead of choosing from the multiple choices at each step. A student inputs her text, after which Communicate! displays the available scripted player statement options at this step. There is also an option *No response matches* displayed at each step. A student indicates which statement is closest to her open-text input, or chooses *No response matches* in case no scripted statement matches her input. If a student chooses *No response matches*, Communicate! thereafter asks her to select one of the scripted statement options to continue the simulation. The information gathered from the 2nd round is used for test data for our method.

We divided the test data in three sets: Training, Validation and Testing. We use the Training and Validation test data for different configurations of SSCM to find a best configuration. The Validation test data is used to validate findings in results with the Training test data. We use the Testing test data as a black box method for our final configuration for comparison with ReaderBench. The Training and Validation test data can openly be addressed to find why a configuration works.

Some Option Matches

SOM tests if a system finds a match in the statement options given a text input and if it is a correct match according to the player. The SOM percentage represents the number of accurate matches, according to a user of Communicate!, divided by the total amount of match requests for a system, such as SSCM. A test case assumes there is a statement option that can be considered a best match to a text input.

No Option Matches

While testing for matches between dialogue options and a text input, it is possible a best match is not plausible enough to use. NOM tests the accuracy of a system if no match is found, given statement options in a scenario. For NOM tests we assume no statement option describes the text input accurately and no option should be selected.

To test for no match occurrences, we will use a threshold. Each of the above methods will use a different threshold. A threshold will be manually determined by incrementing and decrementing it until a maximum is found in the NOM score.

Chapter 5

Results

For each configuration of SSCM we test the accuracy. A configuration consists of a similarity method (euclidean distance or cosine similarity), an information set (emotion or parameter) and vectorization sequence methods used. We compare each configuration of SSCM with the same test data, the Training test data as described in Chapter 4. The total amount of SOM's in the Training test data is 306 and the total amount of NOM's is 277.

Each configuration results in five values: SOM, NOM, SOMasNOM, NOMasSOM, WrongSOM. The SOM value indicates the amount of correct SOM classification for a configuration. The NOM value indicates the amount of correct NOM classifications. The SOMasNOM value indicates the amount of tests that a configuration should classify as a SOM, but is classified as a NOM. The NOMasSOM value indicate the amount of tests that a configuration should classify as a NOM, but is classified as a SOM. The WrongSOM value represents the amount of tests that a configuration classifies as a SOM and matches to the wrong player statement option.

For each configuration we show the SOM and NOM values, as well as the combined value of SOM and NOM, which we indicate by Total Correct. The complete results, with the SOMasNOM, NOMasSOM and WrongSOM values, are found in Appendix C. For each configuration a selection of results from the Training test data is shown. The results in bold are the highest in their respective column (SOM, NOM, Total Correct) in the complete results.

To test whether a configuration is interesting, we calculate a F_1-score. A F_1-score is a measure to test the accuracy of a configuration [34]. It considers both the precision and the recall of a configuration for the score. The precision of a configuration (Eq. 5.1) is calculated by dividing the number of correct positive classifications by the total amount of positive classifications (both correct and incorrect). The recall of a configuration (Eq. 5.2) is calculated by dividing the number of correct positive classifications by the relevant samples (samples that should have been classified as positive). The F_1-score is then calculated with Eq. 5.3.

$$precision = \frac{TP}{TP + FP} \tag{5.1}$$

$$recall = \frac{TP}{TP + FN} \tag{5.2}$$

$$F_1\text{-}score = 2 * \frac{precision * recall}{precision + recall} \tag{5.3}$$

In terms of true and false classifications for both positive and negative results, F_1-score is calculated as shown in Eq. 5.4, where β equal 1 (as shown in Eq. 5.5).

$$F_\beta\text{-}score = \frac{(1 + \beta^2) * TP}{(1 + \beta^2) * TP + \beta^2 * FN + FP} \tag{5.4}$$

$$F_1\text{-}score = \frac{2 * TP}{2 * TP + FN + FP} \tag{5.5}$$

A F_1-score results in a value between 0 and 1. A score of 1 is a perfect test score, a test contains only true positive classifications and has perfect precision and recall, while a score of 0 indicates a test contained only incorrect classifications. For each configuration we calculate a SOM and a NOM F_1-score. A SOM F_1-score indicates the accuracy of a configuration with respect to SOM classifications and player statement option matching. A NOM F_1-score indicates the accuracy with respect to NOM classifications.

For a SOM F_1-score, the true positive classifications (TP) is the value in the Correct SOM column. The value of the false negative classifications (FN) is presented in the SOMasNOM column. The value of the false positive classifications (FP) is the combined value of the WrongSOM and the NOMasSOM columns. False positive classifications are both the NOM classifications in the test data classified as a SOM as well as SOM values matched to a wrong player statement option.

We calculate a NOM F_1-score with the value in the NOM column for TP, the value in the NOMasSOM column for FN and the value in the SOMasNOM column represents FP.

We find a SOM F_1-score more important than a NOM F_1-score when comparing configurations. For a configuration we calculate a Weighted F-score to express this. To calculate a Weighted F-score we use Eq. 5.6.

$$Weighted\ F\text{-}score = 2 * SOM\ F_1\text{-}score + NOM\ F_1\text{-}score \tag{5.6}$$

Base configuration

First, we create a base line of correct SOM and NOM predictions. The configuration of the vector sequence method is the creation of word vectors as described in Chapter 4 for a player input statement, player statement options and a scenario corpus.

Each player statement option vector is matched with the input vector with either a euclidean distance or cosine similarity calculation as the similarity method, resulting in a similarity score between 0 and 1.

We increment a threshold in steps of 0.025 between the values 0 and 1. We consider a statement option with the highest similarity score for a step in a scenario for either SOM or NOM. If the similarity score of an option is higher than the given threshold, we consider it a SOM, otherwise it is a NOM.

We use both information sets (emotion and parameter) for the base line. This will result in a total of 80 tests for each test data set.

For the base line we show both the highest values for the correct number of SOM, NOM and the Total Correct, as well as the values of a threshold value of our interest. For future configurations we only show values of interest.

Table 5.1: Cosine Similarity - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0	**167**	30	197	0.499	0.211	1.208
0.5	138	60	198	0.452	0.344	1.247
0.525	136	62	198	0.448	0.350	1.246
0.550	135	65	200	0.448	0.361	1.257
0.575	133	69	202	0.446	0.375	1.266
0.600	132	76	208	0.448	0.402	1.299
0.625	131	80	211	0.449	0.416	1.313
0.650	130	88	218	0.452	0.443	1.348
0.675	129	95	224	0.455	0.466	1.376
0.700	126	103	229	0.453	0.488	1.395
0.725	122	112	234	0.449	0.511	1.410
0.750	118	120	238	0.444	0.523	1.412
0.775	115	129	244	0.443	0.544	**1.431**
0.800	104	139	243	0.418	0.558	1.394
0.900	60	197	**257**	0.303	0.615	1.221
0.975	15	**225**	240	0.093	0.606	0.792

Table 5.2: Cosine Similarity - Parameter information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0	**90**	42	132	0.310	0.248	0.867
0.5	90	42	132	0.310	0.248	0.867
0.525	90	42	132	0.310	0.248	0.867
0.550	90	42	132	0.310	0.248	0.867
0.575	90	42	132	0.310	0.248	0.867
0.600	90	42	132	0.310	0.248	0.867
0.625	90	42	132	0.310	0.248	0.867
0.650	90	43	133	0.310	0.253	0.874
0.675	90	43	133	0.310	0.253	0.874
0.700	90	43	133	0.310	0.253	0.874
0.725	90	46	136	0.312	0.268	0.892
0.750	89	46	135	0.309	0.267	0.885
0.775	89	47	136	0.310	0.272	0.892
0.800	89	48	137	0.310	0.277	0.898
0.975	74	**89**	**163**	0.286	0.416	**0.987**

Table 5.3: Euclidean Distance - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0	**180**	0	180	0.505	0.000	1.010
0.6	126	19	145	0.394	0.125	0.913
0.625	125	20	145	0.392	0.131	0.914
0.650	125	22	147	0.393	0.143	0.929
0.675	125	31	156	0.399	0.194	0.991
0.700	119	45	164	0.392	0.259	1.043
0.725	115	58	173	0.390	0.311	1.091
0.750	110	73	183	0.386	0.365	1.137
0.775	104	108	212	0.393	0.479	1.265
0.800	84	148	232	0.358	0.557	**1.274**
0.825	60	184	244	0.293	0.595	1.181
0.850	45	209	254	0.244	0.614	1.102
0.975	11	**227**	238	0.069	0.606	0.745

Table 5.4: Euclidean Distance - Parameter information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0	**169**	0	169	0.481	0.000	0.963
0.6	169	3	172	0.484	0.026	0.993
0.625	169	3	172	0.484	0.026	0.993
0.650	169	3	172	0.484	0.026	0.993
0.675	168	3	171	0.481	0.026	0.989
0.700	166	4	170	0.478	0.034	0.989
0.725	165	6	171	0.477	0.050	1.004
0.750	161	21	182	0.478	0.161	1.118
0.775	159	26	185	0.477	0.193	1.148
0.800	158	31	189	0.479	0.224	1.181
0.825	152	38	190	0.470	0.259	**1.199**
0.850	145	43	188	0.457	0.279	1.193
0.975	44	**198**	**242**	0.232	0.599	1.063

Looking at the result above, the configuration with the Cosine similarity matching method in combination with the emotion information set and a threshold value of 0.775 performs best, resulting in a Weighted F-score of 1.431. The parameter information set underperforms for both the Euclidean distance and the Cosine similarity matching, compared to the emotion information set.

We will refer to the best configuration (emotion information set, Cosine similarity, threshold of 0.775) in the results above as Configuration 1 for future configurations.

Word substitution

In this configuration of SSCM, we substitute each word in a statement for their simple form in the vectorization method. A simple form word contains lowercase characters, contains no punctuation (such as hyphens, commas, etc.) and contains no diacritic character. We substitute characters with diacritics for their non-diacritics counterpart. For example we replace the character *ä* with the character *a*.

Table 5.5: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	118	118	236	0.443	0.516	1.402
0.775	115	127	242	0.441	0.539	**1.422**
0.800	103	141	244	0.416	0.563	1.395

Table 5.6: Cosine Similarity - Parameter information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.950	80	65	145	0.292	0.337	0.921
0.975	74	88	**162**	0.285	0.412	**0.983**

Table 5.7: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	101	133	234	0.403	0.544	1.350
0.800	88	164	252	0.385	0.604	**1.374**
0.825	61	190	251	0.302	0.616	1.220

Table 5.8: Euclidean Distance - Parameter information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	86	60	146	0.308	0.323	0.939
0.800	85	65	150	0.307	0.341	**0.956**
0.825	83	66	149	0.302	0.340	0.944

The result above show that the configuration with the cosine similarity matching method, emotion information set and a threshold value of 0.775 performs best with the word substitution method, with a Weighted F-score of 1.422 (Configuration 2). While it underperforms with compared to Configuration 1 (Weighted F-score of 1.431), Configuration 2 has the same amount of correct SOM's and sets up the possibility of using other methods described in later sections.

At this point we see clear evidence that the parameter information set underperforms compared to the emotion information set. For future configurations, we will not use the parameter information set.

Unique word statement vector

Each statement vector in the previous configuration is constructed from all words in a statement. In this configuration only unique words are added to a statement vector. The purpose of this configuration is to see if a multitude of same words decreases an intended sentiment.

Before we add each word to a statement vector, we first check a set if we already added the information set vector of a word to a statement vector. We only add a word if it is absent in the set.

Table 5.9: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	117	120	237	0.442	0.519	1.402
0.775	111	132	243	0.434	0.548	**1.415**
0.800	105	142	247	0.423	0.567	1.414

Table 5.10: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	102	134	236	0.407	0.547	1.361
0.800	87	165	252	0.382	0.600	**1.365**
0.825	65	189	254	0.318	0.615	1.250

The configuration (Configuration 3) containing the emotion information set, cosine similarity matching method and a threshold value of 0.775, performed best with the unique word statements. Compared to Configuration 1 and 2, it underperforms in SOM's (111 versus 115), but performs better for NOM's (132 versus 129 and 127).

We will use unique word statement vectors for future configurations as it allows for single words in a vector to be as influential as a word with multiple entries.

Stemming

For the next configuration, we use stems of words in the vectorization method. We use stems for finding a word vector for a word in a statement. If a word in a statement (both in a player statement option and a player input statement) does not exist in our corpus, we check if its stem exist in our stem corpus. We create a stem corpus by collecting the stems of each word in each statement in a scenario with the corresponding information vector for that statement, only if the stem of the word is not equal to the word itself. Since the non-stemmed word already has an entry in the 'normal' corpus, adding it to the stemmed corpus adds no value. For example a stem of 'blijft' equals 'blijven' and a stem of 'jij' equals 'jij'. We do not add the latter to our stem corpus, since the stem of the word 'jij' is equal to itself.

After we collected all the stems in a scenario, we average the collected information vectors and create a word vector for each stem.

We test two methods for stemming. The first method of stemming we use a database consisting a table obtained from the website Wikitionary[1]. Wiktionary is an open-source multilingual corpus, providing synonyms and stems.

The other method is the use of a Snowball stemmer. Snowball is a small string processing language designed for creating stemming algorithms for use in

[1] `https://www.wiktionary.org/`

information retrieval [30]. Snowball subtracts stems from a word via a pattern matching. We use the Accord.NET framework[2] for our implementation, which uses Snowball.

A difference between the two methods is a resulting stem. A Snowball stemmer gives different stems, as it uses a pattern to subtract a stem. The Wiktionary database contains stem collected from dictionaries. An example can be found with irregular verbs. For the verb 'gebleven' Snowball subtracts the stem 'geblev', while the stem in the Wiktionary database is 'blijven'

Wiktionary Database

Table 5.11: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	114	123	237	0.435	0.527	1.397
0.775	109	133	242	0.428	0.551	**1.407**
0.800	104	141	245	0.419	0.566	1.405

Table 5.12: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	109	115	224	0.414	0.503	1.331
0.775	104	132	236	0.412	0.542	**1.366**
0.800	85	164	249	0.374	0.596	1.345

Snowball stemmer

Table 5.13: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	118	119	237	0.444	0.523	1.410
0.775	111	132	243	0.434	0.552	**1.419**
0.800	105	139	244	0.421	0.563	1.404

Table 5.14: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	108	115	223	0.411	0.505	1.327
0.775	104	131	235	0.411	0.541	**1.363**
0.800	86	163	249	0.377	0.597	1.351

[2] http://accord-framework.net/

In the results above, we see that for both the Wiktionary Database stemming method and the Snowball stemmer, the cosine similarity works best as a matching method. The Wiktionary Database stemming method in combination with the cosine similarity matching method and a threshold value of 0.775 (Configuration 4) has a slightly higher amount of NOM's (133 versus 132), but has a lower amount of SOM's (109 versus 111) compared to the same configuration with the Snowball stemmer (Configuration 5).

We will consider stemming at future configurations, if it shows an increase in performance with a different application.

Stem All

For this configuration we apply the stemming methods to all words in a statement in the vectorization method. We substitute each word in a statement, both player statement options and player input statements, with their stem. We use the same stemming methods as the previous configuration: Wiktionary Database and Snowball Stemmer.

Wiktionary Database

Table 5.15: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.725	125	104	229	0.451	0.492	1.394
0.750	120	114	234	0.445	0.515	**1.405**
0.775	113	120	233	0.430	0.521	1.380

Table 5.16: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	112	115	227	0.423	0.508	1.353
0.800	97	141	238	0.397	0.564	**1.357**
0.825	70	160	230	0.316	0.563	1.195

Snowball stemmer

Table 5.17: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	114	113	227	0.427	0.502	1.356
0.775	109	128	237	0.424	0.540	**1.388**
0.800	103	135	238	0.411	0.549	1.371

Table 5.18: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	109	124	233	0.421	0.524	1.366
0.800	102	152	254	0.422	0.586	**1.430**
0.825	76	180	256	0.354	0.604	1.313

Looking at the result tables, we see two configurations that have a Weighted F-score of at least 1.400. The first configuration contains the Wiktionary Database stemming method, cosine similarity matching method and a threshold value of 0.750 (Configuration 6). The second configuration contains the Snowball stemmer, the euclidean distance matching method and a threshold value of 0.800 (Configuration 7). Configuration 6 has a higher amount of SOM's (120 vs 102), while Configuration 7 has more NOM's (114 versus 152).

We will take this method into consideration for future configurations or combination of methods.

Synonyms

In this configuration we use the synonyms of words in a statement in the vectorization method. We use the same order of selecting a corpus vector as we did with stemming: if our corpus does not contain a corpus vector for a word, we look if a synonym is present in our corpus and use the corresponding corpus vector. These vectors are stored in a synonym corpus.

We use two data sets for the synonyms: Wiktionary and Open Dutch Wordnet. Open Dutch Wordnet[31] is a data set with synonyms, containing 28727 rows of a lemma and its synonym(s). We concluded that a lemma is also a synonym of a synonym, as well as synonyms of a lemma are also synonyms of each other.

We joined both datasets to a single database containing 314858 lemma-synonym sets.

Table 5.19: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	116	116	232	0.435	0.514	1.385
0.775	110	128	238	0.427	0.539	**1.393**
0.800	102	141	243	0.413	0.567	1.393

Table 5.20: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	101	137	238	0.406	0.558	1.371
0.800	87	168	255	0.385	0.609	**1.379**
0.825	62	189	251	0.305	0.611	1.221

When comparing the results, the configuration utilizing synonyms, cosine similarity and a threshold value of 0.775 (Configuration 8), outperforms the configuration with euclidean distance matching method and a threshold value of 0.800 (Configuration 9) when comparing the Weighted F-score and SOM's. However, Configuration 9 has a higher amount NOM's (168 versus 128) compared to Configuration 8.

Second degree synonyms

For this configuration we use second degree synonyms in the vectorization method. A second degree synonym is a word with the same synonym of a different word. For example, the word "trap" is a synonym of the word "trede". The word "schop" is a synonym of the word "trap". The word "schop" is a second degree synonym of the word "trede". In the section Synonyms, the vectorization method uses a synonym of a word if said word is not present in the corpus. For this vectorization method we use a second degree synonym of a word if the word and its synonyms are not present in the synonym corpus.

For each second degree synonym, we check if it is present in the synonym corpus. We add all synonym vectors present to a temporary vector and divide it by the amount of synonym vectors found.

Table 5.21: Cosine Similarity - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0.775	110	127	237	0.426	0.539	1.392
0.800	102	142	244	0.414	0.574	**1.401**
0.825	91	154	245	0.364	0.596	1.324

Table 5.22: Euclidean Distance - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0.650	130	33	163	0.512	0.329	1.353
0.675	127	51	178	0.498	0.365	**1.361**
0.700	121	78	199	0.472	0.413	1.356

When looking at the results, we see that a small increase for the configuration with cosine similarity matching method with a threshold value of 0.800 (Configuration 10) compared to Configuration 8. An increase in SOM's can be seen with the euclidean distance matching method and a threshold value of 0.675, if we compare it to Configuration 9. However, the configuration also has a large decrease in NOM's as well as in Total Correct.

Scenario Statement Option Vector

Previously, we compared our player input statement vector to a player statement option vector calculated in the same manner as the input statement. For the configurations in this section, we compare the input statement vector to a vector containing only a emotional reaction property as stated in a scenario by a author. A scenario statement option vector contains only a value for a specified emotion, in contrast to a calculated statement option vector which contains a value for all emotion afflicted by words in a statement.

Table 5.23: Cosine Similarity - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0.500	120	99	219	0.433	0.477	1.344
0.525	117	105	222	0.429	0.492	**1.351**
0.550	112	109	221	0.418	0.497	1.332

Table 5.24: Euclidean Distance - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0.375	120	95	215	0.430	0.468	1.328
0.400	108	114	222	0.410	0.512	**1.332**
0.425	90	137	227	0.370	0.532	1.273

Looking at the Weighted F-scores in this method, we clearly see that it is lower than the previous configurations. We will not use this method, but will consider it in a different method.

Differential Angle Matching Method

For this configuration we test a matching method using both a scenario statement option vector and a calculated statement option vector. We use the cosine similarity matching method for both statement option vector with a player input statement vector. We also calculate the cosine similarity between the statement option vectors. We only take a statement option into consideration if it has the highest cosine similarity score between a calculated statement option vector and a player input statement of all statement options.

We determine that a statement option is considered a SOM if the cosine similarity between a calculated statement vector and an input statement vector is equal or larger than 0.8.

If this is not the case, we consider looking at the scenario statement option vector. If the cosine similarity of a scenario statement vector and a player input statement vector is higher than a given threshold, we consider it a threshold.

If both cases are false, we lastly check if how large the angle is between a scenario statement vector and a calculated statement vector. We determine that if

the angle between a scenario statement vector and a calculated statement vector is large, than the angle between a scenario statement vector and a player input vector should be small. This comparison ensures that if the similarity between a calculated statement vector and a scenario statement vector is low, a player input statement vector must be more similar to a scenario statement vector. We also compare the cosine similarity between a player input statement vector and a calculated statement vector with a given threshold. If both comparisons are true it is considered a SOM.

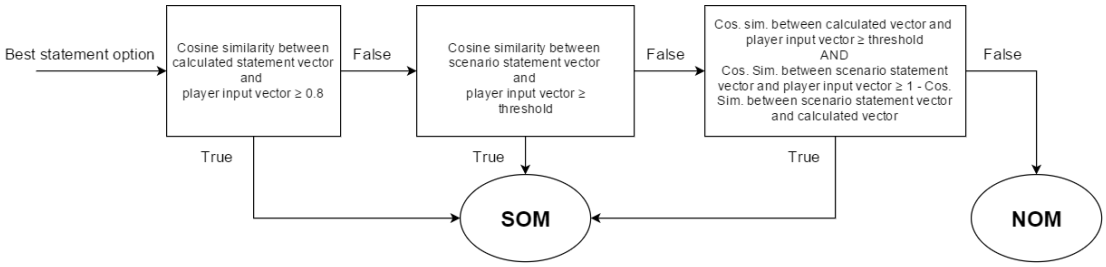


Figure 5.1: Flow chart of the differential angle threshold method

If all three cases are false, a player statement option is considered a NOM. Fig 5.1 shows a flow chart visualizing the process of determining if a statement option is a SOM or a NOM with the Differential Angle Method.

Table 5.25: Cosine Similarity - Emotion information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOM F_1-score**	**NOM F_1-score**	**Weighted F-score**
0.750	118	123	241	0.447	0.530	1.424
0.775	112	134	246	0.438	0.556	**1.433**
0.800	106	142	248	0.427	0.569	1.422

The results above show that the configuration with the Differential Angle threshold method with the emotion information set and a threshold value of 0.775 (Configuration 11) has the highest Weighted F-score. Compared to Configuration 1, it has a lower number of SOM's and a higher number of NOM's as well as a higher Weighted F-score.

We will consider this method in combination with other methods.

Combinations

In this section we combine multiple methods described before to find a best configuration. The methods that we combine are stemming, synonyms, second degree synonyms and Differential Angle threshold method. First we test the combination of synonyms and stems. Then we test the combination of Differential Angle threshold method with stemming. The third combination set is Differential Angle threshold method with synonyms. The last combination set contains stemming, synonyms, and Differential Angle threshold method.

The combination can be found in Appendix C, each combination has a unique number.

Synonyms and Stems

In this configuration we test the usage of stems and synonyms combined in the vectorization method. We test six sequences of stemming and synonyms. In the first sequence, if a vector for a word is not present in the corpus, we check if a word vector is present for its stem. If no vector is present for its stem, we check if a word vector is present in the synonym corpus (Combination 1). In the second sequence to swap the order of stem and synonym checking (Combination 2).

For the combination of stemming and synonyms, we only use the Wiktionary Database stemming method. Synonyms use Wiktionary as a source

Stemming → Synonyms

Combination 1

Table 5.26: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	117	115	232	0.437	0.513	1.388
0.775	110	128	238	0.427	0.539	**1.393**
0.800	100	139	239	0.405	0.559	1.369

Table 5.27: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	111	115	226	0.420	0.510	1.349
0.775	102	136	238	0.409	0.554	**1.372**
0.800	86	164	250	0.378	0.600	1.356

Synonyms → Stemming

Combination 2

Table 5.28: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	116	114	230	0.434	0.511	1.379
0.775	110	127	237	0.426	0.539	**1.392**
0.800	102	139	241	0.411	0.564	1.386

Table 5.29: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	111	113	224	0.418	0.504	1.341
0.775	103	135	238	0.411	0.553	**1.376**
0.800	87	165	252	0.382	0.602	1.367

For the third (Combination 3) and fourth (Combination 4) sequences of synonyms and stemming we also use second degree synonyms. We look at second degree synonyms directly after normal synonyms if no synonym is found. The orders are the same as the previous two sequences.

Stemming → Synonyms → Second degree Synonyms

Combination 3

Table 5.30: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	115	110	225	0.428	0.494	1.349
0.775	109	128	237	0.424	0.540	**1.388**
0.800	101	139	240	0.408	0.562	1.378

Table 5.31: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	108	118	226	0.413	0.519	1.345
0.775	100	134	234	0.401	0.546	**1.347**
0.800	84	165	249	0.372	0.596	1.339

Synonyms → Second degree Synonyms → Stemming

Combination 4

Table 5.32: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	110	127	237	0.426	0.539	1.392
0.800	102	142	244	0.414	0.574	**1.401**
0.825	91	154	245	0.391	0.574	1.356

Table 5.33: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	100	135	235	0.402	0.550	1.353
0.800	86	166	252	0.380	0.599	**1.359**
0.825	62	190	252	0.306	0.612	1.224

The fifth sequence (Combination 5) uses the Stem All method with synonyms. We stem all words in all statements and if our corpus does not contain a word, we check if it is a synonym in our synonym corpus.

The last sequence (Combination 6) for this combination, we also use second degree synonyms. If a synonym is also not available in our synonym corpus, we check for any second degree synonyms.

Stem All → Synonyms

Combination 5

Table 5.34: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.725	123	102	225	0.444	0.486	1.374
0.750	122	109	231	0.447	0.503	**1.397**
0.775	115	118	233	0.434	0.519	1.387

Table 5.35: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	113	116	229	0.426	0.511	1.364
0.800	100	140	240	0.406	0.562	**1.374**
0.825	74	162	236	0.333	0.572	1.238

Stem All → Synonyms → Second degree Synonyms

Combination 6

Table 5.36: Cosine Similarity - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	124	106	230	0.450	0.494	1.394
0.775	117	117	234	0.439	0.517	**1.395**
0.800	108	131	239	0.424	0.546	1.393

Table 5.37: Euclidean Distance - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	113	114	227	0.425	0.507	1.356
0.800	100	137	237	0.403	0.552	**1.359**
0.825	74	161	235	0.332	0.568	1.232

Looking at the results of the six sequences, four configurations stand out for us. The first configuration (Configuration 12) consists of Combination 1 with a threshold value of 0.775 while using cosine similarity as the matching method. The second configuration (Configuration 13) is Combination 4 with the cosine similarity matching method and a threshold value of 0.800. The third configuration (Configuration 14) consists of Combination 5 with a threshold value of 0.750 and the cosine similarity matching method. The last configuration (Configuration 15) consists Combination 6, cosine similarity matching method and a threshold value of 0.775.

Configuration 13 has the highest number of NOM's (142) compared to the other selected configurations in this section as well as the highest Weighted F-score. Configuration 14 has the highest number of SOM's (122) of the selected configurations. The other two configurations (Configurations 12 and 15) have a trade-off between SOM's and NOM's while having a Weighted F-score of at least 1.39.

Differential Angle threshold method and Stemming

In this section we test the combination of Differential Angle threshold method and stemming. We test two variants of stemming, as described in Sections *Stemming* (page 23) and *Stem All* (page 25). For both stemming methods, we use the Wiktionary Database and the Snowball stemming methods. The normal threshold method is replaced with Differential Angle threshold method.

Because Differential Angle threshold method uses only cosine similarity, we will not use euclidean distance as a matching method.

Differential Angle and Wiktionary Database Stemming

Combination 7

Table 5.38: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	116	124	240	0.442	0.533	1.417
0.775	111	134	245	0.435	0.557	**1.428**
0.800	106	141	247	0.426	0.570	1.421

Differential Angle and Snowball Stemming

Combination 8

Table 5.39: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	119	121	240	0.448	0.531	1.427
0.775	112	133	245	0.438	0.556	**1.431**
0.800	106	139	245	0.424	0.564	1.412

Differential Angle and Wiktionary Database Stem all

Combination 9

Table 5.40: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.725	119	104	223	0.434	0.481	1.350
0.750	115	114	229	0.431	0.508	**1.369**
0.775	111	120	231	0.424	0.519	1.367

Differential Angle and Snowball Stem All

Combination 10

Table 5.41: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.675	125	99	224	0.447	0.474	1.368
0.700	125	104	229	0.451	0.491	1.393
0.725	121	109	230	0.444	0.499	1.387
0.750	116	114	230	0.434	0.508	1.375
0.775	111	129	240	0.431	0.545	**1.408**
0.800	105	135	240	0.417	0.551	1.386

Looking at the results in this section we see three configurations that interest us. The first configuration (Configuration 16) consists of Combination 7 with a threshold value of 0.775. The second configuration (Configuration 17) consists of Combination 8 and a threshold value of 0.775. The last configurations (Configuration 18) consist of Combination 10 and a threshold value of 0.700. All selected configurations in this section use cosine similarity as matching method.

Configuration 16 has 1 more NOM compared to Configuration 17 (134 versus 133) while Configuration 17 has 1 more SOM (112 versus 111). Both configurations have a high Weighted F-score (1.428 and 1.431). Configuration 18 has a high number of SOM's (125). While the same combination with a threshold value of 0.775 has a higher Weighted F-score (1.408 versus 1.393), Configuration 16 has the same number of SOM's (111) while having a higher number of NOM's (134 versus 129).

Differential Angle threshold method and Synonyms

In this section we test the combination of Differential Angle threshold method and synonyms. We use the synonym methods as described in Sections *Synonyms* (page 26) and *Second Degree Synonyms* (page 27). We use the Differential Angle threshold method instead of the normal threshold method.

Differential Angle threshold method and Synonyms

Combination 11

Table 5.42: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	117	116	233	0.438	0.514	1.391
0.775	112	128	240	0.433	0.541	**1.408**
0.800	104	139	243	0.418	0.564	1.399

Differential Angle threshold method and Second Degree Synonyms

Combination 12

Table 5.43: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	112	126	238	0.432	0.538	1.402
0.800	104	140	244	0.419	0.570	**1.407**
0.825	104	140	244	0.419	0.570	1.407
0.850	103	140	243	0.415	0.569	1.400

Both combinations (Combinations 11 and 12) in this section have a interesting configuration. Combination 11 with a threshold value of 0.775 (Configuration 19) and Combination 12 with a threshold value of 0.800 (Configuration 20). Both configurations have a Weighted F-score of at least 1.400.

Differential Angle threshold, Stemming and Synonyms

In this section we use the variations described in Section *Stemming and Synonyms* (page 30) and replace the normal threshold method with Differential Angle threshold method.

Stemming → Synonyms

Combination 13

Table 5.44: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	118	114	232	0.439	0.510	1.389
0.775	112	127	239	0.432	0.538	**1.403**
0.800	102	137	239	0.410	0.556	1.375

Synonyms → Stemming

Combination 14

Table 5.45: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	117	113	230	0.436	0.508	1.379
0.775	112	126	238	0.432	0.538	**1.402**
0.800	104	137	241	0.416	0.560	1.392

Stemming → Synonyms → Second degree Synonyms

Combination 15

Table 5.46: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.750	116	109	225	0.430	0.492	1.351
0.775	111	127	238	0.429	0.539	**1.398**
0.800	103	137	240	0.413	0.558	1.384

Synonyms → Second degree Synonyms → Stemming

Combination 16

Table 5.47: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	112	126	238	0.432	0.538	1.402
0.800	104	140	244	0.419	0.570	**1.407**
0.825	103	140	243	0.415	0.569	1.400

Stem All → Synonyms

Combination 17

Table 5.48: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	113	116	229	0.426	0.509	1.362
0.800	107	128	235	0.418	0.537	**1.373**
0.975	107	128	235	0.418	0.537	1.373

Stem All → Synonyms → Second degree Synonyms

Combination 18

Table 5.49: Differential Angle - Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOM F_1-score	NOM F_1-score	Weighted F-score
0.775	115	115	230	0.432	0.507	1.370
0.800	108	129	237	0.422	0.540	**1.383**
0.975	108	129	237	0.422	0.540	1.383

In this section, we found two interesting configurations. The first configuration (Configuration 21) consists of Combination 13 with a threshold value of 0.775. Configuration 21 has a NOM match more compare to the configuration consisting of Combination 14 with a threshold value of 0.775 (127 versus 126). The second configuration consists of Combination 16 with a threshold value of 0.800 (Configuration 22).

Validation

In the previous sections we selected 22 configurations based on the Training test data. In this section we calculate the Weighted F-score based on the Validation test data. Based on the combined calculations of both the Training test data and Validation test data, we will select four configurations. A configuration with the highest combined number of SOM's, a configuration with the highest combined number of NOM's, a configuration with the highest combined number of Total Correct and a configuration with the highest combined Weighted F-score.

Table 5.50: Configuration Comparison

	Training			Validation		
Configuration	**SOM**	**NOM**	**Weighted F-score**	**SOM**	**NOM**	**Weighted F-score**
1	115	129	1.431	36	62	1.320
2	115	127	1.422	36	60	1.299
3	111	132	1.415	34	62	1.281
4	109	133	1.407	32	60	1.238
5	111	132	1.419	33	58	1.225
6	120	114	1.405	32	52	1.194
7	102	152	1.430	31	65	1.245
8	110	128	1.393	26	65	1.147
9	87	168	1.379	18	82	1.074
10	102	142	1.401	20	70	1.053
11	112	134	1.433	34	62	1.281
12	110	128	1.393	23	62	1.063
13	102	142	1.401	20	70	1.053
14	122	109	1.397	27	53	1.095
15	117	117	1.395	24	58	1.069
16	111	134	1.428	32	60	1.238
17	112	133	1.431	33	58	1.225
18	125	104	1.393	45	44	1.343
19	112	128	1.408	26	65	1.147
20	104	140	1.407	20	70	1.053
21	112	127	1.403	23	62	1.063
22	104	140	1.407	23	63	1.075

Looking at the comparison table of the selected configurations, we found three configurations of interest. The configuration with the highest number of SOM's is Configuration 18 (170 Correct SOM's). The configuration with the highest number of NOM's is Configuration 9 (250 Correct NOM's) and also has the highest Total Correct (355 SOM and NOM classifications). Configuration 1 has the highest combined Weighted F-score value.

We use these three configurations for comparison with other methods.

Comparison with ReaderBench

In this section we compare our best configurations of SSCM with several configurations of ReaderBench. We will test SSCM and ReaderBench with the Testing test data as a black box method. Ștefan Rușeți et al. have conducted a experiment with ReaderBench with the same test data before [9]. The variants used by ReaderBench are: SpaCy, Path length, Leacock-Chodorow [28], Wu-Palmer [7], String kernels and Average of SpaCy and string kernels.

The Testing data contains 147 SOM matches and 77 NOM matches, making a total of 224 tests. The results are shown in the table below.

			SOM		**NOM**		**All**	
Totals			147		77		224	
Method			#	%	#	%	#	%
RB	SpaCy		38	26	77	**100**	115	51
	Path length (WordNet)		25	17	74	96	99	44
	Leacock-Chodorow (WordNet)		19	13	74	96	93	42
	Wu-Palmer (WordNet)		19	13	76	99	95	42
	String kernels		72	49	64	77	136	61
	Average of SpaCy and string kernels		33	22	77	**100**	110	49
	Neural network		72	49	65	77	137	**61**
SSCM	Best SOM	Config 18.	74	**50**	36	47	110	49
	Best NOM	Config 9.	35	24	67	87	102	46
	Best Weighted F-score	Config 1.	64	44	49	64	113	50

Table 5.51: Comparison with ReaderBench

Looking at the results, SSCM performs best compared to ReaderBench with Configuration 18, resulting in the highest number of SOM's. ReaderBench performs best with variation SpaCy and Average of path length and string kernels for NOM's and for overall accuracy variation String kernels works best.

Discussion

The configuration with all words stemmed by a Snowball Stemmer, emotion information set, the Differential Angle matching method and a threshold value of 0.700 works best with respect to SOM values. It outperforms ReaderBench, which utilizes a neural network, with a small amount in terms of SOM matches. The other configurations of SSCM performed comparable to the non-neural network methods of ReaderBench.

We also see that the configuration with the highest SOM matches according to the Training test data, is the configuration that uses the *Stem All* method as described on page 25, euclidean distance as matching method and a threshold of 0.000 with the emotion information set. This configuration had 185 SOM matches out of 306, a percentage of 60.5%. We did not select the configuration as it had 0 NOM matches.

During our research, we omitted the use of the user-defined parameter information set. The information set in our test scenario contained two parameters. We suspect that the number of parameters affects the usability of the information set. For future testing, we could use a scenario with a larger amount of user-defined parameters.

The Validation data set is skewed towards NOM matches, while the Testing data set has more SOM matches than NOM matches. Because of the higher amount of NOM matches in the Validation data, we may have selected a suboptimal configuration for the Testing data set. In the future it is best if we create an evenly distributed test data for testing our methods.

SSCM is comparable to ReaderBench in Total Accuracy, while providing a method without much setup. SSCM can be adjusted to use the method with different types of scenarios, for example in games. ReaderBench requires a large data set of text and sentiment values to work, while SSCM only needs local data from a scenario.

Chapter 6

Conclusion & Future Work

This paper describes the research of a text-matching method using only data present in a scripted scenario. Using information extracted from a scripted scenario we find it possible to match a player input statement to a player statement option using only local information. Using the emotion property information in a Communicate! scenario, we created a method to use the emotion property information for matching a player input statement to a player statement option. We compared our method with an available method, ReaderBench [11].Overall, the configuration with all words stemmed by a Snowball Stemmer, emotion information set, the Differential Angle matching method and a threshold value of 0.700 is a slightly better method for matching SOM's compared to ReaderBench.

For future work, we propose the use of machine learnig methods. Machine learning is a method we omitted in Chapter 3. The argument against using machine learning was the adaptability of a scenario. An author could change a scenario, leading to restarting a machine learning method as previously learning strategies would be invalid. We propose the usage of classification methods for scenarios that are used for a longer period of time. An example of such method as Support Vector Machines [20]. A Support Vector Machine classifies a vector by calculating if a vector is on the positive or negative field of a multidimensional plane. A vector on the positive plane means the Support Vector Machine positively classifies a vector for that class. A class for a scenario would be a player statement option, where a vector would be a calculated player input statement.

For the use of machine learning methods, we propose a method for users of a scripted scenario in Communicate! to add a player statement option to an existing option. A player could add a statement if a statement option is similar, but not exactly how a player would express herself in a scenario step. By adding an option to add statements to a player statement option, we have a larger data set of words for our corpus. A larger data set would add to context of a player statement option, without losing its original sentiment.

Acknowledgements

This paper would not have existed without the support and patience of J. T. Jeuring and Raja Lala. They helped in the process of making this paper, even in lesser times. I would also like to thank my colleagues for their support and allowing me to take my time to finish this paper.

I also want to thank my fianc'e for the love and support she gave me, even in dire times.

Bibliography

[1] B. Abu Shawar and E. Atwell. Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics*, 10(4):489–516, 2005. Cited By :17.

[2] L. W. Anderson and D. R. Krathwohl. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, 2001. Cited By :3983.

[3] David B Aragon. Fuzzy string matcher, January 4 1994. US Patent 5,276,741.

[4] Agnese Augello, Manuel Gentile, and Frank Dignum. Social agents for learning in virtual environments. In *International Conference on Games and Learning Alliance*, pages 133–143. Springer, 2016.

[5] Agnese Augello, Manuel Gentile, and Frank Dignum. An overview of open-source chatbots social skills. In Sotiris Diplaris, Anna Satsiou, Asbjørn Følstad, Michail Vafopoulos, and Thomas Vilarinho, editors, *Internet Science*, pages 236–248, Cham, 2018. Springer International Publishing.

[6] Mikhail Bilenko and Raymond J Mooney. Adaptive duplicate detection using learnable string similarity measures. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 39–48. ACM, 2003.

[7] Alexander Budanitsky and Graeme Hirst. Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1):13–47, 2006.

[8] Osmar A Carvalho Júnior, Renato F Guimarães, Alan R Gillespie, Nilton C Silva, and Roberto AT Gomes. A new approach to change vector analysis using distance and similarity measures. *Remote Sensing*, 3(11):2473–2493, 2011.

[9] Ștefan Ruseți, Raja Lala, Gabriel Guțu-Robu, Mihai Dascălu, Johan Jeuring, and Marcell van Geest. Semantic matching of open texts to pre-scripted answers in dialogue-based learning. 2019.

[10] Gunther Cox. Chatterbot. `https://github.com/gunthercox/ChatterBot`. Accessed: May 2018.

[11] Mihai Dascalu, Philippe Dessus, Maryse Bianco, Stefan Trausan-Matu, and Aurélie Nardy. Mining texts, learner productions and strategies with readerbench. In *Educational Data Mining*, pages 345–377. Springer, 2014.

[12] Mihai Dascalu, Gabriel Gutu, Stefan Ruseti, Ionut Cristian Paraschiv, Philippe Dessus, Danielle S. McNamara, Scott A. Crossley, and Stefan Trausan-Matu. Readerbench: A multi-lingual framework for analyzing text complexity. In Élise Lavoué, Hendrik Drachsler, Katrien Verbert, Julien Broisin, and Mar Pérez-Sanagustín, editors, *Data Driven Approaches in Digital Education*, pages 495–499, Cham, 2017. Springer International Publishing.

[13] Susan T Dumais, George W Furnas, Thomas K Landauer, Scott Deerwester, and Richard Harshman. Using latent semantic analysis to improve access to textual information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 281–285. Acm, 1988.

[14] Uli Fechner, Lutz Franke, Steffen Renner, Petra Schneider, and Gisbert Schneider. Comparison of correlation vector methods for ligand-based similarity searching. *Journal of computer-aided molecular design*, 17(10):687–698, 2003.

[15] Susan Feldman. Nlp meets the jabberwocky: Natural language processing in information retrieval. *ONLINE-WESTON THEN WILTON-*, 23:62–73, 1999.

[16] Sudeep Gandhe, Nicolle Whitman, David Traum, and Ron Artstein. An integrated authoring tool for tactical questioning dialogue systems. In *6th IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, page 10, 2009.

[17] Patrick Gebhard, Gregor Mehlmann, and Michael Kipp. Visual scenemaker—a tool for authoring interactive virtual characters. *Journal on Multimodal User Interfaces*, 6(1-2):3–11, 2012.

[18] Gabriel Gutu, Mihai Dascalu, Stefan Trausan-Matu, and Philippe Dessus. Readerbench goes online: A comprehension-centered framework for educational purposes. 09 2016.

[19] William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. Inducing domain-specific sentiment lexicons from unlabeled corpora. *CoRR*, abs/1606.02820, 2016.

[20] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[21] Johan Jeuring, Frans Grosfeld, Bastiaan Heeren, Michiel Hulsbergen, Richta IJntema, Vincent Jonker, Nicole Mastenbroek, Maarten van der Smagt, Frank Wijmans, Majanne Wolters, and Henk van Zeijts. Communicate! — a serious game for communication skills —. In Gráinne Conole, Tomaž Klobučar, Christoph Rensing, Johannes Konert, and Elise Lavoué, editors, *Design for Teaching and Learning in a Networked World*, pages 513–517, Cham, 2015. Springer International Publishing.

[22] Richard V Kadison. The pythagorean theorem: I. the finite case. *Proceedings of the National Academy of Sciences*, 99(7):4178–4184, 2002.

[23] Elizabeth D Liddy. Natural language processing. 2001.

[24] DH Lie, J Hulstijn, R op den Akker, and A Nijholt. A transformational approach to natural language understanding in dialogue systems. *Proceedings Natural Language Processing and Industrial Applications (NLP+ IA'98)*, 2:163–168, 1998.

[25] Edward Loper and Steven Bird. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*, 2002.

[26] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

[27] Maria das Graças Bruno Marietto, Rafael Varago de Aguiar, Gislene de Oliveira Barbosa, Wagner Tanaka Botelho, Edson Pimentel, Robson dos Santos França, and Vera Lúcia da Silva. Artificial intelligence markup language: A brief tutorial. *arXiv preprint arXiv:1307.3091*, 2013.

[28] Rada Mihalcea, Courtney Corley, Carlo Strapparava, et al. Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI*, volume 6, pages 775–780, 2006.

[29] B. Nørgaard, J. Ammentorp, P. Kofoed, and K. O. Kyvik. Training improves inter-collegial communication. *Clinical Teacher*, 9(3):173–177, 2012. Cited By :2.

[30] Martin F Porter. Snowball: A language for stemming algorithms, 2001.

[31] Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vosse. Open dutch wordnet. In *Proceedings of the Eight Global Wordnet Conference*, Bucharest, Romania, 2016.

[32] Victoria Rubin, Yimin Chen, and Lynne Marie Thorimbert. Artificially intelligent conversational agents in libraries. 28:496–522, 11 2010.

[33] S. L. Ruyak, L. Migliaccio, A. Levi, and S. Patel. Role development in midwifery education: A place for simulation. *Midwifery*, 59:141–143, 2018.

[34] Yutaka Sasaki et al. The truth of the f-measure. *Teach Tutor mater*, 1(5):1–5, 2007.

[35] Bruce Wilcox. Chatscript documentation. `https://github.com/bwilcox-1234/ChatScript/`. Accessed: May 2018.

[36] Michael Wooldridge and Nicholas R Jennings. Intelligent agents: Theory and practice. *The knowledge engineering review*, 10(2):115–152, 1995.

[37] Jun Ye. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling*, 53(1-2):91–97, 2011.

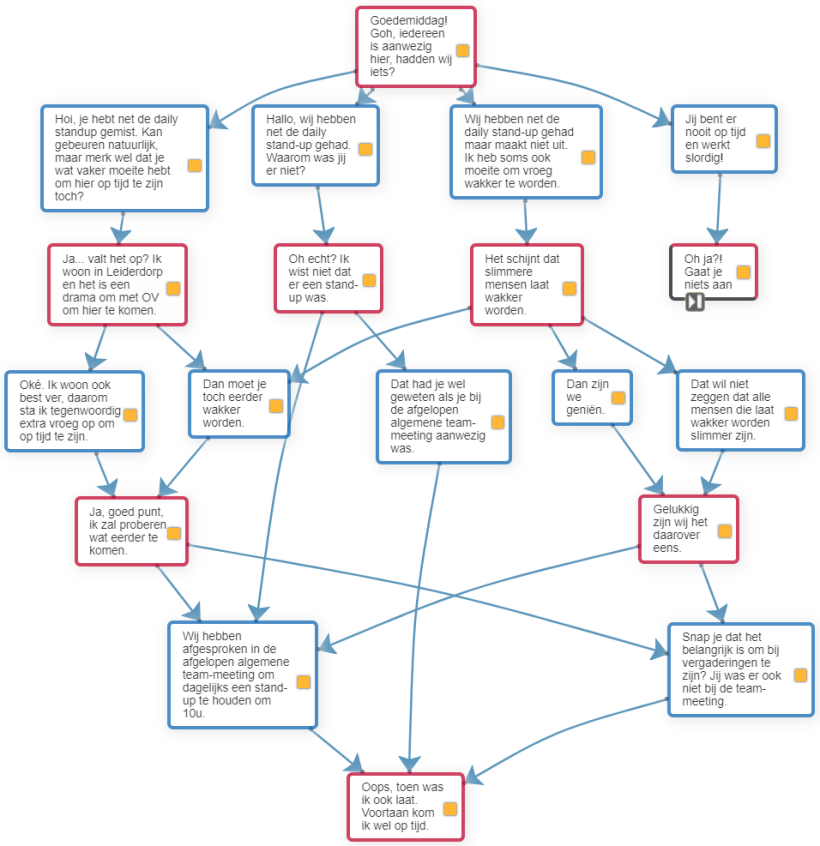
[38] Bo Yu, Zong-ben Xu, and Cheng-hua Li. Latent semantic analysis for text categorization using neural network. *Knowledge-Based Systems*, 21(8):900–904, 2008.

[39] Yin Zhang, Rong Jin, and Zhi-Hua Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, 2010.

Appendix A

Example Scenario

This graph shows the flow of a scenario. Red blocks represent computer statements and blue block player statements. Red/black blocks are scenario endings.



Appendix B

Example Scenario Specific Corpus

A selection from a corpus from a scenario, based on the emotion property information set.

```
<?xml version="1.0"?>
        <CBC_Storage xmlns:xsi="http://www.w3.org/2001/
            XMLSchema-instance" xmlns:xsd="http://www.w3.
            org/2001/XMLSchema">
        <ID>100</ID>
        <Corpus>
        <str_fla>
        <key>jij</key>
        <value>
        <float>0.166666672</float>
        <float>0</float>
        <float>0.166666672</float>
        <float>0.166666672</float>
        <float>0</float>
        <float>0.166666672</float>
        <float>0</float>
        <float>0.166666672</float>
        <float>0</float>
        <float>0</float>
        <float>0</float>
        <float>0.166666672</float>
        <float>0</float>
        <float>0</float>
        </value>
        </str_fla>
        <str_fla>
```

<key>bent</key>
<value>
<float>1</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
<str_fla>
<key>er</key>
<value>
<float>0.25</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0.25</float>
<float>0.25</float>
<float>0.25</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
<str_fla>
<key>nooit</key>
<value>
<float>1</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>

```
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
<str_fla>
<key>op</key>
<value>
<float>0.2</float>
<float>0.2</float>
<float>0</float>
<float>0</float>
<float>0.4</float>
<float>0</float>
<float>0.2</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
<str_fla>
<key>tijd</key>
<value>
<float>0.25</float>
<float>0.25</float>
<float>0</float>
<float>0</float>
<float>0.25</float>
<float>0</float>
<float>0.25</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
```

```
</value>
</str_fla>
</Corpus>
<Stems>
<str_fla>
<key>werken</key>
<value>
<float>0.25</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0.5</float>
<float>0.25</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
<str_fla>
<key>mist</key>
<value>
<float>0</float>
<float>1</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
</Stems>
<Statements>
<str_fla>
```

```
<key>jij bent er nooit op tijd en werkt slordig</
    key>
<value>
<float>1</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
<str_fla>
<key>hoi je hebt net de daily standup gemist kan
    gebeuren natuurlijk maar merk wel dat je wat
    vaker moeite hebt om hier op tijd te zijn toch
    </key>
<value>
<float>0</float>
<float>1</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
<float>0</float>
</value>
</str_fla>
</Statements>
<Parameters>
<str_int>
<key>anger</key>
<value>0</value>
```

```
</str_int>
<str_int>
<key>fear</key>
<value>1</value>
</str_int>
<str_int>
<key>surprise</key>
<value>2</value>
</str_int>
<str_int>
<key>happiness</key>
<value>3</value>
</str_int>
<str_int>
<key>interested</key>
<value>4</value>
</str_int>
<str_int>
<key>uninterested</key>
<value>5</value>
</str_int>
<str_int>
<key>trusting</key>
<value>6</value>
</str_int>
<str_int>
<key>yes</key>
<value>7</value>
</str_int>
<str_int>
<key>confused</key>
<value>8</value>
</str_int>
<str_int>
<key>sadness</key>
<value>9</value>
</str_int>
<str_int>
<key>littlesad</key>
<value>10</value>
</str_int>
<str_int>
<key>shock</key>
<value>11</value>
</str_int>
<str_int>
```

<key>no</key>
<value>12</value>
</str_int>
<str_int>
<key>base</key>
<value>13</value>
</str_int>
</Parameters>
<Average>
<float>0.01819728</float>
<float>0.08935903</float>
<float>0.0534633771</float>
<float>0.107425027</float>
<float>0.13611643</float>
<float>0.0234754626</float>
<float>0.165341839</float>
<float>0.192983881</float>
<float>0.0365478061</float>
<float>0.0320057124</float>
<float>0.0308614951</float>
<float>0.0587379448</float>
<float>0.045280613</float>
<float>0.0102040814</float>
</Average>
<Language>NL</Language>
</CBC_Storage>

Appendix C

Results

Base results

Cosine similarity

Table C.1: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**167**	30	197	27	197	112
0.025	167	30	197	27	197	112
0.050	167	30	197	27	197	112
0.075	166	30	196	29	197	111
0.100	164	31	195	31	196	111
0.125	160	34	194	35	193	111
0.150	153	36	189	42	191	111
0.175	152	36	188	43	191	111
0.200	151	38	189	44	189	111
0.225	151	42	193	44	185	111
0.250	150	42	192	45	185	111
0.275	150	42	192	45	185	111
0.300	147	45	192	48	182	111
0.325	146	45	191	49	182	111
0.350	144	45	189	52	182	110
0.375	144	46	190	54	181	108
0.400	144	51	195	54	176	108
0.425	144	54	198	54	173	108
0.450	143	55	198	55	172	108
0.475	140	58	198	59	169	107
0.500	138	60	198	62	167	106
0.525	136	62	198	65	165	105
0.550	135	65	200	68	162	103
0.575	133	69	202	72	158	101
0.600	132	76	208	75	151	99
0.625	131	80	211	78	147	97
0.650	130	88	218	82	139	94
0.675	129	95	224	86	132	91
0.700	126	103	229	92	124	88
0.725	122	112	234	99	115	85
0.750	118	120	238	112	107	76
0.775	115	129	244	118	98	73
0.800	104	139	243	132	88	70
0.825	94	148	242	147	79	65
0.850	79	168	247	172	59	55
0.875	68	181	249	201	46	37
0.900	60	197	**257**	217	30	29
0.925	47	210	257	242	17	17
0.950	26	219	245	273	8	7
0.975	15	**225**	240	290	2	1

Table C.2: Parameter information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**90**	42	132	70	185	146
0.025	90	42	132	70	185	146
0.050	90	42	132	70	185	146
0.075	90	42	132	70	185	146
0.100	90	42	132	70	185	146
0.125	90	42	132	70	185	146
0.150	90	42	132	70	185	146
0.175	90	42	132	70	185	146
0.200	90	42	132	70	185	146
0.225	90	42	132	70	185	146
0.250	90	42	132	70	185	146
0.275	90	42	132	70	185	146
0.300	90	42	132	70	185	146
0.325	90	42	132	70	185	146
0.350	90	42	132	70	185	146
0.375	90	42	132	70	185	146
0.400	90	42	132	70	185	146
0.425	90	42	132	70	185	146
0.450	90	42	132	70	185	146
0.475	90	42	132	70	185	146
0.500	90	42	132	70	185	146
0.525	90	42	132	70	185	146
0.550	90	42	132	70	185	146
0.575	90	42	132	70	185	146
0.600	90	42	132	70	185	146
0.625	90	42	132	70	185	146
0.650	90	43	133	70	184	146
0.675	90	43	133	70	184	146
0.700	90	43	133	70	184	146
0.725	90	46	136	70	181	146
0.750	89	46	135	71	181	146
0.775	89	47	136	71	180	146
0.800	89	48	137	71	179	146
0.825	87	51	138	75	176	144
0.850	85	53	138	79	174	142
0.875	84	58	142	82	169	140
0.900	84	61	145	84	166	138
0.925	83	63	146	89	164	134
0.950	80	66	146	94	161	132
0.975	74	**89**	**163**	112	138	120

Table C.3: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**56**	5	61	1	94	61
0.025	56	5	61	1	94	61
0.050	56	5	61	1	94	61
0.075	56	5	61	1	94	61
0.100	56	5	61	1	94	61
0.125	56	5	61	1	94	61
0.150	56	5	61	1	94	61
0.175	56	5	61	1	94	61
0.200	56	5	61	1	94	61
0.225	56	5	61	1	94	61
0.250	56	5	61	1	94	61
0.275	56	5	61	1	94	61
0.300	56	5	61	1	94	61
0.325	56	5	61	1	94	61
0.350	56	6	62	1	93	61
0.375	56	7	63	1	92	61
0.400	56	7	63	1	92	61
0.425	56	7	63	1	92	61
0.450	56	10	66	3	89	59
0.475	56	11	67	3	88	59
0.500	56	15	71	3	84	59
0.525	56	18	74	3	81	59
0.550	56	20	76	5	79	57
0.575	55	23	78	8	76	55
0.600	55	26	81	8	73	55
0.625	54	32	86	10	67	54
0.650	49	40	89	19	59	50
0.675	44	42	86	26	57	48
0.700	41	47	88	30	52	47
0.725	39	49	88	36	50	43
0.750	39	58	97	45	41	34
0.775	36	62	98	58	37	24
0.800	33	68	101	67	31	18
0.825	31	72	103	70	27	17
0.850	23	82	105	80	17	15
0.875	17	88	105	90	11	11
0.900	8	92	100	101	7	9
0.925	5	96	101	108	3	5
0.950	2	97	**99**	112	2	4
0.975	0	**99**	99	115	0	3

Euclidean distance

Table C.4: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**180**	0	180	2	227	124
0.025	180	0	180	2	227	124
0.050	180	0	180	2	227	124
0.075	180	0	180	2	227	124
0.100	180	0	180	2	227	124
0.125	180	0	180	2	227	124
0.150	180	0	180	2	227	124
0.175	180	0	180	2	227	124
0.200	180	0	180	2	227	124
0.225	177	1	178	5	226	124
0.250	166	1	167	16	226	124
0.275	159	4	163	23	223	124
0.300	154	13	167	28	214	124
0.325	139	15	154	43	212	124
0.350	137	15	152	45	212	124
0.375	135	16	151	47	211	124
0.400	134	17	151	48	210	124
0.425	130	17	147	53	210	123
0.450	130	17	147	53	210	123
0.475	129	19	148	54	208	123
0.500	128	19	147	55	208	123
0.525	126	19	145	57	208	123
0.550	126	19	145	57	208	123
0.575	126	19	145	57	208	123
0.600	126	19	145	58	208	122
0.625	125	20	145	59	207	122
0.650	125	22	147	59	205	122
0.675	125	31	156	62	196	119
0.700	119	45	164	76	182	111
0.725	115	58	173	88	169	103
0.750	110	73	183	100	154	96
0.775	104	108	212	116	119	86
0.800	84	148	232	156	79	66
0.825	60	184	244	208	43	38
0.850	45	209	**254**	245	18	16
0.875	32	219	251	266	8	8
0.900	14	224	238	291	3	1
0.925	12	**227**	239	294	0	0
0.950	12	227	239	294	0	0
0.975	11	227	238	295	0	0

Table C.5: Parameter information set - Training test data

Threshold	**SOM**	**NOM**	**Total Correct**	**SOMasNOM**	**NOMasSOM**	**WrongSOM**
0	**169**	0	169	0	227	137
0.025	169	0	169	0	227	137
0.050	169	0	169	0	227	137
0.075	169	0	169	0	227	137
0.100	169	0	169	0	227	137
0.125	169	0	169	0	227	137
0.150	169	0	169	0	227	137
0.175	169	0	169	0	227	137
0.200	169	0	169	0	227	137
0.225	169	0	169	0	227	137
0.250	169	0	169	0	227	137
0.275	169	0	169	0	227	137
0.300	169	0	169	0	227	137
0.325	169	0	169	0	227	137
0.350	169	0	169	0	227	137
0.375	169	0	169	0	227	137
0.400	169	0	169	0	227	137
0.425	169	0	169	0	227	137
0.450	169	0	169	0	227	137
0.475	169	0	169	0	227	137
0.500	169	0	169	0	227	137
0.525	169	2	171	0	225	137
0.550	169	2	171	0	225	137
0.575	169	2	171	0	225	137
0.600	169	3	172	1	224	136
0.625	169	3	172	1	224	136
0.650	169	3	172	2	224	135
0.675	168	3	171	3	224	135
0.700	166	4	170	5	223	135
0.725	165	6	171	7	221	134
0.750	161	21	182	13	206	132
0.775	159	26	185	16	201	131
0.800	158	31	189	19	196	129
0.825	152	38	190	28	189	126
0.850	145	43	188	38	184	123
0.875	138	49	187	52	178	116
0.900	131	75	206	71	152	104
0.925	104	99	203	112	128	90
0.950	75	149	224	167	78	64
0.975	44	**198**	**242**	236	29	26

Substituted words

Cosine similarity

Table C.6: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**168**	27	195	27	200	111
0.025	168	27	195	27	200	111
0.050	168	27	195	27	200	111
0.075	165	28	193	31	199	110
0.100	165	28	193	31	199	110
0.125	161	30	191	35	197	110
0.150	154	33	187	42	194	110
0.175	153	33	186	43	194	110
0.200	152	35	187	44	192	110
0.225	152	39	191	44	188	110
0.250	151	41	192	45	186	110
0.275	151	41	192	45	186	110
0.300	148	44	192	48	183	110
0.325	147	44	191	49	183	110
0.350	145	44	189	52	183	109
0.375	145	46	191	54	181	107
0.400	145	51	196	54	176	107
0.425	145	54	199	54	173	107
0.450	144	55	199	55	172	107
0.475	141	59	200	59	168	106
0.500	139	62	201	62	165	105
0.525	138	65	203	64	162	104
0.550	136	68	204	68	159	102
0.575	134	71	205	72	156	100
0.600	134	79	213	74	148	98
0.625	131	81	212	79	146	96
0.650	130	90	220	82	137	94
0.675	129	95	224	86	132	91
0.700	126	103	229	92	124	88
0.725	122	111	233	100	116	84
0.750	118	118	236	112	109	76
0.775	115	127	242	117	100	74
0.800	103	141	244	133	86	70
0.825	94	151	245	147	76	65
0.850	78	169	247	177	58	51
0.875	68	180	248	201	47	37
0.900	60	198	**258**	217	29	29
0.925	47	210	257	242	17	17
0.950	26	219	245	273	8	7
0.975	15	**225**	240	290	2	1

Table C.7: Parameter information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**90**	39	129	70	188	146
0.025	90	39	129	70	188	146
0.050	90	39	129	70	188	146
0.075	90	39	129	70	188	146
0.100	90	39	129	70	188	146
0.125	90	39	129	70	188	146
0.150	90	39	129	70	188	146
0.175	90	39	129	70	188	146
0.200	90	39	129	70	188	146
0.225	90	39	129	70	188	146
0.250	90	39	129	70	188	146
0.275	90	39	129	70	188	146
0.300	90	39	129	70	188	146
0.325	90	39	129	70	188	146
0.350	90	39	129	70	188	146
0.375	90	39	129	70	188	146
0.400	90	39	129	70	188	146
0.425	90	39	129	70	188	146
0.450	90	39	129	70	188	146
0.475	90	39	129	70	188	146
0.500	90	39	129	70	188	146
0.525	90	39	129	70	188	146
0.550	90	39	129	70	188	146
0.575	90	39	129	70	188	146
0.600	90	39	129	70	188	146
0.625	90	39	129	70	188	146
0.650	90	39	129	70	188	146
0.675	90	39	129	70	188	146
0.700	90	39	129	70	188	146
0.725	90	43	133	70	184	146
0.750	89	43	132	71	184	146
0.775	89	44	133	71	183	146
0.800	89	47	136	71	180	146
0.825	87	50	137	75	177	144
0.850	85	52	137	79	175	142
0.875	84	58	142	82	169	140
0.900	84	60	144	84	167	138
0.925	82	62	144	90	165	134
0.950	80	65	145	94	162	132
0.975	74	**88**	**162**	112	139	120

Table C.8: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**56**	4	60	1	95	61
0.025	56	4	60	1	95	61
0.050	56	4	60	1	95	61
0.075	56	4	60	1	95	61
0.100	56	4	60	1	95	61
0.125	56	4	60	1	95	61
0.150	56	4	60	1	95	61
0.175	56	4	60	1	95	61
0.200	56	4	60	1	95	61
0.225	56	4	60	1	95	61
0.250	56	4	60	1	95	61
0.275	56	4	60	1	95	61
0.300	56	4	60	1	95	61
0.325	56	4	60	1	95	61
0.350	56	5	61	1	94	61
0.375	56	6	62	1	93	61
0.400	56	6	62	2	93	60
0.425	56	6	62	2	93	60
0.450	56	9	65	4	90	58
0.475	56	9	65	4	90	58
0.500	56	12	68	4	87	58
0.525	56	16	72	4	83	58
0.550	56	20	76	5	79	57
0.575	55	22	77	7	77	56
0.600	55	24	79	7	75	56
0.625	54	27	81	10	72	54
0.650	49	36	85	19	63	50
0.675	44	41	85	26	58	48
0.700	41	45	86	29	54	48
0.725	38	48	86	35	51	45
0.750	38	58	96	46	41	34
0.775	36	60	96	58	39	24
0.800	33	68	101	67	31	18
0.825	32	72	104	70	27	16
0.850	24	79	103	80	20	14
0.875	19	87	**106**	88	12	11
0.900	8	91	99	103	8	7
0.925	6	95	101	109	4	3
0.950	2	97	99	113	2	3
0.975	0	**99**	99	115	0	3

Euclidean distance

Table C.9: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**179**	0	179	3	227	124
0.025	179	0	179	3	227	124
0.050	179	0	179	3	227	124
0.075	179	0	179	3	227	124
0.100	179	0	179	3	227	124
0.125	178	0	178	4	227	124
0.150	177	0	177	5	227	124
0.175	176	0	176	6	227	124
0.200	175	0	175	7	227	124
0.225	167	3	170	15	224	124
0.250	156	5	161	26	222	124
0.275	153	6	159	29	221	124
0.300	139	17	156	43	210	124
0.325	139	17	156	43	210	124
0.350	137	18	155	45	209	124
0.375	136	20	156	46	207	124
0.400	135	20	155	47	207	124
0.425	133	20	153	50	207	123
0.450	131	20	151	52	207	123
0.475	130	24	154	53	203	123
0.500	128	24	152	55	203	123
0.525	125	24	149	58	203	123
0.550	125	24	149	58	203	123
0.575	125	26	151	59	201	122
0.600	125	27	152	61	200	120
0.625	125	28	153	63	199	118
0.650	124	32	156	66	195	116
0.675	120	59	179	76	168	110
0.700	119	83	202	81	144	106
0.725	112	97	209	96	130	98
0.750	105	119	224	114	108	87
0.775	101	133	234	129	94	76
0.800	88	164	252	152	63	66
0.825	61	190	251	200	37	45
0.850	46	210	**256**	235	17	25
0.875	33	218	251	264	9	9
0.900	22	223	245	281	4	3
0.925	14	226	240	291	1	1
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Table C.10: Parameter information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**93**	15	108	67	212	146
0.025	93	15	108	67	212	146
0.050	93	15	108	67	212	146
0.075	93	15	108	67	212	146
0.100	93	15	108	67	212	146
0.125	93	15	108	67	212	146
0.150	93	15	108	67	212	146
0.175	93	15	108	67	212	146
0.200	93	15	108	67	212	146
0.225	93	15	108	67	212	146
0.250	93	15	108	67	212	146
0.275	93	15	108	67	212	146
0.300	93	15	108	67	212	146
0.325	93	15	108	67	212	146
0.350	93	15	108	67	212	146
0.375	93	15	108	67	212	146
0.400	93	15	108	67	212	146
0.425	93	15	108	67	212	146
0.450	93	22	115	67	205	146
0.475	93	22	115	67	205	146
0.500	93	22	115	67	205	146
0.525	93	37	130	67	190	146
0.550	93	37	130	68	190	145
0.575	93	37	130	68	190	145
0.600	93	39	132	69	188	144
0.625	93	41	134	69	186	144
0.650	93	45	138	71	182	142
0.675	91	45	136	74	182	141
0.700	90	49	139	75	178	141
0.725	89	51	140	78	176	139
0.750	87	55	142	81	172	138
0.775	86	60	146	84	167	136
0.800	85	65	150	89	162	132
0.825	83	66	149	95	161	128
0.850	78	75	153	106	152	122
0.875	75	92	167	115	135	116
0.900	70	112	182	134	115	102
0.925	58	145	203	172	82	76
0.950	37	174	211	215	53	54
0.975	18	**215**	**233**	268	12	20

Unique words

Cosine similarity

Table C.11: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**166**	27	193	27	200	113
0.025	166	27	193	27	200	113
0.050	166	27	193	27	200	113
0.075	163	28	191	31	199	112
0.100	163	28	191	31	199	112
0.125	160	30	190	34	197	112
0.150	152	33	185	42	194	112
0.175	151	33	184	43	194	112
0.200	150	34	184	44	193	112
0.225	150	38	188	44	189	112
0.250	149	41	190	45	186	112
0.275	149	41	190	45	186	112
0.300	146	44	190	48	183	112
0.325	145	44	189	49	183	112
0.350	143	45	188	52	182	111
0.375	143	46	189	54	181	109
0.400	143	51	194	54	176	109
0.425	143	55	198	54	172	109
0.450	142	56	198	55	171	109
0.475	139	61	200	59	166	108
0.500	138	65	203	61	162	107
0.525	136	67	203	65	160	105
0.550	135	70	205	67	157	104
0.575	133	75	208	73	152	100
0.600	133	82	215	75	145	98
0.625	130	83	213	79	144	97
0.650	127	93	220	87	134	92
0.675	125	99	224	92	128	89
0.700	124	105	229	97	122	85
0.725	121	112	233	105	115	80
0.750	117	120	237	115	107	74
0.775	111	132	243	123	95	72
0.800	105	142	247	132	85	69
0.825	96	156	252	145	71	65
0.850	79	168	247	174	59	53
0.875	69	185	254	199	42	38
0.900	60	199	**259**	218	28	28
0.925	44	209	253	247	18	15
0.950	27	218	245	273	9	6
0.975	15	**225**	240	290	2	1

Table C.12: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**56**	4	60	1	95	61
0.025	56	4	60	1	95	61
0.050	56	4	60	1	95	61
0.075	56	4	60	1	95	61
0.100	56	4	60	1	95	61
0.125	56	4	60	1	95	61
0.150	56	4	60	1	95	61
0.175	56	4	60	1	95	61
0.200	56	4	60	1	95	61
0.225	56	4	60	1	95	61
0.250	56	4	60	1	95	61
0.275	56	4	60	1	95	61
0.300	56	4	60	1	95	61
0.325	56	6	62	1	93	61
0.350	56	6	62	1	93	61
0.375	56	6	62	1	93	61
0.400	56	7	63	2	92	60
0.425	56	7	63	2	92	60
0.450	56	10	66	3	89	59
0.475	56	11	67	3	88	59
0.500	56	15	71	4	84	58
0.525	56	19	75	5	80	57
0.550	56	22	78	6	77	56
0.575	56	23	79	6	76	56
0.600	55	28	83	9	71	54
0.625	50	34	84	16	65	52
0.650	48	39	87	20	60	50
0.675	41	41	82	30	58	47
0.700	39	45	84	32	54	47
0.725	37	51	88	40	48	41
0.750	37	61	98	47	38	34
0.775	34	62	96	60	37	24
0.800	32	69	101	68	30	18
0.825	30	72	102	72	27	16
0.850	24	79	**103**	80	20	14
0.875	18	85	103	90	14	10
0.900	8	90	98	103	9	7
0.925	6	95	101	108	4	4
0.950	3	97	100	112	2	3
0.975	0	**99**	99	115	0	3

Euclidean distance

Table C.13: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**179**	0	179	3	227	124
0.025	179	0	179	3	227	124
0.050	179	0	179	3	227	124
0.075	179	0	179	3	227	124
0.100	179	0	179	3	227	124
0.125	178	0	178	4	227	124
0.150	177	0	177	5	227	124
0.175	176	0	176	6	227	124
0.200	175	0	175	7	227	124
0.225	167	3	170	15	224	124
0.250	156	6	162	26	221	124
0.275	153	6	159	29	221	124
0.300	139	16	155	43	211	124
0.325	139	17	156	43	210	124
0.350	137	18	155	45	209	124
0.375	136	20	156	46	207	124
0.400	136	20	156	46	207	124
0.425	134	20	154	49	207	123
0.450	131	20	151	52	207	123
0.475	130	24	154	53	203	123
0.500	129	24	153	54	203	123
0.525	125	24	149	58	203	123
0.550	125	24	149	58	203	123
0.575	125	26	151	59	201	122
0.600	125	27	152	61	200	120
0.625	124	28	152	64	199	118
0.650	123	34	157	68	193	115
0.675	120	62	182	80	165	106
0.700	118	87	205	85	140	103
0.725	110	100	210	100	127	96
0.750	106	118	224	116	109	84
0.775	102	134	236	129	93	75
0.800	87	165	252	158	62	61
0.825	65	189	254	199	38	42
0.850	47	209	**256**	235	18	24
0.875	33	218	251	265	9	8
0.900	20	223	243	283	4	3
0.925	14	226	240	291	1	1
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Stemming

Wiktionary Database

Cosine similarity

Table C.14: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**164**	27	191	27	200	115
0.025	164	27	191	27	200	115
0.050	164	27	191	27	200	115
0.075	161	28	189	31	199	114
0.100	161	28	189	31	199	114
0.125	156	31	187	36	196	114
0.150	149	33	182	43	194	114
0.175	148	33	181	44	194	114
0.200	148	34	182	44	193	114
0.225	148	39	187	44	188	114
0.250	147	42	189	45	185	114
0.275	147	42	189	45	185	114
0.300	144	45	189	48	182	114
0.325	143	45	188	49	182	114
0.350	141	46	187	52	181	113
0.375	141	48	189	54	179	111
0.400	141	51	192	54	176	111
0.425	141	55	196	54	172	111
0.450	140	57	197	56	170	110
0.475	137	62	199	60	165	109
0.500	137	66	203	62	161	107
0.525	135	69	204	64	158	107
0.550	134	72	206	67	155	105
0.575	133	76	209	72	151	101
0.600	133	82	215	74	145	99
0.625	131	83	214	76	144	99
0.650	126	92	218	86	135	94
0.675	123	99	222	94	128	89
0.700	122	104	226	98	123	86
0.725	119	112	231	105	115	82
0.750	114	123	237	117	104	75
0.775	109	133	242	123	94	74
0.800	104	141	245	130	86	72
0.825	90	152	242	151	75	65
0.850	80	165	245	176	62	50
0.875	70	181	251	200	46	36
0.900	61	195	**256**	218	32	27
0.925	43	205	248	246	22	17
0.950	26	216	242	274	11	6
0.975	16	**223**	239	289	4	1

Table C.15: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**49**	4	53	1	95	68
0.025	49	4	53	1	95	68
0.050	49	4	53	1	95	68
0.075	49	4	53	1	95	68
0.100	49	4	53	1	95	68
0.125	49	4	53	1	95	68
0.150	49	4	53	1	95	68
0.175	49	4	53	1	95	68
0.200	49	4	53	1	95	68
0.225	49	4	53	1	95	68
0.250	49	4	53	1	95	68
0.275	49	4	53	1	95	68
0.300	49	4	53	1	95	68
0.325	49	6	55	1	93	68
0.350	49	6	55	1	93	68
0.375	49	6	55	1	93	68
0.400	49	7	56	1	92	68
0.425	49	7	56	1	92	68
0.450	49	10	59	2	89	67
0.475	49	11	60	2	88	67
0.500	49	16	65	3	83	66
0.525	49	21	70	4	78	65
0.550	49	24	73	5	75	64
0.575	48	25	73	7	74	63
0.600	47	31	78	10	68	61
0.625	44	35	79	14	64	60
0.650	43	40	83	18	59	57
0.675	38	43	81	27	56	53
0.700	38	46	84	28	53	52
0.725	33	51	84	37	48	48
0.750	33	58	91	45	41	40
0.775	32	60	92	55	39	31
0.800	30	69	99	63	30	25
0.825	29	76	**105**	67	23	22
0.850	23	82	105	76	17	19
0.875	18	87	105	88	12	12
0.900	9	91	100	101	8	8
0.925	7	96	103	107	3	4
0.950	2	98	100	113	1	3
0.975	0	**99**	99	115	0	3

Euclidean distance

Table C.16: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**183**	0	183	4	227	119
0.025	183	0	183	4	227	119
0.050	183	0	183	4	227	119
0.075	183	0	183	4	227	119
0.100	183	0	183	4	227	119
0.125	182	0	182	5	227	119
0.150	181	0	181	6	227	119
0.175	181	0	181	6	227	119
0.200	176	0	176	11	227	119
0.225	171	2	173	16	225	119
0.250	160	6	166	27	221	119
0.275	156	7	163	31	220	119
0.300	144	16	160	43	211	119
0.325	144	17	161	43	210	119
0.350	142	18	160	45	209	119
0.375	141	20	161	46	207	119
0.400	141	20	161	46	207	119
0.425	139	20	159	49	207	118
0.450	136	20	156	52	207	118
0.475	135	24	159	53	203	118
0.500	134	24	158	54	203	118
0.525	130	24	154	58	203	118
0.550	130	24	154	58	203	118
0.575	130	27	157	59	200	117
0.600	130	28	158	61	199	115
0.625	129	28	157	64	199	113
0.650	128	34	162	68	193	110
0.675	124	64	188	80	163	102
0.700	119	87	206	89	140	98
0.725	114	97	211	99	130	93
0.750	109	115	224	115	112	82
0.775	104	132	236	128	95	74
0.800	85	164	249	159	63	62
0.825	65	188	253	200	39	41
0.850	47	207	**254**	236	20	23
0.875	33	218	251	264	9	9
0.900	21	224	245	282	3	3
0.925	14	226	240	291	1	1
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Snowball stemmer

Cosine similarity

Table C.17: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**165**	27	192	27	200	114
0.025	165	27	192	27	200	114
0.050	165	27	192	27	200	114
0.075	162	28	190	31	199	113
0.100	162	29	191	31	198	113
0.125	159	31	190	34	196	113
0.150	151	33	184	42	194	113
0.175	150	33	183	43	194	113
0.200	149	34	183	44	193	113
0.225	149	37	186	44	190	113
0.250	148	40	188	45	187	113
0.275	148	40	188	45	187	113
0.300	145	43	188	48	184	113
0.325	144	43	187	49	184	113
0.350	142	43	185	52	184	112
0.375	142	44	186	54	183	110
0.400	142	49	191	54	178	110
0.425	142	54	196	54	173	110
0.450	141	55	196	55	172	110
0.475	138	59	197	59	168	109
0.500	138	63	201	60	164	108
0.525	135	66	201	65	161	106
0.550	134	71	205	67	156	105
0.575	132	78	210	73	149	101
0.600	132	83	215	76	144	98
0.625	129	84	213	80	143	97
0.650	127	94	221	86	133	93
0.675	125	99	224	91	128	90
0.700	124	105	229	95	122	87
0.725	121	111	232	101	116	84
0.750	118	119	237	109	108	79
0.775	111	132	243	119	95	76
0.800	105	139	244	128	88	73
0.825	95	154	249	143	73	68
0.850	81	167	248	170	60	55
0.875	71	182	253	197	45	38
0.900	62	196	**258**	218	31	26
0.925	44	206	250	247	21	15
0.950	27	217	244	273	10	6
0.975	15	**224**	239	290	3	1

Table C.18: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**55**	4	59	1	95	62
0.025	55	4	59	1	95	62
0.050	55	4	59	1	95	62
0.075	55	4	59	1	95	62
0.100	55	4	59	1	95	62
0.125	55	4	59	1	95	62
0.150	55	4	59	1	95	62
0.175	55	4	59	1	95	62
0.200	55	4	59	1	95	62
0.225	55	4	59	1	95	62
0.250	55	4	59	1	95	62
0.275	55	4	59	1	95	62
0.300	55	4	59	1	95	62
0.325	55	6	61	1	93	62
0.350	55	6	61	1	93	62
0.375	55	6	61	1	93	62
0.400	55	6	61	2	93	61
0.425	55	6	61	2	93	61
0.450	55	10	65	3	89	60
0.475	55	12	67	3	87	60
0.500	55	15	70	5	84	58
0.525	55	20	75	7	79	56
0.550	55	24	79	7	75	56
0.575	55	26	81	8	73	55
0.600	53	29	82	13	70	52
0.625	50	34	84	17	65	51
0.650	48	40	88	20	59	50
0.675	41	42	83	30	57	47
0.700	39	47	86	33	52	46
0.725	34	51	85	43	48	41
0.750	34	58	92	50	41	34
0.775	33	58	91	59	41	26
0.800	31	66	97	66	33	21
0.825	29	72	101	70	27	19
0.850	23	76	99	79	23	16
0.875	17	85	**102**	91	14	10
0.900	8	89	97	103	10	7
0.925	6	95	101	108	4	4
0.950	4	97	101	111	2	3
0.975	1	**99**	100	114	0	3

Euclidean distance

Table C.19: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**182**	0	182	3	227	121
0.025	182	0	182	3	227	121
0.050	182	0	182	3	227	121
0.075	182	0	182	3	227	121
0.100	182	0	182	3	227	121
0.125	181	0	181	4	227	121
0.150	180	0	180	5	227	121
0.175	179	0	179	6	227	121
0.200	178	0	178	7	227	121
0.225	170	3	173	15	224	121
0.250	159	6	165	26	221	121
0.275	156	6	162	29	221	121
0.300	142	16	158	43	211	121
0.325	142	17	159	43	210	121
0.350	140	18	158	45	209	121
0.375	139	20	159	46	207	121
0.400	138	20	158	47	207	121
0.425	136	20	156	50	207	120
0.450	134	20	154	52	207	120
0.475	133	23	156	53	204	120
0.500	132	23	155	54	204	120
0.525	128	23	151	58	204	120
0.550	128	23	151	58	204	120
0.575	128	26	154	59	201	119
0.600	128	27	155	61	200	117
0.625	127	28	155	64	199	115
0.650	126	31	157	68	196	112
0.675	123	60	183	79	167	104
0.700	120	85	205	85	142	101
0.725	112	98	210	98	129	96
0.750	108	115	223	113	112	85
0.775	104	131	235	126	96	76
0.800	86	163	249	156	64	64
0.825	66	189	255	199	38	41
0.850	47	209	**256**	236	18	23
0.875	33	217	250	265	10	8
0.900	20	224	244	283	3	3
0.925	14	226	240	291	1	1
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Stem All

Wiktionary Database

Cosine similarity

Table C.20: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**168**	26	194	25	201	113
0.025	168	26	194	25	201	113
0.050	167	27	194	27	200	112
0.075	164	28	192	30	199	112
0.100	163	28	191	31	199	112
0.125	153	30	183	41	197	112
0.150	152	33	185	42	194	112
0.175	151	33	184	43	194	112
0.200	151	35	186	43	192	112
0.225	149	35	184	45	192	112
0.250	149	38	187	45	189	112
0.275	149	40	189	45	187	112
0.300	148	43	191	46	184	112
0.325	146	43	189	48	184	112
0.350	144	44	188	50	183	112
0.375	143	45	188	51	182	112
0.400	143	46	189	52	181	111
0.425	143	48	191	54	179	109
0.450	142	54	196	55	173	109
0.475	141	57	198	57	170	108
0.500	141	62	203	58	165	107
0.525	138	64	202	61	163	107
0.550	137	67	204	63	160	106
0.575	137	71	208	65	156	104
0.600	135	74	209	68	153	103
0.625	135	80	215	72	147	99
0.650	132	82	214	79	145	95
0.675	130	90	220	83	137	93
0.700	129	96	225	86	131	91
0.725	125	104	229	92	123	89
0.750	120	114	234	102	113	84
0.775	113	120	233	114	107	79
0.800	104	129	233	125	98	77
0.825	92	141	233	140	86	74
0.850	89	156	245	160	71	57
0.875	75	171	246	187	56	44
0.900	61	193	**254**	215	34	30
0.925	42	202	244	248	25	16
0.950	25	214	239	276	13	5
0.975	15	**224**	239	291	3	0

Table C.21: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**48**	3	51	1	96	69
0.025	48	3	51	1	96	69
0.050	48	3	51	1	96	69
0.075	48	3	51	1	96	69
0.100	48	3	51	1	96	69
0.125	48	3	51	1	96	69
0.150	48	3	51	1	96	69
0.175	48	3	51	1	96	69
0.200	48	3	51	1	96	69
0.225	48	3	51	1	96	69
0.250	48	3	51	1	96	69
0.275	48	3	51	1	96	69
0.300	48	3	51	1	96	69
0.325	48	3	51	1	96	69
0.350	48	3	51	1	96	69
0.375	48	3	51	1	96	69
0.400	48	3	51	1	96	69
0.425	48	3	51	1	96	69
0.450	48	6	54	1	93	69
0.475	48	9	57	2	90	68
0.500	48	13	61	2	86	68
0.525	48	13	61	2	86	68
0.550	47	17	64	4	82	67
0.575	47	23	70	7	76	64
0.600	47	24	71	9	75	62
0.625	45	27	72	15	72	58
0.650	43	34	77	18	65	57
0.675	41	38	79	21	61	56
0.700	36	42	78	30	57	52
0.725	36	45	81	35	54	47
0.750	32	52	84	40	47	46
0.775	30	57	87	49	42	39
0.800	28	60	88	55	39	35
0.825	25	69	94	64	30	29
0.850	23	82	105	71	17	24
0.875	18	88	**106**	83	11	17
0.900	12	89	101	96	10	10
0.925	6	94	100	106	5	6
0.950	4	97	101	110	2	4
0.975	0	**99**	99	115	0	3

Euclidean distance

Table C.22: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**184**	0	184	4	227	118
0.025	184	0	184	4	227	118
0.050	184	0	184	4	227	118
0.075	184	0	184	4	227	118
0.100	184	0	184	4	227	118
0.125	183	0	183	5	227	118
0.150	182	0	182	6	227	118
0.175	182	0	182	6	227	118
0.200	177	0	177	11	227	118
0.225	176	2	178	12	225	118
0.250	164	5	169	24	222	118
0.275	156	7	163	32	220	118
0.300	147	15	162	41	212	118
0.325	145	17	162	43	210	118
0.350	143	18	161	45	209	118
0.375	142	19	161	46	208	118
0.400	140	20	160	48	207	118
0.425	137	20	157	52	207	117
0.450	136	20	156	53	207	117
0.475	136	20	156	53	207	117
0.500	134	23	157	55	204	117
0.525	132	23	155	57	204	117
0.550	132	23	155	57	204	117
0.575	132	23	155	58	204	116
0.600	131	26	157	59	201	116
0.625	131	27	158	61	200	114
0.650	131	30	161	62	197	113
0.675	128	49	177	67	178	111
0.700	127	67	194	74	160	105
0.725	122	86	208	85	141	99
0.750	118	102	220	96	125	92
0.775	112	115	227	111	112	83
0.800	97	141	238	132	86	77
0.825	70	160	230	181	67	55
0.850	51	194	245	220	33	35
0.875	38	212	**250**	251	15	17
0.900	21	220	241	280	7	5
0.925	15	227	242	291	0	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Snowball stemmer

Cosine similarity

Table C.23: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**168**	27	195	26	200	112
0.025	168	27	195	26	200	112
0.050	168	27	195	26	200	112
0.075	165	28	193	30	199	111
0.100	164	29	193	31	198	111
0.125	156	31	187	39	196	111
0.150	154	33	187	41	194	111
0.175	153	33	186	42	194	111
0.200	152	34	186	43	193	111
0.225	152	34	186	43	193	111
0.250	149	40	189	46	187	111
0.275	149	40	189	46	187	111
0.300	148	40	188	47	187	111
0.325	148	40	188	47	187	111
0.350	145	40	185	50	187	111
0.375	144	41	185	52	186	110
0.400	143	45	188	53	182	110
0.425	143	51	194	54	176	109
0.450	141	53	194	56	174	109
0.475	140	54	194	57	173	109
0.500	140	59	199	58	168	108
0.525	136	64	200	64	163	106
0.550	135	68	203	65	159	106
0.575	133	75	208	68	152	105
0.600	131	78	209	74	149	101
0.625	130	80	210	78	147	98
0.650	126	88	214	86	139	94
0.675	124	96	220	90	131	92
0.700	124	102	226	91	125	91
0.725	120	107	227	100	120	86
0.750	114	113	227	110	114	82
0.775	109	128	237	119	99	78
0.800	103	135	238	130	92	73
0.825	92	149	241	145	78	69
0.850	84	163	247	167	64	55
0.875	74	180	254	194	47	38
0.900	66	191	**257**	212	36	28
0.925	46	206	252	247	21	13
0.950	26	215	241	275	12	5
0.975	16	**224**	240	289	3	1

Euclidean distance

Table C.24: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**185**	0	185	3	227	118
0.025	185	0	185	3	227	118
0.050	185	0	185	3	227	118
0.075	185	0	185	3	227	118
0.100	185	0	185	3	227	118
0.125	184	0	184	4	227	118
0.150	183	0	183	5	227	118
0.175	182	0	182	6	227	118
0.200	180	0	180	8	227	118
0.225	167	3	170	21	224	118
0.250	162	6	168	26	221	118
0.275	157	6	163	31	221	118
0.300	146	13	159	42	214	118
0.325	146	14	160	42	213	118
0.350	143	15	158	45	212	118
0.375	142	19	161	46	208	118
0.400	141	20	161	47	207	118
0.425	138	20	158	51	207	117
0.450	136	20	156	53	207	117
0.475	135	20	155	54	207	117
0.500	133	23	156	56	204	117
0.525	131	23	154	58	204	117
0.550	131	23	154	58	204	117
0.575	131	26	157	58	201	117
0.600	131	27	158	61	200	114
0.625	131	27	158	63	200	112
0.650	130	28	158	64	199	112
0.675	126	55	181	73	172	107
0.700	123	77	200	82	150	101
0.725	117	95	212	94	132	95
0.750	113	110	223	104	117	89
0.775	109	124	233	122	103	75
0.800	102	152	254	140	75	64
0.825	76	180	256	189	47	41
0.850	52	205	**257**	233	22	21
0.875	37	217	254	261	10	8
0.900	20	224	244	284	3	2
0.925	14	226	240	292	1	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Table C.25: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**52**	0	52	0	99	66
0.025	52	0	52	0	99	66
0.050	52	0	52	0	99	66
0.075	52	0	52	0	99	66
0.100	52	0	52	0	99	66
0.125	52	0	52	0	99	66
0.150	52	0	52	0	99	66
0.175	52	0	52	0	99	66
0.200	52	0	52	0	99	66
0.225	52	0	52	0	99	66
0.250	52	0	52	0	99	66
0.275	52	0	52	0	99	66
0.300	52	0	52	0	99	66
0.325	52	0	52	0	99	66
0.350	52	0	52	0	99	66
0.375	52	0	52	0	99	66
0.400	52	0	52	0	99	66
0.425	52	0	52	0	99	66
0.450	52	0	52	0	99	66
0.475	52	0	52	0	99	66
0.500	52	0	52	0	99	66
0.525	52	0	52	0	99	66
0.550	52	0	52	0	99	66
0.575	52	1	53	0	98	66
0.600	52	3	55	0	96	66
0.625	51	6	57	3	93	64
0.650	51	10	61	5	89	62
0.675	48	24	72	11	75	59
0.700	45	33	78	22	66	51
0.725	42	46	88	28	53	48
0.750	34	56	90	50	43	34
0.775	31	70	101	63	29	24
0.800	23	83	106	77	16	18
0.825	18	89	**107**	92	10	8
0.850	13	92	105	99	7	6
0.875	5	98	103	109	1	4
0.900	2	98	100	114	1	2
0.925	1	99	100	116	0	1
0.950	0	99	99	118	0	0
0.975	0	**99**	99	118	0	0

Synonyms

Cosine similarity

Table C.26: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**170**	20	190	21	207	115
0.025	170	20	190	21	207	115
0.050	169	21	190	22	206	115
0.075	165	21	186	27	206	114
0.100	165	23	188	27	204	114
0.125	162	26	188	30	201	114
0.150	152	28	180	40	199	114
0.175	151	30	181	41	197	114
0.200	151	31	182	41	196	114
0.225	151	34	185	41	193	114
0.250	149	37	186	43	190	114
0.275	148	37	185	44	190	114
0.300	146	40	186	46	187	114
0.325	145	40	185	48	187	113
0.350	143	41	184	51	186	112
0.375	142	42	184	53	185	111
0.400	142	46	188	54	181	110
0.425	142	47	189	54	180	110
0.450	141	52	193	55	175	110
0.475	137	55	192	60	172	109
0.500	137	60	197	61	167	108
0.525	136	65	201	62	162	108
0.550	135	66	201	64	161	107
0.575	132	69	201	68	158	106
0.600	130	75	205	76	152	100
0.625	127	80	207	79	147	100
0.650	124	89	213	85	138	97
0.675	124	95	219	89	132	93
0.700	124	98	222	95	129	87
0.725	120	105	225	104	122	82
0.750	116	116	232	108	111	82
0.775	110	128	238	120	99	76
0.800	102	141	243	129	86	75
0.825	90	155	245	149	72	67
0.850	79	174	253	173	53	54
0.875	68	187	**255**	201	40	37
0.900	56	199	255	223	28	27
0.925	42	212	254	252	15	12
0.950	22	222	244	278	5	6
0.975	14	**225**	239	292	2	0

Table C.27: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**49**	3	52	1	96	68
0.025	49	3	52	1	96	68
0.050	49	3	52	1	96	68
0.075	49	3	52	1	96	68
0.100	49	3	52	1	96	68
0.125	49	3	52	1	96	68
0.150	49	3	52	1	96	68
0.175	49	3	52	1	96	68
0.200	49	3	52	1	96	68
0.225	49	3	52	1	96	68
0.250	49	3	52	1	96	68
0.275	49	3	52	1	96	68
0.300	49	3	52	1	96	68
0.325	49	4	53	1	95	68
0.350	49	4	53	1	95	68
0.375	49	4	53	1	95	68
0.400	49	5	54	2	94	67
0.425	49	5	54	2	94	67
0.450	49	7	56	2	92	67
0.475	49	9	58	2	90	67
0.500	49	12	61	2	87	67
0.525	49	14	63	3	85	66
0.550	49	18	67	4	81	65
0.575	48	21	69	5	78	65
0.600	47	24	71	9	75	62
0.625	44	28	72	17	71	57
0.650	41	32	73	23	67	54
0.675	38	38	76	32	61	48
0.700	36	43	79	37	56	45
0.725	32	50	82	48	49	38
0.750	31	57	88	54	42	33
0.775	26	65	91	67	34	25
0.800	22	69	91	77	30	19
0.825	19	73	92	82	26	17
0.850	17	78	95	87	21	14
0.875	12	84	96	97	15	9
0.900	9	89	98	104	10	5
0.925	6	93	**99**	109	6	3
0.950	2	96	98	113	3	3
0.975	0	**98**	98	116	1	2

Euclidean distance

Table C.28: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**181**	0	181	5	227	120
0.025	181	0	181	5	227	120
0.050	181	0	181	5	227	120
0.075	181	0	181	5	227	120
0.100	181	0	181	5	227	120
0.125	180	0	180	6	227	120
0.150	179	0	179	7	227	120
0.175	179	0	179	7	227	120
0.200	178	0	178	8	227	120
0.225	173	1	174	13	226	120
0.250	159	5	164	27	222	120
0.275	155	8	163	31	219	120
0.300	147	18	165	39	209	120
0.325	145	19	164	41	208	120
0.350	141	21	162	45	206	120
0.375	139	22	161	47	205	120
0.400	139	22	161	47	205	120
0.425	135	22	157	52	205	119
0.450	134	24	158	53	203	119
0.475	134	27	161	53	200	119
0.500	133	27	160	54	200	119
0.525	129	27	156	58	200	119
0.550	129	27	156	58	200	119
0.575	129	29	158	58	198	119
0.600	129	30	159	60	197	117
0.625	129	32	161	62	195	115
0.650	127	35	162	65	192	114
0.675	124	54	178	73	173	109
0.700	120	79	199	83	148	103
0.725	111	91	202	98	136	97
0.750	108	115	223	111	112	87
0.775	101	137	238	127	90	78
0.800	87	168	**255**	157	59	62
0.825	62	189	251	203	38	41
0.850	45	208	253	236	19	25
0.875	31	216	247	265	11	10
0.900	17	224	241	286	3	3
0.925	13	226	239	293	1	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Table C.29: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**48**	0	48	0	99	70
0.025	48	0	48	0	99	70
0.050	48	0	48	0	99	70
0.075	48	0	48	0	99	70
0.100	48	0	48	0	99	70
0.125	48	0	48	0	99	70
0.150	48	0	48	0	99	70
0.175	48	0	48	0	99	70
0.200	48	0	48	0	99	70
0.225	48	0	48	0	99	70
0.250	48	0	48	0	99	70
0.275	48	0	48	0	99	70
0.300	48	0	48	0	99	70
0.325	48	0	48	0	99	70
0.350	48	0	48	0	99	70
0.375	48	0	48	0	99	70
0.400	48	0	48	0	99	70
0.425	48	0	48	0	99	70
0.450	48	0	48	0	99	70
0.475	48	0	48	0	99	70
0.500	48	0	48	0	99	70
0.525	48	0	48	0	99	70
0.550	48	0	48	0	99	70
0.575	48	0	48	0	99	70
0.600	48	1	49	1	98	69
0.625	48	6	54	1	93	69
0.650	47	16	63	8	83	63
0.675	44	24	68	14	75	60
0.700	40	40	80	30	59	48
0.725	37	53	90	38	46	43
0.750	31	68	99	58	31	29
0.775	23	76	99	76	23	19
0.800	16	82	98	88	17	14
0.825	10	86	96	100	13	8
0.850	6	92	98	106	7	6
0.875	5	97	**102**	110	2	3
0.900	0	98	98	118	1	0
0.925	0	99	99	118	0	0
0.950	0	99	99	118	0	0
0.975	0	**99**	99	118	0	0

Second degree synonyms

Cosine Similarity

Table C.30: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**170**	19	189	21	208	115
0.025	170	19	189	21	208	115
0.050	169	20	189	22	207	115
0.075	165	20	185	27	207	114
0.100	165	22	187	27	205	114
0.125	162	25	187	30	202	114
0.150	152	25	177	40	202	114
0.175	151	27	178	41	200	114
0.200	151	28	179	41	199	114
0.225	151	31	182	41	196	114
0.250	149	36	185	43	191	114
0.275	148	37	185	44	190	114
0.300	146	40	186	46	187	114
0.325	145	40	185	48	187	113
0.350	143	40	183	51	187	112
0.375	142	43	185	52	184	112
0.400	142	45	187	53	182	111
0.425	142	46	188	53	181	111
0.450	141	49	190	55	178	110
0.475	138	51	189	59	176	109
0.500	138	56	194	59	171	109
0.525	138	60	198	60	167	108
0.550	136	63	199	62	164	108
0.575	134	66	200	66	161	106
0.600	132	73	205	72	154	102
0.625	128	80	208	77	147	101
0.650	125	89	214	86	138	95
0.675	125	93	218	90	134	91
0.700	125	95	220	95	132	86
0.725	118	100	218	104	127	84
0.750	116	110	226	107	117	83
0.775	110	127	237	117	100	79
0.800	102	142	244	126	85	78
0.825	91	154	245	146	73	69
0.850	81	169	250	171	58	54
0.875	68	187	**255**	200	40	38
0.900	55	199	254	224	28	27
0.925	40	213	253	257	14	9
0.950	22	222	244	279	5	5
0.975	14	**226**	240	292	1	0

Table C.31: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**49**	3	52	0	96	69
0.025	49	3	52	0	96	69
0.050	49	3	52	0	96	69
0.075	49	3	52	0	96	69
0.100	49	3	52	0	96	69
0.125	49	3	52	0	96	69
0.150	49	3	52	0	96	69
0.175	49	3	52	0	96	69
0.200	49	3	52	0	96	69
0.225	49	3	52	0	96	69
0.250	49	3	52	0	96	69
0.275	49	3	52	0	96	69
0.300	49	3	52	0	96	69
0.325	49	4	53	0	95	69
0.350	49	4	53	0	95	69
0.375	49	4	53	0	95	69
0.400	49	4	53	0	95	69
0.425	49	5	54	0	94	69
0.450	49	7	56	0	92	69
0.475	49	10	59	0	89	69
0.500	49	12	61	1	87	68
0.525	49	14	63	1	85	68
0.550	49	17	66	2	82	67
0.575	47	21	68	4	78	67
0.600	46	26	72	7	73	65
0.625	44	27	71	16	72	58
0.650	42	30	72	22	69	54
0.675	38	37	75	31	62	49
0.700	35	41	76	36	58	47
0.725	30	48	78	45	51	43
0.750	28	57	85	54	42	36
0.775	24	62	86	65	37	29
0.800	20	70	90	75	29	23
0.825	18	74	92	83	25	17
0.850	14	78	92	89	21	15
0.875	10	83	93	99	16	9
0.900	7	88	95	106	11	5
0.925	5	93	98	110	6	3
0.950	1	98	**99**	115	1	2
0.975	0	**98**	98	116	1	2

Euclidean distance

Table C.32: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0	**183**	0	183	5	227	118
0.025	183	0	183	5	227	118
0.050	183	0	183	5	227	118
0.075	183	0	183	5	227	118
0.100	183	0	183	5	227	118
0.125	182	0	182	6	227	118
0.150	181	0	181	7	227	118
0.175	181	0	181	7	227	118
0.200	180	0	180	8	227	118
0.225	175	1	176	13	226	118
0.250	161	4	165	27	223	118
0.275	157	7	164	31	220	118
0.300	149	16	165	39	211	118
0.325	147	19	166	41	208	118
0.350	143	21	164	45	206	118
0.375	141	22	163	47	205	118
0.400	140	22	162	48	205	118
0.425	136	22	158	53	205	117
0.450	136	24	160	53	203	117
0.475	136	27	163	53	200	117
0.500	135	27	162	54	200	117
0.525	132	27	159	57	200	117
0.550	132	27	159	57	200	117
0.575	131	27	158	58	200	117
0.600	131	28	159	60	199	115
0.625	131	29	160	62	198	113
0.650	130	33	163	63	194	113
0.675	127	51	178	71	176	108
0.700	121	78	199	81	149	104
0.725	112	90	202	95	137	99
0.750	108	117	225	109	110	89
0.775	100	135	235	129	92	77
0.800	87	166	253	160	61	59
0.825	62	190	252	204	37	40
0.850	44	210	**254**	237	17	25
0.875	29	219	248	269	8	8
0.900	17	225	242	287	2	2
0.925	13	227	240	293	0	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Scenario Vector

Cosine similarity

Table C.33: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**159**	37	196	28	190	119
0.025	159	37	196	28	190	119
0.050	159	37	196	28	190	119
0.075	156	38	194	31	189	119
0.100	156	38	194	31	189	119
0.125	152	39	191	35	188	119
0.150	144	42	186	44	185	118
0.175	143	42	185	45	185	118
0.200	141	46	187	47	181	118
0.225	140	50	190	48	177	118
0.250	139	52	191	49	175	118
0.275	139	54	193	53	173	114
0.300	135	61	196	59	166	112
0.325	135	65	200	60	162	111
0.350	131	68	199	64	159	111
0.375	131	74	205	65	153	110
0.400	128	79	207	69	148	109
0.425	127	84	211	72	143	107
0.450	125	90	215	76	137	105
0.475	123	96	219	83	131	100
0.500	120	99	219	89	128	97
0.525	117	105	222	95	122	94
0.550	112	109	221	103	118	91
0.575	106	116	222	115	111	85
0.600	99	126	225	132	101	75
0.625	90	138	228	149	89	67
0.650	83	145	228	163	82	60
0.675	75	151	226	179	76	52
0.700	71	159	230	187	68	48
0.725	59	164	223	208	63	39
0.750	53	173	226	221	54	32
0.775	48	187	235	228	40	30
0.800	41	195	236	243	32	22
0.825	34	206	**240**	259	21	13
0.850	30	210	240	269	17	7
0.875	25	215	240	276	12	5
0.900	19	220	239	284	7	3
0.925	14	220	234	290	7	2
0.950	13	222	235	293	5	0
0.975	12	**224**	236	294	3	0

Euclidean distance

Table C.34: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**179**	5	184	3	222	124
0.025	179	5	184	3	222	124
0.050	179	5	184	3	222	124
0.075	179	5	184	3	222	124
0.100	179	5	184	3	222	124
0.125	178	5	183	4	222	124
0.150	176	9	185	6	218	124
0.175	175	10	185	7	217	124
0.200	174	10	184	8	217	124
0.225	166	13	179	16	214	124
0.250	153	15	168	30	212	123
0.275	150	16	166	33	211	123
0.300	133	50	183	53	177	120
0.325	132	63	195	59	164	115
0.350	126	80	206	68	147	112
0.375	120	95	215	84	132	102
0.400	108	114	222	104	113	94
0.425	90	137	227	151	90	65
0.450	77	150	227	178	77	51
0.475	57	172	229	215	55	34
0.500	49	191	**240**	230	36	27
0.525	32	204	236	256	23	18
0.550	27	212	239	271	15	8
0.575	19	217	236	283	10	4
0.600	17	219	236	286	8	3
0.625	14	222	236	290	5	2
0.650	13	223	236	292	4	1
0.675	13	224	237	292	3	1
0.700	13	224	237	292	3	1
0.725	12	225	237	293	2	1
0.750	11	226	237	295	1	0
0.775	11	226	237	295	1	0
0.800	11	226	237	295	1	0
0.825	11	226	237	295	1	0
0.850	11	226	237	295	1	0
0.875	11	226	237	295	1	0
0.900	11	227	238	295	0	0
0.925	11	227	238	295	0	0
0.950	11	227	238	295	0	0
0.975	11	**227**	238	295	0	0

Angle Difference

Table C.35: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**166**	27	193	27	200	113
0.025	166	38	204	28	189	112
0.050	166	38	204	28	189	112
0.075	163	39	202	32	188	111
0.100	163	40	203	33	187	110
0.125	160	41	201	37	186	109
0.150	152	45	197	45	182	109
0.175	151	47	198	47	180	108
0.200	149	49	198	49	178	108
0.225	149	50	199	49	177	108
0.250	148	51	199	53	176	105
0.275	148	51	199	53	176	105
0.300	145	54	199	56	173	105
0.325	144	55	199	57	172	105
0.350	142	56	198	60	171	104
0.375	142	56	198	62	171	102
0.400	142	59	201	62	168	102
0.425	142	63	205	62	164	102
0.450	141	64	205	63	163	102
0.475	139	69	208	66	158	101
0.500	138	73	211	68	154	100
0.525	135	75	210	72	152	99
0.550	134	78	212	74	149	98
0.575	132	82	214	79	145	95
0.600	132	88	220	81	139	93
0.625	130	89	219	84	138	92
0.650	127	98	225	91	129	88
0.675	125	104	229	94	123	87
0.700	124	109	233	99	118	83
0.725	121	115	236	106	112	79
0.750	118	123	241	114	104	74
0.775	112	134	246	121	93	73
0.800	106	142	**248**	130	85	70
0.825	106	142	248	131	85	69
0.850	105	142	247	132	85	69
0.875	105	142	247	132	85	69
0.900	105	142	247	132	85	69
0.925	105	142	247	132	85	69
0.950	105	142	247	132	85	69
0.975	105	**142**	247	132	85	69

Table C.36: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**56**	4	60	1	95	61
0.025	56	4	60	1	95	61
0.050	56	5	61	1	94	61
0.075	56	5	61	1	94	61
0.100	56	5	61	1	94	61
0.125	56	5	61	1	94	61
0.150	56	5	61	1	94	61
0.175	56	5	61	1	94	61
0.200	56	5	61	1	94	61
0.225	55	6	61	2	93	61
0.250	55	6	61	2	93	61
0.275	55	6	61	2	93	61
0.300	55	6	61	2	93	61
0.325	55	7	62	2	92	61
0.350	55	7	62	2	92	61
0.375	55	7	62	2	92	61
0.400	55	8	63	3	91	60
0.425	55	8	63	3	91	60
0.450	55	11	66	4	88	59
0.475	55	12	67	4	87	59
0.500	55	16	71	5	83	58
0.525	55	20	75	6	79	57
0.550	55	23	78	7	76	56
0.575	55	24	79	7	75	56
0.600	55	29	84	9	70	54
0.625	50	34	84	16	65	52
0.650	48	39	87	20	60	50
0.675	41	41	82	30	58	47
0.700	39	45	84	32	54	47
0.725	37	51	88	40	48	41
0.750	37	61	98	47	38	34
0.775	34	62	96	60	37	24
0.800	32	69	**101**	68	30	18
0.825	32	69	101	68	30	18
0.850	32	69	101	68	30	18
0.875	32	69	101	68	30	18
0.900	32	69	101	68	30	18
0.925	32	69	101	68	30	18
0.950	32	69	101	68	30	18
0.975	32	**69**	101	68	30	18

Combinations

Combination 1

Cosine similarity

Table C.37: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**165**	21	186	26	206	115
0.025	165	21	186	26	206	115
0.050	165	21	186	26	206	115
0.075	162	22	184	30	205	114
0.100	162	24	186	30	203	114
0.125	158	27	185	34	200	114
0.150	150	29	179	42	198	114
0.175	149	30	179	43	197	114
0.200	149	31	180	43	196	114
0.225	149	34	183	43	193	114
0.250	147	37	184	45	190	114
0.275	147	37	184	45	190	114
0.300	145	40	185	47	187	114
0.325	145	40	185	48	187	113
0.350	143	41	184	51	186	112
0.375	142	43	185	53	184	111
0.400	142	48	190	54	179	110
0.425	142	49	191	54	178	110
0.450	141	55	196	55	172	110
0.475	138	57	195	61	170	107
0.500	138	62	200	61	165	107
0.525	137	67	204	62	160	107
0.550	136	68	204	64	159	106
0.575	134	72	206	67	155	105
0.600	131	77	208	75	150	100
0.625	129	81	210	77	146	100
0.650	125	88	213	85	139	96
0.675	125	94	219	89	133	92
0.700	125	97	222	94	130	87
0.725	122	104	226	101	123	83
0.750	117	115	232	106	112	83
0.775	110	128	238	120	99	76
0.800	100	139	239	131	88	75
0.825	88	151	239	151	76	67
0.850	79	168	247	175	59	52
0.875	69	180	249	201	47	36
0.900	57	196	**253**	220	31	29
0.925	42	210	252	251	17	13
0.950	23	220	243	277	7	6
0.975	15	**225**	240	291	2	0

Table C.38: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**47**	3	50	1	96	70
0.025	47	3	50	1	96	70
0.050	47	3	50	1	96	70
0.075	47	3	50	1	96	70
0.100	47	3	50	1	96	70
0.125	47	3	50	1	96	70
0.150	47	3	50	1	96	70
0.175	47	3	50	1	96	70
0.200	47	3	50	1	96	70
0.225	47	3	50	1	96	70
0.250	47	3	50	1	96	70
0.275	47	3	50	1	96	70
0.300	47	3	50	1	96	70
0.325	47	4	51	1	95	70
0.350	47	4	51	1	95	70
0.375	47	4	51	1	95	70
0.400	47	5	52	1	94	70
0.425	47	5	52	1	94	70
0.450	47	7	54	1	92	70
0.475	47	9	56	1	90	70
0.500	47	13	60	2	86	69
0.525	47	16	63	2	83	69
0.550	47	20	67	2	79	69
0.575	45	23	68	5	76	68
0.600	45	25	70	8	74	65
0.625	43	32	75	14	67	61
0.650	40	35	75	21	64	57
0.675	38	42	80	28	57	52
0.700	34	47	81	35	52	49
0.725	29	50	79	44	49	45
0.750	28	57	85	53	42	37
0.775	23	62	85	66	37	29
0.800	20	70	90	74	29	24
0.825	20	76	96	76	23	22
0.850	18	81	99	81	18	19
0.875	13	85	98	94	14	11
0.900	9	89	98	103	10	6
0.925	6	94	**100**	109	5	3
0.950	2	97	99	113	2	3
0.975	0	**99**	99	116	0	2

Euclidean distance

Table C.39: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**184**	0	184	6	227	116
0.025	184	0	184	6	227	116
0.050	184	0	184	6	227	116
0.075	184	0	184	6	227	116
0.100	184	0	184	6	227	116
0.125	183	0	183	7	227	116
0.150	182	0	182	8	227	116
0.175	182	0	182	8	227	116
0.200	178	0	178	12	227	116
0.225	172	1	173	18	226	116
0.250	160	6	166	30	221	116
0.275	157	9	166	33	218	116
0.300	149	18	167	41	209	116
0.325	148	19	167	42	208	116
0.350	145	21	166	45	206	116
0.375	143	22	165	47	205	116
0.400	143	22	165	47	205	116
0.425	139	22	161	52	205	115
0.450	138	24	162	53	203	115
0.475	138	27	165	53	200	115
0.500	137	27	164	54	200	115
0.525	133	27	160	58	200	115
0.550	133	27	160	58	200	115
0.575	133	29	162	58	198	115
0.600	133	30	163	60	197	113
0.625	133	33	166	62	194	111
0.650	131	36	167	65	191	110
0.675	128	59	187	72	168	106
0.700	121	81	202	86	146	99
0.725	115	94	209	98	133	93
0.750	111	115	226	109	112	86
0.775	102	136	238	128	91	76
0.800	86	164	250	156	63	64
0.825	64	188	**252**	202	39	40
0.850	46	206	252	235	21	25
0.875	31	217	248	265	10	10
0.900	18	225	243	285	2	3
0.925	13	226	239	293	1	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Combination 2

Cosine similarity

Table C.40: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**168**	20	188	21	207	117
0.025	168	20	188	21	207	117
0.050	167	21	188	22	206	117
0.075	163	21	184	27	206	116
0.100	163	23	186	27	204	116
0.125	160	26	186	30	201	116
0.150	150	28	178	40	199	116
0.175	149	30	179	41	197	116
0.200	149	31	180	41	196	116
0.225	149	34	183	41	193	116
0.250	147	37	184	43	190	116
0.275	146	37	183	44	190	116
0.300	144	40	184	46	187	116
0.325	143	40	183	48	187	115
0.350	141	41	182	51	186	114
0.375	140	42	182	53	185	113
0.400	140	46	186	54	181	112
0.425	140	47	187	54	180	112
0.450	139	52	191	55	175	112
0.475	136	55	191	59	172	111
0.500	136	60	196	60	167	110
0.525	135	66	201	61	161	110
0.550	134	67	201	63	160	109
0.575	131	69	200	67	158	108
0.600	129	76	205	75	151	102
0.625	126	80	206	78	147	102
0.650	123	88	211	84	139	99
0.675	123	94	217	88	133	95
0.700	123	96	219	94	131	89
0.725	120	104	224	101	123	85
0.750	116	114	230	105	113	85
0.775	110	127	237	117	100	79
0.800	102	139	241	127	88	77
0.825	90	152	242	147	75	69
0.850	81	170	251	169	57	56
0.875	70	183	253	197	44	39
0.900	58	197	**255**	221	30	27
0.925	41	211	252	252	16	13
0.950	22	220	242	278	7	6
0.975	14	**225**	239	292	2	0

Euclidean distance

Table C.41: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**184**	0	184	5	227	117
0.025	184	0	184	5	227	117
0.050	184	0	184	5	227	117
0.075	184	0	184	5	227	117
0.100	184	0	184	5	227	117
0.125	183	0	183	6	227	117
0.150	182	0	182	7	227	117
0.175	182	0	182	7	227	117
0.200	181	0	181	8	227	117
0.225	176	1	177	13	226	117
0.250	162	5	167	27	222	117
0.275	158	8	166	31	219	117
0.300	150	18	168	39	209	117
0.325	148	19	167	41	208	117
0.350	144	21	165	45	206	117
0.375	142	22	164	47	205	117
0.400	142	22	164	47	205	117
0.425	138	22	160	52	205	116
0.450	137	24	161	53	203	116
0.475	137	27	164	53	200	116
0.500	136	27	163	54	200	116
0.525	132	27	159	58	200	116
0.550	132	27	159	58	200	116
0.575	132	29	161	58	198	116
0.600	132	30	162	60	197	114
0.625	132	32	164	62	195	112
0.650	130	34	164	65	193	111
0.675	127	56	183	73	171	106
0.700	122	79	201	84	148	100
0.725	114	92	206	97	135	95
0.750	111	113	224	108	114	87
0.775	103	135	238	126	92	77
0.800	87	165	**252**	156	62	63
0.825	63	189	252	202	38	41
0.850	45	207	252	237	20	24
0.875	30	218	248	266	9	10
0.900	17	225	242	286	2	3
0.925	13	226	239	293	1	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Combination 3

Cosine similarity

Table C.42: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**165**	20	185	26	207	115
0.025	165	20	185	26	207	115
0.050	165	20	185	26	207	115
0.075	162	21	183	30	206	114
0.100	162	23	185	30	204	114
0.125	158	26	184	34	201	114
0.150	150	26	176	42	201	114
0.175	149	27	176	43	200	114
0.200	149	28	177	43	199	114
0.225	149	31	180	43	196	114
0.250	147	36	183	45	191	114
0.275	147	37	184	45	190	114
0.300	145	40	185	47	187	114
0.325	145	40	185	48	187	113
0.350	143	40	183	51	187	112
0.375	142	44	186	52	183	112
0.400	142	47	189	53	180	111
0.425	142	48	190	53	179	111
0.450	141	52	193	56	175	109
0.475	138	54	192	60	173	108
0.500	138	59	197	60	168	108
0.525	138	62	200	61	165	107
0.550	136	64	200	63	163	107
0.575	135	68	203	65	159	106
0.600	132	75	207	72	152	102
0.625	130	81	211	76	146	100
0.650	125	88	213	87	139	94
0.675	125	92	217	90	135	91
0.700	125	96	221	95	131	86
0.725	118	101	219	104	126	84
0.750	115	110	225	108	117	83
0.775	109	128	237	119	99	78
0.800	101	139	240	129	88	76
0.825	90	153	243	149	74	67
0.850	79	167	246	176	60	51
0.875	67	184	251	203	43	36
0.900	55	198	**253**	222	29	29
0.925	41	212	253	255	15	10
0.950	23	222	245	278	5	5
0.975	15	**225**	240	291	2	0

Euclidean distance

Table C.43: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**183**	0	183	6	227	117
0.025	183	0	183	6	227	117
0.050	183	0	183	6	227	117
0.075	183	0	183	6	227	117
0.100	183	0	183	6	227	117
0.125	182	0	182	7	227	117
0.150	181	0	181	8	227	117
0.175	181	0	181	8	227	117
0.200	177	0	177	12	227	117
0.225	171	1	172	18	226	117
0.250	159	5	164	30	222	117
0.275	156	8	164	33	219	117
0.300	148	16	164	41	211	117
0.325	147	19	166	42	208	117
0.350	144	21	165	45	206	117
0.375	142	22	164	47	205	117
0.400	141	22	163	48	205	117
0.425	137	22	159	53	205	116
0.450	137	24	161	53	203	116
0.475	137	27	164	53	200	116
0.500	136	27	163	54	200	116
0.525	133	27	160	57	200	116
0.550	133	27	160	57	200	116
0.575	132	27	159	58	200	116
0.600	132	28	160	60	199	114
0.625	132	30	162	62	197	112
0.650	131	34	165	63	193	112
0.675	128	56	184	70	171	108
0.700	121	79	200	82	148	103
0.725	113	92	205	97	135	96
0.750	108	118	226	110	109	88
0.775	100	134	234	130	93	76
0.800	84	165	249	162	62	60
0.825	63	191	254	203	36	40
0.850	45	210	**255**	235	17	26
0.875	30	219	249	268	8	8
0.900	18	226	244	286	1	2
0.925	13	227	240	293	0	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Combination 4

Cosine similarity

Table C.44: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**170**	19	189	21	208	115
0.025	170	19	189	21	208	115
0.050	169	20	189	22	207	115
0.075	165	20	185	27	207	114
0.100	165	22	187	27	205	114
0.125	162	25	187	30	202	114
0.150	152	25	177	40	202	114
0.175	151	27	178	41	200	114
0.200	151	28	179	41	199	114
0.225	151	31	182	41	196	114
0.250	149	36	185	43	191	114
0.275	148	37	185	44	190	114
0.300	146	40	186	46	187	114
0.325	145	40	185	48	187	113
0.350	143	40	183	51	187	112
0.375	142	43	185	52	184	112
0.400	142	45	187	53	182	111
0.425	142	46	188	53	181	111
0.450	141	49	190	55	178	110
0.475	138	51	189	59	176	109
0.500	138	56	194	59	171	109
0.525	138	60	198	60	167	108
0.550	136	63	199	62	164	108
0.575	134	66	200	66	161	106
0.600	132	73	205	72	154	102
0.625	128	80	208	77	147	101
0.650	125	89	214	86	138	95
0.675	125	93	218	90	134	91
0.700	125	95	220	95	132	86
0.725	118	101	219	104	126	84
0.750	116	110	226	107	117	83
0.775	110	127	237	117	100	79
0.800	102	142	244	126	85	78
0.825	91	154	245	146	73	69
0.850	81	169	250	171	58	54
0.875	68	187	**255**	200	40	38
0.900	55	199	254	224	28	27
0.925	40	213	253	256	14	10
0.950	22	222	244	279	5	5
0.975	14	**225**	239	292	2	0

Table C.45: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**49**	3	52	0	96	69
0.025	49	3	52	0	96	69
0.050	49	3	52	0	96	69
0.075	49	3	52	0	96	69
0.100	49	3	52	0	96	69
0.125	49	3	52	0	96	69
0.150	49	3	52	0	96	69
0.175	49	3	52	0	96	69
0.200	49	3	52	0	96	69
0.225	49	3	52	0	96	69
0.250	49	3	52	0	96	69
0.275	49	3	52	0	96	69
0.300	49	3	52	0	96	69
0.325	49	4	53	0	95	69
0.350	49	4	53	0	95	69
0.375	49	4	53	0	95	69
0.400	49	4	53	0	95	69
0.425	49	5	54	0	94	69
0.450	49	7	56	0	92	69
0.475	49	10	59	0	89	69
0.500	49	12	61	1	87	68
0.525	49	14	63	1	85	68
0.550	49	17	66	2	82	67
0.575	47	21	68	4	78	67
0.600	46	26	72	7	73	65
0.625	44	28	72	16	71	58
0.650	41	31	72	23	68	54
0.675	38	37	75	31	62	49
0.700	35	41	76	36	58	47
0.725	30	48	78	45	51	43
0.750	28	57	85	54	42	36
0.775	24	62	86	65	37	29
0.800	20	70	90	75	29	23
0.825	18	74	92	83	25	17
0.850	14	78	92	89	21	15
0.875	10	83	93	99	16	9
0.900	7	88	95	106	11	5
0.925	5	93	98	110	6	3
0.950	1	98	**99**	115	1	2
0.975	0	**98**	98	116	1	2

Euclidean distance

Table C.46: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**183**	0	183	5	227	118
0.025	183	0	183	5	227	118
0.050	183	0	183	5	227	118
0.075	183	0	183	5	227	118
0.100	183	0	183	5	227	118
0.125	182	0	182	6	227	118
0.150	181	0	181	7	227	118
0.175	181	0	181	7	227	118
0.200	180	0	180	8	227	118
0.225	175	1	176	13	226	118
0.250	161	4	165	27	223	118
0.275	157	7	164	31	220	118
0.300	149	16	165	39	211	118
0.325	147	19	166	41	208	118
0.350	143	21	164	45	206	118
0.375	141	22	163	47	205	118
0.400	140	22	162	48	205	118
0.425	136	22	158	53	205	117
0.450	136	24	160	53	203	117
0.475	136	27	163	53	200	117
0.500	135	27	162	54	200	117
0.525	132	27	159	57	200	117
0.550	132	27	159	57	200	117
0.575	131	27	158	58	200	117
0.600	131	28	159	60	199	115
0.625	131	29	160	62	198	113
0.650	130	32	162	63	195	113
0.675	127	51	178	71	176	108
0.700	121	78	199	81	149	104
0.725	112	90	202	95	137	99
0.750	108	117	225	109	110	89
0.775	100	135	235	129	92	77
0.800	86	166	252	161	61	59
0.825	62	190	252	204	37	40
0.850	44	210	**254**	237	17	25
0.875	29	220	249	268	7	9
0.900	17	226	243	287	1	2
0.925	13	227	240	293	0	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Combination 5

Cosine similarity

Table C.47: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**169**	20	189	24	207	113
0.025	169	20	189	24	207	113
0.050	168	21	189	26	206	112
0.075	165	22	187	29	205	112
0.100	164	24	188	30	203	112
0.125	154	26	180	40	201	112
0.150	153	29	182	41	198	112
0.175	152	30	182	42	197	112
0.200	152	32	184	42	195	112
0.225	151	32	183	43	195	112
0.250	150	35	185	44	192	112
0.275	150	37	187	44	190	112
0.300	149	40	189	45	187	112
0.325	147	40	187	47	187	112
0.350	145	41	186	49	186	112
0.375	143	43	186	52	184	111
0.400	143	44	187	52	183	111
0.425	143	44	187	53	183	110
0.450	141	49	190	55	178	110
0.475	141	53	194	58	174	107
0.500	141	57	198	58	170	107
0.525	139	58	197	61	169	106
0.550	137	61	198	64	166	105
0.575	136	66	202	66	161	104
0.600	135	69	204	68	158	103
0.625	133	74	207	72	153	101
0.650	131	81	212	76	146	99
0.675	128	88	216	82	139	96
0.700	127	94	221	86	133	93
0.725	123	102	225	91	125	92
0.750	122	109	231	97	118	87
0.775	115	118	233	110	109	81
0.800	107	130	237	122	97	77
0.825	95	142	237	135	85	76
0.850	84	156	240	163	71	59
0.875	72	168	240	187	59	47
0.900	56	192	**248**	219	35	31
0.925	38	204	242	250	23	18
0.950	24	221	245	278	6	4
0.975	15	**224**	239	291	3	0

Table C.48: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**45**	2	47	1	97	72
0.025	45	2	47	1	97	72
0.050	45	2	47	1	97	72
0.075	45	2	47	1	97	72
0.100	45	2	47	1	97	72
0.125	45	2	47	1	97	72
0.150	45	2	47	1	97	72
0.175	45	2	47	1	97	72
0.200	45	2	47	1	97	72
0.225	45	2	47	1	97	72
0.250	45	2	47	1	97	72
0.275	45	2	47	1	97	72
0.300	45	2	47	1	97	72
0.325	45	2	47	1	97	72
0.350	45	2	47	1	97	72
0.375	45	2	47	1	97	72
0.400	45	2	47	1	97	72
0.425	45	2	47	1	97	72
0.450	45	4	49	1	95	72
0.475	45	6	51	1	93	72
0.500	45	12	57	1	87	72
0.525	45	12	57	1	87	72
0.550	44	15	59	3	84	71
0.575	43	19	62	6	80	69
0.600	42	21	63	10	78	66
0.625	40	27	67	14	72	64
0.650	38	34	72	21	65	59
0.675	37	36	73	26	63	55
0.700	35	43	78	33	56	50
0.725	33	48	81	41	51	44
0.750	27	53	80	48	46	43
0.775	25	56	81	55	43	38
0.800	22	61	83	62	38	34
0.825	21	68	89	71	31	26
0.850	20	76	96	76	23	22
0.875	14	84	98	87	15	17
0.900	11	86	97	96	13	11
0.925	5	92	97	107	7	6
0.950	4	96	**100**	110	3	4
0.975	0	**99**	99	116	0	2

Euclidean distance

Table C.49: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**185**	0	185	6	227	115
0.025	185	0	185	6	227	115
0.050	185	0	185	6	227	115
0.075	185	0	185	6	227	115
0.100	185	0	185	6	227	115
0.125	184	0	184	7	227	115
0.150	183	0	183	8	227	115
0.175	183	0	183	8	227	115
0.200	179	0	179	12	227	115
0.225	176	1	177	15	226	115
0.250	164	5	169	27	222	115
0.275	158	9	167	33	218	115
0.300	152	17	169	39	210	115
0.325	149	19	168	42	208	115
0.350	146	21	167	45	206	115
0.375	144	22	166	47	205	115
0.400	141	22	163	50	205	115
0.425	139	22	161	53	205	114
0.450	139	24	163	53	203	114
0.475	139	24	163	53	203	114
0.500	137	27	164	55	200	114
0.525	135	27	162	57	200	114
0.550	135	27	162	57	200	114
0.575	135	27	162	57	200	114
0.600	134	30	164	59	197	113
0.625	134	31	165	61	196	111
0.650	134	32	166	62	195	110
0.675	131	48	179	65	179	110
0.700	130	60	190	73	167	103
0.725	126	79	205	83	148	97
0.750	119	100	219	97	127	90
0.775	113	116	229	111	111	82
0.800	100	140	240	131	87	75
0.825	74	162	236	177	65	55
0.850	50	194	244	221	33	35
0.875	36	209	**245**	252	18	18
0.900	20	222	242	282	5	4
0.925	13	226	239	293	1	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Combination 6

Cosine similarity

Table C.50: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**168**	27	195	26	200	112
0.025	168	37	205	27	190	111
0.050	167	37	204	28	190	111
0.075	164	38	202	32	189	110
0.100	163	39	202	33	188	110
0.125	155	40	195	43	187	108
0.150	153	43	196	46	184	107
0.175	152	44	196	47	183	107
0.200	150	46	196	49	181	107
0.225	150	46	196	50	181	106
0.250	147	47	194	54	180	105
0.275	147	47	194	54	180	105
0.300	146	47	193	55	180	105
0.325	146	49	195	55	178	105
0.350	143	49	192	58	178	105
0.375	143	49	192	59	178	104
0.400	142	51	193	60	176	104
0.425	142	57	199	61	170	103
0.450	140	59	199	63	168	103
0.475	140	60	200	63	167	103
0.500	140	65	205	64	162	102
0.525	135	70	205	70	157	101
0.550	134	74	208	71	153	101
0.575	132	80	212	74	147	100
0.600	131	82	213	79	145	96
0.625	130	84	214	83	143	93
0.650	126	91	217	90	136	90
0.675	125	99	224	92	128	89
0.700	125	104	229	93	123	88
0.725	121	109	230	101	118	84
0.750	116	114	230	108	113	82
0.775	111	129	**240**	117	98	78
0.800	105	135	240	128	92	73
0.825	105	135	240	128	92	73
0.850	104	135	239	129	92	73
0.875	104	135	239	129	92	73
0.900	104	135	239	129	92	73
0.925	104	135	239	129	92	73
0.950	103	135	238	130	92	73
0.975	103	**135**	238	130	92	73

Table C.51: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**54**	4	58	1	95	63
0.025	54	4	58	1	95	63
0.050	54	5	59	1	94	63
0.075	54	5	59	1	94	63
0.100	54	5	59	1	94	63
0.125	54	5	59	1	94	63
0.150	54	5	59	1	94	63
0.175	54	5	59	1	94	63
0.200	54	6	60	1	93	63
0.225	53	6	59	2	93	63
0.250	53	7	60	2	92	63
0.275	53	7	60	2	92	63
0.300	53	7	60	2	92	63
0.325	53	7	60	2	92	63
0.350	53	7	60	2	92	63
0.375	53	7	60	2	92	63
0.400	53	8	61	3	91	62
0.425	53	8	61	3	91	62
0.450	53	10	63	3	89	62
0.475	53	13	66	4	86	61
0.500	53	15	68	4	84	61
0.525	53	19	72	5	80	60
0.550	53	23	76	8	76	57
0.575	53	27	80	9	72	56
0.600	52	27	79	12	72	54
0.625	49	34	83	16	65	53
0.650	48	37	85	20	62	50
0.675	45	40	85	23	59	50
0.700	45	44	89	27	55	46
0.725	36	50	86	39	49	43
0.750	33	56	89	49	43	36
0.775	33	59	92	59	40	26
0.800	31	65	**96**	65	34	22
0.825	31	65	96	65	34	22
0.850	31	65	96	65	34	22
0.875	31	65	96	65	34	22
0.900	31	65	96	65	34	22
0.925	31	65	96	65	34	22
0.950	31	65	96	65	34	22
0.975	31	**65**	96	65	34	22

Euclidean distance

Table C.52: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	183	0	183	6	227	117
0.025	183	0	183	6	227	117
0.050	183	0	183	6	227	117
0.075	183	0	183	6	227	117
0.100	183	0	183	6	227	117
0.125	182	0	182	7	227	117
0.150	181	0	181	8	227	117
0.175	181	0	181	8	227	117
0.200	177	0	177	12	227	117
0.225	174	1	175	15	226	117
0.250	162	4	166	27	223	117
0.275	156	8	164	33	219	117
0.300	150	15	165	39	212	117
0.325	147	19	166	42	208	117
0.350	144	21	165	45	206	117
0.375	142	22	164	47	205	117
0.400	139	22	161	50	205	117
0.425	137	22	159	53	205	116
0.450	137	24	161	53	203	116
0.475	137	24	161	53	203	116
0.500	135	27	162	55	200	116
0.525	133	27	160	57	200	116
0.550	133	27	160	57	200	116
0.575	133	27	160	57	200	116
0.600	132	28	160	59	199	115
0.625	132	29	161	61	198	113
0.650	132	29	161	62	198	112
0.675	129	48	177	65	179	112
0.700	127	59	186	72	168	107
0.725	123	76	199	83	151	100
0.750	118	98	216	94	129	94
0.775	113	114	227	109	113	84
0.800	100	137	237	132	90	74
0.825	74	161	235	179	66	53
0.850	52	193	245	220	34	34
0.875	36	211	247	252	16	18
0.900	20	224	244	283	3	3
0.925	13	227	240	293	0	0
0.950	12	227	239	294	0	0
0.975	12	227	239	294	0	0

Combination 7

Table C.53: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**164**	27	191	27	200	115
0.025	164	38	202	28	189	114
0.050	164	38	202	28	189	114
0.075	161	39	200	32	188	113
0.100	161	40	201	34	187	111
0.125	156	42	198	39	185	111
0.150	149	45	194	46	182	111
0.175	148	46	194	48	181	110
0.200	147	47	194	49	180	110
0.225	147	48	195	49	179	110
0.250	146	49	195	52	178	108
0.275	146	49	195	52	178	108
0.300	143	52	195	55	175	108
0.325	142	53	195	56	174	108
0.350	140	54	194	59	173	107
0.375	140	55	195	61	172	105
0.400	140	57	197	61	170	105
0.425	140	61	201	61	166	105
0.450	139	63	202	62	164	105
0.475	137	68	205	65	159	104
0.500	137	72	209	67	155	102
0.525	134	75	209	70	152	102
0.550	133	78	211	73	149	100
0.575	132	81	213	77	146	97
0.600	132	86	218	79	141	95
0.625	131	87	218	80	140	95
0.650	126	95	221	89	132	91
0.675	124	102	226	94	125	88
0.700	123	106	229	98	121	85
0.725	120	113	233	104	114	82
0.750	116	124	240	114	103	76
0.775	111	134	245	120	93	75
0.800	106	141	**247**	127	86	73
0.825	106	141	247	128	86	72
0.850	105	141	246	129	86	72
0.875	105	141	246	129	86	72
0.900	105	141	246	129	86	72
0.925	105	141	246	129	86	72
0.950	104	141	245	130	86	72
0.975	104	**141**	245	130	86	72

Table C.54: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**49**	4	53	1	95	68
0.025	49	4	53	1	95	68
0.050	49	5	54	1	94	68
0.075	49	5	54	1	94	68
0.100	49	5	54	1	94	68
0.125	49	5	54	1	94	68
0.150	49	5	54	1	94	68
0.175	49	5	54	1	94	68
0.200	49	6	55	1	93	68
0.225	49	6	55	1	93	68
0.250	49	6	55	1	93	68
0.275	49	6	55	1	93	68
0.300	49	6	55	1	93	68
0.325	49	7	56	1	92	68
0.350	49	7	56	1	92	68
0.375	49	7	56	1	92	68
0.400	49	8	57	1	91	68
0.425	49	8	57	1	91	68
0.450	49	11	60	2	88	67
0.475	49	12	61	2	87	67
0.500	49	17	66	3	82	66
0.525	49	22	71	4	77	65
0.550	49	25	74	5	74	64
0.575	48	26	74	7	73	63
0.600	47	31	78	10	68	61
0.625	44	35	79	14	64	60
0.650	43	40	83	18	59	57
0.675	38	43	81	27	56	53
0.700	38	46	84	28	53	52
0.725	33	51	84	37	48	48
0.750	33	58	91	45	41	40
0.775	32	60	92	55	39	31
0.800	30	69	**99**	63	30	25
0.825	30	69	99	63	30	25
0.850	30	69	99	63	30	25
0.875	30	69	99	63	30	25
0.900	30	69	99	63	30	25
0.925	30	69	99	63	30	25
0.950	30	69	99	63	30	25
0.975	30	**69**	99	63	30	25

Combination 8

Table C.55: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**165**	27	192	27	200	114
0.025	165	37	202	28	190	113
0.050	165	37	202	28	190	113
0.075	162	38	200	32	189	112
0.100	162	39	201	33	188	111
0.125	159	40	199	37	187	110
0.150	151	43	194	45	184	110
0.175	150	45	195	47	182	109
0.200	148	47	195	49	180	109
0.225	148	48	196	49	179	109
0.250	147	49	196	52	178	107
0.275	147	49	196	52	178	107
0.300	144	52	196	55	175	107
0.325	143	53	196	56	174	107
0.350	141	53	194	59	174	106
0.375	141	53	194	61	174	104
0.400	141	56	197	61	171	104
0.425	141	61	202	61	166	104
0.450	140	62	202	62	165	104
0.475	138	66	204	65	161	103
0.500	138	70	208	66	157	102
0.525	134	73	207	71	154	101
0.550	133	78	211	73	149	100
0.575	131	84	215	78	143	97
0.600	131	88	219	81	139	94
0.625	129	89	218	84	138	93
0.650	127	98	225	89	129	90
0.675	125	103	228	92	124	89
0.700	124	108	232	96	119	86
0.725	121	113	234	101	114	84
0.750	119	121	240	108	106	79
0.775	112	133	**245**	118	94	76
0.800	106	139	245	127	88	73
0.825	106	139	245	127	88	73
0.850	105	139	244	128	88	73
0.875	105	139	244	128	88	73
0.900	105	139	244	128	88	73
0.925	105	139	244	128	88	73
0.950	105	139	244	128	88	73
0.975	105	**139**	244	128	88	73

Table C.56: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**55**	4	59	1	95	62
0.025	55	4	59	1	95	62
0.050	55	5	60	1	94	62
0.075	55	5	60	1	94	62
0.100	55	5	60	1	94	62
0.125	55	5	60	1	94	62
0.150	55	5	60	1	94	62
0.175	55	5	60	1	94	62
0.200	55	6	61	1	93	62
0.225	54	6	60	2	93	62
0.250	54	6	60	2	93	62
0.275	54	6	60	2	93	62
0.300	54	6	60	2	93	62
0.325	54	7	61	2	92	62
0.350	54	7	61	2	92	62
0.375	54	7	61	2	92	62
0.400	54	7	61	3	92	61
0.425	54	7	61	3	92	61
0.450	54	11	65	4	88	60
0.475	54	13	67	4	86	60
0.500	54	16	70	6	83	58
0.525	54	21	75	8	78	56
0.550	54	25	79	8	74	56
0.575	54	26	80	9	73	55
0.600	53	29	82	13	70	52
0.625	50	34	84	17	65	51
0.650	48	40	88	20	59	50
0.675	41	42	83	30	57	47
0.700	39	47	86	33	52	46
0.725	34	51	85	43	48	41
0.750	34	58	92	50	41	34
0.775	33	58	91	59	41	26
0.800	31	66	**97**	66	33	21
0.825	31	66	97	66	33	21
0.850	31	66	97	66	33	21
0.875	31	66	97	66	33	21
0.900	31	66	97	66	33	21
0.925	31	66	97	66	33	21
0.950	31	66	97	66	33	21
0.975	31	**66**	97	66	33	21

Cosine similarity

Table C.57: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**168**	19	187	24	208	114
0.025	168	19	187	24	208	114
0.050	167	20	187	26	207	113
0.075	164	21	185	29	206	113
0.100	163	23	186	30	204	113
0.125	153	25	178	40	202	113
0.150	152	26	178	41	201	113
0.175	151	27	178	42	200	113
0.200	151	29	180	42	198	113
0.225	150	29	179	43	198	113
0.250	149	34	183	44	193	113
0.275	149	37	186	44	190	113
0.300	148	40	188	45	187	113
0.325	146	40	186	47	187	113
0.350	144	41	185	49	186	113
0.375	142	43	185	52	184	112
0.400	142	44	186	52	183	112
0.425	142	44	186	53	183	111
0.450	140	48	188	55	179	111
0.475	139	53	192	57	174	110
0.500	139	55	194	58	172	109
0.525	137	56	193	61	171	108
0.550	135	58	193	63	169	108
0.575	135	63	198	65	164	106
0.600	134	65	199	68	162	104
0.625	132	71	203	71	156	103
0.650	130	81	211	77	146	99
0.675	129	87	216	83	140	94
0.700	128	92	220	87	135	91
0.725	126	98	224	91	129	89
0.750	124	106	230	96	121	86
0.775	117	117	234	109	110	80
0.800	108	131	239	122	96	76
0.825	97	143	240	134	84	75
0.850	84	154	238	164	73	58
0.875	74	170	244	186	57	46
0.900	58	193	**251**	219	34	29
0.925	40	204	244	248	23	18
0.950	24	221	245	278	6	4
0.975	15	**224**	239	291	3	0

Table C.58: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**45**	2	47	0	97	73
0.025	45	2	47	0	97	73
0.050	45	2	47	0	97	73
0.075	45	2	47	0	97	73
0.100	45	2	47	0	97	73
0.125	45	2	47	0	97	73
0.150	45	2	47	0	97	73
0.175	45	2	47	0	97	73
0.200	45	2	47	0	97	73
0.225	45	2	47	0	97	73
0.250	45	2	47	0	97	73
0.275	45	2	47	0	97	73
0.300	45	2	47	0	97	73
0.325	45	2	47	0	97	73
0.350	45	2	47	0	97	73
0.375	45	2	47	0	97	73
0.400	45	2	47	0	97	73
0.425	45	2	47	0	97	73
0.450	45	4	49	0	95	73
0.475	45	5	50	0	94	73
0.500	45	12	57	0	87	73
0.525	45	12	57	0	87	73
0.550	44	14	58	2	85	72
0.575	43	18	61	5	81	70
0.600	42	20	62	8	79	68
0.625	39	25	64	14	74	65
0.650	37	33	70	21	66	60
0.675	36	35	71	26	64	56
0.700	34	43	77	32	56	52
0.725	31	47	78	42	52	45
0.750	26	53	79	49	46	43
0.775	24	58	82	56	41	38
0.800	22	62	84	61	37	35
0.825	21	69	90	71	30	26
0.850	18	76	94	79	23	21
0.875	13	84	97	90	15	15
0.900	10	88	98	99	11	9
0.925	4	92	96	109	7	5
0.950	3	98	**101**	112	1	3
0.975	0	**99**	99	116	0	2

Euclidean distance

Table C.59: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**183**	0	183	6	227	117
0.025	183	0	183	6	227	117
0.050	183	0	183	6	227	117
0.075	183	0	183	6	227	117
0.100	183	0	183	6	227	117
0.125	182	0	182	7	227	117
0.150	181	0	181	8	227	117
0.175	181	0	181	8	227	117
0.200	177	0	177	12	227	117
0.225	174	1	175	15	226	117
0.250	162	4	166	27	223	117
0.275	156	8	164	33	219	117
0.300	150	15	165	39	212	117
0.325	147	19	166	42	208	117
0.350	144	21	165	45	206	117
0.375	142	22	164	47	205	117
0.400	139	22	161	50	205	117
0.425	137	22	159	53	205	116
0.450	137	24	161	53	203	116
0.475	137	24	161	53	203	116
0.500	135	27	162	55	200	116
0.525	133	27	160	57	200	116
0.550	133	27	160	57	200	116
0.575	133	27	160	57	200	116
0.600	132	28	160	59	199	115
0.625	132	29	161	61	198	113
0.650	132	29	161	62	198	112
0.675	129	48	177	65	179	112
0.700	127	59	186	72	168	107
0.725	123	76	199	83	151	100
0.750	118	98	216	94	129	94
0.775	113	114	227	109	113	84
0.800	100	137	237	132	90	74
0.825	74	161	235	179	66	53
0.850	52	193	245	220	34	34
0.875	36	211	**247**	252	16	18
0.900	20	224	244	283	3	3
0.925	13	227	240	293	0	0
0.950	12	227	239	294	0	0
0.975	12	**227**	239	294	0	0

Table C.60: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**38**	0	38	0	99	80
0.025	38	0	38	0	99	80
0.050	38	0	38	0	99	80
0.075	38	0	38	0	99	80
0.100	38	0	38	0	99	80
0.125	38	0	38	0	99	80
0.150	38	0	38	0	99	80
0.175	38	0	38	0	99	80
0.200	38	0	38	0	99	80
0.225	38	0	38	0	99	80
0.250	38	0	38	0	99	80
0.275	38	0	38	0	99	80
0.300	38	0	38	0	99	80
0.325	38	0	38	0	99	80
0.350	38	0	38	0	99	80
0.375	38	0	38	0	99	80
0.400	38	0	38	0	99	80
0.425	38	0	38	0	99	80
0.450	38	0	38	0	99	80
0.475	38	0	38	0	99	80
0.500	38	0	38	0	99	80
0.525	38	0	38	0	99	80
0.550	38	0	38	0	99	80
0.575	38	0	38	0	99	80
0.600	38	0	38	0	99	80
0.625	38	4	42	0	95	80
0.650	37	12	49	2	87	79
0.675	35	18	53	6	81	77
0.700	33	27	60	18	72	67
0.725	31	41	72	25	58	62
0.750	29	54	83	34	45	55
0.775	25	68	93	52	31	41
0.800	17	77	94	75	22	26
0.825	13	86	99	90	13	15
0.850	6	92	98	104	7	8
0.875	6	96	**102**	108	3	4
0.900	2	98	100	114	1	2
0.925	0	99	99	116	0	2
0.950	0	99	99	118	0	0
0.975	0	**99**	99	118	0	0

Combination 9

Table C.61: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**168**	26	194	25	201	113
0.025	168	35	203	26	192	112
0.050	167	37	204	27	190	112
0.075	164	38	202	32	189	110
0.100	163	38	201	35	189	108
0.125	153	40	193	45	187	108
0.150	152	45	197	47	182	107
0.175	151	46	197	48	181	107
0.200	151	47	198	48	180	107
0.225	149	49	198	50	178	107
0.250	149	49	198	50	178	107
0.275	147	49	196	54	178	105
0.300	143	52	195	59	175	104
0.325	141	53	194	61	174	104
0.350	138	53	191	64	174	104
0.375	137	54	191	65	173	104
0.400	137	54	191	66	173	103
0.425	137	57	194	67	170	102
0.450	136	62	198	68	165	102
0.475	135	65	200	69	162	102
0.500	135	70	205	70	157	101
0.525	132	72	204	73	155	101
0.550	131	75	206	75	152	100
0.575	131	76	207	76	151	99
0.600	129	78	207	79	149	98
0.625	129	84	213	83	143	94
0.650	126	85	211	89	142	91
0.675	124	92	216	93	135	89
0.700	123	97	220	95	130	88
0.725	119	104	223	101	123	86
0.750	115	114	229	108	113	83
0.775	111	120	231	115	107	80
0.800	104	129	**233**	124	98	78
0.825	104	129	233	125	98	77
0.850	104	129	233	125	98	77
0.875	104	129	233	125	98	77
0.900	104	129	233	125	98	77
0.925	104	129	233	125	98	77
0.950	104	129	233	125	98	77
0.975	104	**129**	233	125	98	77

Combination 10

Combination 11

Table C.62: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**170**	20	190	21	207	115
0.025	170	32	202	23	195	113
0.050	169	33	202	24	194	113
0.075	165	34	199	29	193	112
0.100	165	36	201	30	191	111
0.125	162	38	200	33	189	111
0.150	152	40	192	44	187	110
0.175	151	41	192	46	186	109
0.200	150	42	192	48	185	108
0.225	150	42	192	48	185	108
0.250	148	44	192	51	183	107
0.275	147	44	191	52	183	107
0.300	145	47	192	54	180	107
0.325	144	47	191	56	180	106
0.350	142	48	190	59	179	105
0.375	141	48	189	61	179	104
0.400	141	52	193	62	175	103
0.425	141	53	194	62	174	103
0.450	140	58	198	63	169	103
0.475	137	61	198	67	166	102
0.500	137	65	202	68	162	101
0.525	135	69	204	70	158	101
0.550	134	70	204	72	157	100
0.575	131	73	204	75	154	100
0.600	129	79	208	82	148	95
0.625	126	81	207	85	146	95
0.650	123	90	213	90	137	93
0.675	123	95	218	93	132	90
0.700	123	98	221	98	129	85
0.725	119	105	224	106	122	81
0.750	117	116	233	108	111	81
0.775	112	128	240	118	99	76
0.800	104	139	**243**	127	88	75
0.825	104	139	243	127	88	75
0.850	103	139	242	128	88	75
0.875	103	139	242	128	88	75
0.900	103	139	242	128	88	75
0.925	102	139	241	129	88	75
0.950	102	139	241	129	88	75
0.975	102	**139**	241	129	88	75

Table C.63: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**49**	3	52	1	96	68
0.025	49	3	52	1	96	68
0.050	49	4	53	1	95	68
0.075	49	4	53	1	95	68
0.100	49	4	53	1	95	68
0.125	49	4	53	1	95	68
0.150	49	4	53	1	95	68
0.175	49	4	53	1	95	68
0.200	48	4	52	2	95	68
0.225	48	4	52	3	95	67
0.250	48	4	52	3	95	67
0.275	48	4	52	3	95	67
0.300	48	4	52	3	95	67
0.325	48	4	52	3	95	67
0.350	48	4	52	3	95	67
0.375	48	4	52	3	95	67
0.400	48	5	53	4	94	66
0.425	48	5	53	4	94	66
0.450	48	7	55	4	92	66
0.475	48	9	57	4	90	66
0.500	48	12	60	4	87	66
0.525	48	14	62	5	85	65
0.550	48	18	66	6	81	64
0.575	47	21	68	7	78	64
0.600	47	24	71	10	75	61
0.625	44	28	72	18	71	56
0.650	41	32	73	23	67	54
0.675	38	38	76	32	61	48
0.700	36	43	79	37	56	45
0.725	32	50	82	48	49	38
0.750	31	57	88	54	42	33
0.775	26	65	**91**	67	34	25
0.800	22	69	91	77	30	19
0.825	22	69	91	77	30	19
0.850	22	69	91	77	30	19
0.875	22	69	91	77	30	19
0.900	22	69	91	77	30	19
0.925	22	69	91	77	30	19
0.950	22	69	91	77	30	19
0.975	22	**69**	91	77	30	19

Combination 12

Table C.64: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**170**	19	189	21	208	115
0.025	170	31	201	23	196	113
0.050	169	32	201	24	195	113
0.075	165	32	197	29	195	112
0.100	165	33	198	29	194	112
0.125	162	35	197	32	192	112
0.150	152	35	187	44	192	110
0.175	151	36	187	46	191	109
0.200	150	37	187	49	190	107
0.225	150	37	187	49	190	107
0.250	148	41	189	52	186	106
0.275	147	42	189	53	185	106
0.300	145	45	190	55	182	106
0.325	144	45	189	57	182	105
0.350	142	45	187	60	182	104
0.375	141	47	188	61	180	104
0.400	141	49	190	62	178	103
0.425	141	50	191	62	177	103
0.450	140	53	193	64	174	102
0.475	138	55	193	67	172	101
0.500	138	59	197	67	168	101
0.525	137	63	200	69	164	100
0.550	135	66	201	71	161	100
0.575	133	69	202	74	158	99
0.600	131	76	207	79	151	96
0.625	127	80	207	84	147	95
0.650	124	89	213	91	138	91
0.675	124	92	216	94	135	88
0.700	124	94	218	98	133	84
0.725	117	99	216	106	128	83
0.750	117	109	226	106	118	83
0.775	112	126	238	115	101	79
0.800	104	140	**244**	124	87	78
0.825	104	140	244	124	87	78
0.850	103	140	243	125	87	78
0.875	103	140	243	125	87	78
0.900	103	140	243	125	87	78
0.925	102	140	242	126	87	78
0.950	102	140	242	126	87	78
0.975	102	**140**	242	126	87	78

Table C.65: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**49**	3	52	0	96	69
0.025	49	3	52	0	96	69
0.050	49	4	53	0	95	69
0.075	49	4	53	0	95	69
0.100	49	4	53	0	95	69
0.125	49	4	53	0	95	69
0.150	49	4	53	0	95	69
0.175	49	4	53	0	95	69
0.200	48	4	52	1	95	69
0.225	48	5	53	2	94	68
0.250	48	5	53	2	94	68
0.275	48	5	53	2	94	68
0.300	48	5	53	2	94	68
0.325	48	5	53	2	94	68
0.350	48	5	53	2	94	68
0.375	48	5	53	2	94	68
0.400	48	5	53	2	94	68
0.425	48	6	54	2	93	68
0.450	48	8	56	2	91	68
0.475	48	11	59	2	88	68
0.500	48	13	61	3	86	67
0.525	48	15	63	3	84	67
0.550	48	18	66	4	81	66
0.575	46	22	68	6	77	66
0.600	46	26	72	8	73	64
0.625	44	27	71	17	72	57
0.650	42	30	72	22	69	54
0.675	38	37	75	31	62	49
0.700	35	41	76	36	58	47
0.725	30	48	78	45	51	43
0.750	28	57	85	54	42	36
0.775	24	62	86	65	37	29
0.800	20	70	**90**	75	29	23
0.825	20	70	90	75	29	23
0.850	20	70	90	75	29	23
0.875	20	70	90	75	29	23
0.900	20	70	90	75	29	23
0.925	20	70	90	75	29	23
0.950	20	70	90	75	29	23
0.975	20	**70**	90	75	29	23

Combination 13

Table C.66: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**165**	21	186	26	206	115
0.025	165	33	198	28	194	113
0.050	165	33	198	28	194	113
0.075	162	34	196	32	193	112
0.100	162	36	198	33	191	111
0.125	158	38	196	37	189	111
0.150	150	40	190	45	187	111
0.175	149	40	189	48	187	109
0.200	148	41	189	50	186	108
0.225	147	41	188	51	186	108
0.250	145	43	188	54	184	107
0.275	145	43	188	54	184	107
0.300	143	46	189	56	181	107
0.325	143	46	189	57	181	106
0.350	141	47	188	60	180	105
0.375	140	48	188	62	179	104
0.400	140	53	193	63	174	103
0.425	140	54	194	63	173	103
0.450	139	60	199	64	167	103
0.475	137	62	199	68	165	101
0.500	137	66	203	68	161	101
0.525	135	70	205	70	157	101
0.550	134	71	205	72	156	100
0.575	132	75	207	74	152	100
0.600	130	80	210	80	147	96
0.625	128	81	209	82	146	96
0.650	124	88	212	89	139	93
0.675	124	93	217	92	134	90
0.700	124	96	220	97	131	85
0.725	121	103	224	103	124	82
0.750	118	114	232	106	113	82
0.775	112	127	**239**	118	100	76
0.800	102	137	239	129	90	75
0.825	102	137	239	129	90	75
0.850	101	137	238	130	90	75
0.875	101	137	238	130	90	75
0.900	101	137	238	130	90	75
0.925	100	137	237	131	90	75
0.950	100	137	237	131	90	75
0.975	100	**137**	237	131	90	75

Table C.67: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**47**	3	50	1	96	70
0.025	47	3	50	1	96	70
0.050	47	4	51	1	95	70
0.075	47	4	51	1	95	70
0.100	47	4	51	1	95	70
0.125	47	4	51	1	95	70
0.150	47	4	51	1	95	70
0.175	47	4	51	1	95	70
0.200	47	4	51	1	95	70
0.225	47	4	51	2	95	69
0.250	47	4	51	2	95	69
0.275	47	4	51	2	95	69
0.300	47	4	51	2	95	69
0.325	47	4	51	2	95	69
0.350	47	4	51	2	95	69
0.375	47	4	51	2	95	69
0.400	47	5	52	2	94	69
0.425	47	5	52	2	94	69
0.450	47	7	54	2	92	69
0.475	47	9	56	2	90	69
0.500	47	13	60	3	86	68
0.525	47	16	63	3	83	68
0.550	47	20	67	3	79	68
0.575	45	23	68	6	76	67
0.600	45	25	70	9	74	64
0.625	43	32	75	15	67	60
0.650	40	35	75	21	64	57
0.675	38	42	80	28	57	52
0.700	34	47	81	35	52	49
0.725	29	50	79	44	49	45
0.750	28	57	85	53	42	37
0.775	23	62	85	66	37	29
0.800	20	70	**90**	74	29	24
0.825	20	70	90	74	29	24
0.850	20	70	90	74	29	24
0.875	20	70	90	74	29	24
0.900	20	70	90	74	29	24
0.925	20	70	90	74	29	24
0.950	20	70	90	74	29	24
0.975	20	**70**	90	74	29	24

Combination 14

Table C.68: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**168**	20	188	21	207	117
0.025	168	32	200	23	195	115
0.050	167	33	200	24	194	115
0.075	163	33	196	29	194	114
0.100	163	35	198	30	192	113
0.125	160	37	197	33	190	113
0.150	150	39	189	44	188	112
0.175	149	40	189	46	187	111
0.200	148	41	189	48	186	110
0.225	148	41	189	48	186	110
0.250	146	43	189	51	184	109
0.275	145	43	188	52	184	109
0.300	143	46	189	54	181	109
0.325	142	46	188	56	181	108
0.350	140	47	187	59	180	107
0.375	139	47	186	61	180	106
0.400	139	51	190	62	176	105
0.425	139	52	191	62	175	105
0.450	138	57	195	63	170	105
0.475	136	60	196	66	167	104
0.500	136	64	200	67	163	103
0.525	134	69	203	69	158	103
0.550	133	70	203	71	157	102
0.575	130	72	202	74	155	102
0.600	128	79	207	81	148	97
0.625	125	80	205	84	147	97
0.650	122	88	210	89	139	95
0.675	122	93	215	92	134	92
0.700	122	95	217	97	132	87
0.725	119	103	222	103	124	84
0.750	117	113	230	105	114	84
0.775	112	126	238	115	101	79
0.800	104	137	**241**	125	90	77
0.825	104	137	241	125	90	77
0.850	103	137	240	126	90	77
0.875	103	137	240	126	90	77
0.900	103	137	240	126	90	77
0.925	102	137	239	127	90	77
0.950	102	137	239	127	90	77
0.975	102	**137**	239	127	90	77

Combination 15

Table C.69: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**165**	20	185	26	207	115
0.025	165	32	197	28	195	113
0.050	165	32	197	28	195	113
0.075	162	33	195	32	194	112
0.100	162	34	196	32	193	112
0.125	158	36	194	36	191	112
0.150	150	36	186	46	191	110
0.175	149	36	185	48	191	109
0.200	147	37	184	51	190	108
0.225	147	37	184	51	190	108
0.250	145	41	186	54	186	107
0.275	145	42	187	54	185	107
0.300	143	45	188	56	182	107
0.325	143	45	188	57	182	106
0.350	141	45	186	60	182	105
0.375	140	48	188	61	179	105
0.400	140	51	191	62	176	104
0.425	140	52	192	62	175	104
0.450	139	56	195	64	171	103
0.475	137	58	195	67	169	102
0.500	137	62	199	67	165	102
0.525	136	65	201	69	162	101
0.550	134	67	201	71	160	101
0.575	133	71	204	73	156	100
0.600	131	78	209	78	149	97
0.625	129	81	210	82	146	95
0.650	124	88	212	91	139	91
0.675	124	91	215	93	136	89
0.700	124	95	219	98	132	84
0.725	117	100	217	106	127	83
0.750	116	109	225	107	118	83
0.775	111	127	238	117	100	78
0.800	103	137	**240**	127	90	76
0.825	103	137	240	127	90	76
0.850	102	137	239	128	90	76
0.875	102	137	239	128	90	76
0.900	102	137	239	128	90	76
0.925	101	137	238	129	90	76
0.950	101	137	238	129	90	76
0.975	101	**137**	238	129	90	76

Combination 16

Table C.70: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**170**	19	189	21	208	115
0.025	170	31	201	23	196	113
0.050	169	32	201	24	195	113
0.075	165	32	197	29	195	112
0.100	165	33	198	29	194	112
0.125	162	35	197	32	192	112
0.150	152	35	187	44	192	110
0.175	151	36	187	46	191	109
0.200	150	37	187	49	190	107
0.225	150	37	187	49	190	107
0.250	148	41	189	52	186	106
0.275	147	42	189	53	185	106
0.300	145	45	190	55	182	106
0.325	144	45	189	57	182	105
0.350	142	45	187	60	182	104
0.375	141	47	188	61	180	104
0.400	141	49	190	62	178	103
0.425	141	50	191	62	177	103
0.450	140	53	193	64	174	102
0.475	138	55	193	67	172	101
0.500	138	59	197	67	168	101
0.525	137	63	200	69	164	100
0.550	135	66	201	71	161	100
0.575	133	69	202	74	158	99
0.600	131	76	207	79	151	96
0.625	127	80	207	84	147	95
0.650	124	89	213	91	138	91
0.675	124	92	216	94	135	88
0.700	124	94	218	98	133	84
0.725	117	100	217	106	127	83
0.750	117	109	226	106	118	83
0.775	112	126	238	115	101	79
0.800	104	140	**244**	124	87	78
0.825	104	140	244	124	87	78
0.850	103	140	243	125	87	78
0.875	103	140	243	125	87	78
0.900	103	140	243	125	87	78
0.925	102	140	242	126	87	78
0.950	102	140	242	126	87	78
0.975	102	**140**	242	126	87	78

Table C.71: Emotion information set - Validation test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**49**	3	52	0	96	69
0.025	49	3	52	0	96	69
0.050	49	4	53	0	95	69
0.075	49	4	53	0	95	69
0.100	49	4	53	0	95	69
0.125	49	4	53	0	95	69
0.150	49	4	53	0	95	69
0.175	49	4	53	0	95	69
0.200	48	4	52	1	95	69
0.225	48	5	53	2	94	68
0.250	48	5	53	2	94	68
0.275	48	5	53	2	94	68
0.300	48	5	53	2	94	68
0.325	48	5	53	2	94	68
0.350	48	5	53	2	94	68
0.375	48	5	53	2	94	68
0.400	48	5	53	2	94	68
0.425	48	6	54	2	93	68
0.450	48	8	56	2	91	68
0.475	48	11	59	2	88	68
0.500	48	13	61	3	86	67
0.525	48	15	63	3	84	67
0.550	48	18	66	4	81	66
0.575	46	22	68	6	77	66
0.600	46	26	72	8	73	64
0.625	44	28	72	17	71	57
0.650	41	31	72	23	68	54
0.675	38	37	75	31	62	49
0.700	35	41	76	36	58	47
0.725	30	48	78	45	51	43
0.750	28	57	85	54	42	36
0.775	24	62	86	65	37	29
0.800	20	70	**90**	75	29	23
0.825	20	70	90	75	29	23
0.850	20	70	90	75	29	23
0.875	20	70	90	75	29	23
0.900	20	70	90	75	29	23
0.925	20	70	90	75	29	23
0.950	20	70	90	75	29	23
0.975	20	**70**	90	75	29	23

Combination 17

Table C.72: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**169**	20	189	24	207	113
0.025	169	31	200	25	196	112
0.050	168	33	201	26	194	112
0.075	165	34	199	31	193	110
0.100	164	35	199	33	192	109
0.125	154	37	191	43	190	109
0.150	153	40	193	45	187	108
0.175	152	40	192	46	187	108
0.200	152	41	193	46	186	108
0.225	151	43	194	48	184	107
0.250	150	43	193	49	184	107
0.275	148	43	191	52	184	106
0.300	146	46	192	55	181	105
0.325	144	46	190	57	181	105
0.350	141	46	187	61	181	104
0.375	139	48	187	64	179	103
0.400	139	48	187	64	179	103
0.425	139	49	188	65	178	102
0.450	137	54	191	67	173	102
0.475	137	58	195	69	169	100
0.500	137	62	199	69	165	100
0.525	135	63	198	72	164	99
0.550	133	66	199	75	161	98
0.575	132	69	201	76	158	98
0.600	131	72	203	78	155	97
0.625	129	76	205	82	151	95
0.650	127	79	206	86	148	93
0.675	124	86	210	92	141	90
0.700	123	92	215	94	135	89
0.725	119	100	219	99	127	88
0.750	118	107	225	104	120	84
0.775	113	116	229	113	111	80
0.800	107	128	**235**	122	99	77
0.825	107	128	235	122	99	77
0.850	107	128	235	122	99	77
0.875	107	128	235	122	99	77
0.900	107	128	235	122	99	77
0.925	107	128	235	122	99	77
0.950	107	128	235	122	99	77
0.975	107	**128**	235	122	99	77

Combination 18

Table C.73: Emotion information set - Training test data

Threshold	SOM	NOM	Total Correct	SOMasNOM	NOMasSOM	WrongSOM
0.0	**168**	19	187	24	208	114
0.025	168	30	198	25	197	113
0.050	167	32	199	26	195	113
0.075	164	33	197	31	194	111
0.100	162	34	196	33	193	111
0.125	152	36	188	43	191	111
0.150	151	37	188	46	190	109
0.175	150	37	187	47	190	109
0.200	150	38	188	47	189	109
0.225	149	40	189	49	187	108
0.250	148	42	190	50	185	108
0.275	146	43	189	53	184	107
0.300	144	46	190	56	181	106
0.325	142	46	188	59	181	105
0.350	139	46	185	62	181	105
0.375	137	48	185	65	179	104
0.400	137	48	185	65	179	104
0.425	137	49	186	66	178	103
0.450	135	53	188	68	174	103
0.475	135	58	193	69	169	102
0.500	135	60	195	70	167	101
0.525	133	61	194	73	166	100
0.550	131	63	194	75	164	100
0.575	131	66	197	76	161	99
0.600	130	68	198	79	159	97
0.625	128	73	201	82	154	96
0.650	126	79	205	88	148	92
0.675	125	85	210	93	142	88
0.700	124	90	214	95	137	87
0.725	122	96	218	99	131	85
0.750	120	104	224	103	123	83
0.775	115	115	230	112	112	79
0.800	108	129	**237**	122	98	76
0.825	108	129	237	122	98	76
0.850	108	129	237	122	98	76
0.875	108	129	237	122	98	76
0.900	108	129	237	122	98	76
0.925	108	129	237	122	98	76
0.950	108	129	237	122	98	76
0.975	108	**129**	237	122	98	76