

Drug Name Similarity Index for Sound-Alikeness

Tomoyuki Nagata^{1*}, Masaomi Kimura², Michiko Ohkura², and Fumito Tsuchiya³

¹ Graduate School, Shibaura Institute of Technology, Tokyo, Japan
ma12078@shibaura-it.ac.jp

² Shibaura Institute of Technology, Tokyo, Japan
{masaomi, ohkura}@sic.shibaura-it.ac.jp

³ International University of Health and Welfare, Japan
iuhwfumito@gmail.com

Abstract. Drug name confusion is one of major medical errors. Some similar drug names can cause medical accidents. In order to solve this problem, the Ministry of Health, Labour and Welfare developed the drug name database system to prevent from authorizing drugs whose names are similar to existing drug names. Previous studies have been proposed the drug name similarity index based on look-alikeness. Despite of these efforts, the studies do not take account of drug name confusion caused by sound-alike. In this paper, we proposed the phonetic similarity index based on the features used in articulatory phonetics.

Keywords: Drug name confusion, Sound-alikeness, Name similarity index, Medical error, Human error.

1 Introduction

Though drugs are to promote people's health, they can be toxic if used improperly. People who use them are not perfect and can cause a medical error. One of the promising ways to prevent such an error is to device the systems which reduce human errors. Drug name confusion is one of major human errors related to safe use of drugs. Actually, some drug names are so similar that it makes medical experts confused unintentionally. For example, at some hospital in Japan, Succin, instead of Saxizon, was mistakenly administered to a patient and it unfortunately made him die. In order to prevent such accidents, the Ministry of Health, Labour and Welfare developed the system, Medicine Similar Name Search Engine, in order to prevent authorizing drugs whose names are similar to existing drug names. The system computes similarity of names based on character matching, such as edit distance and head and tail cosine similarity (htco). However, the system is inadequate to compute similarity more close to human sensitivity. Because the system computes similarity based on only matching character but without consideration the character shapes and the pronunciation. For example, ツルメリン and ソルノミン are also drug names in Japanese. The existing system computes the similarity of them low because the characters of them are almost different. However, because their shapes are very similar, people feel that

these names are very similar in many cases. In order to solve this problem, Nabeta et al. proposed the drug name similarity index by improving the Letter-Sequence-Kernel (LSK)[1-2].

Despite of these efforts, the studies cannot perfectly prevent confusing drug names. This is because they only measure the look-alike similarity based on character matching. As Lambert et al. mentioned, drug name confusion can be caused by both sound-alike and look-alike similarity [3]. Doctors are often forced to do it in emergency, though they are prohibited from oral ordering. In such a situation, the drugs sound-alike with other drugs can cause mixing-up. We should also note that pronunciations can affect the name similarity, if we read them carefully. For this reason, in this study, we focus on drug name confusion caused by sound-alikeness and propose the phonetic similarity index of drug names.

2 Method

Before we define phonetic similarity index, we have to clarify the structure of speech sounds. Phonetics is the study of speech sounds and classifies speech sounds to three fields, articulatory phonetics, acoustic phonetics and auditory phonetics. Articulatory phonetics deal with how humans produce speech sounds on the basis of biological organ structures. Acoustic phonetics deal with how speech sounds is propagated in terms of frequency. Auditory phonetics deal with how humans hear speech sounds. Considering this, we might say that the similarity of speech sounds is consisted of the similarity of pronunciation, sound vibration, and perceptual process. In these fields, the considerable factor in first is similarity of pronunciation because if pronunciations of drug names are similar, these speech sounds are also similar in other fields. Thus, we just focus on articulatory phonetics and classify speech sounds based on their production process.

Articulatory phonetics defines speech sounds as a set of consonant and vowel. Consonant is a speech sound that is articulated with complete or partial closure of the vocal tract, and vowel is pronounced with an open vocal tract. These sounds are determined involved in the manner of articulation and the place of articulation. Manner of articulation describes how speech organs produce sounds, and place of articulation describes the points where block air stream. The International Phonetic Alphabet (IPA) defines speech sounds based on these features.

Especially, in Japanese, a character of Japanese phonetic alphabet consists of pair of consonant and vowel. Thus, we defined each similarity of consonant part and vowel part. When we defined the similarity of consonant part, we considered the similarity of each consonant because it is considered that a pair of characters which has similar consonant each other is more similar than the other pair. For example, Japanese character ‘タ’ (/ta/) is more similar to ‘ダ’ (/da/) than to ‘マ’ (/ma/). Since these characters have the same vowels, we can ignore the contribution of vowels to similarity in this case. This leads the fact that the consonant /t/ is more similar to /d/ than /m/. The difference comes from the features of pronunciations. For this reason, we

defined the similarity of each consonant based on the IPA definition and classified them into some groups.

As for definition process of the similarity of consonant part, we defined the similarity of each consonant based on the features of consonant. We use this similarity to define phonetic version of Letter Sequence Kernel index.

In many languages, consonants are produced by discharging of the air from lungs. These are described by five factors: state of glottis, places of articulation, path of airflow, degree of occlusion, state of palatine velum. State of glottis is presence or absence of the vibration of vocal folds. The voiced sounds are the sounds which involve the vibration, and the voiceless sounds are the sounds which do not involve the vibration. Places of articulation are the places where block the airflow from lungs. We pronounce consonants by blocking the airflow by means of a combination of upper and lower speech organs such as bottom lip and upper lip, blade of tongue and teeth-ridge, and so on. For example, the consonant of ‘ p^h ’ (/pa/x) in Japanese is pronounced by the combination of bottom lip and upper lip, whose sounds are called bilabial. If paths of airflow are not blocked, airflow runs through a center or edges of oral cavity to avoid tongue. Degree of occlusion can be expressed as a distance of two speech organs that block the airflow. Palatine velum is organ to regulate the airflow to nose. The sounds produced by the airflow flown to nose are called nasal sound. If the sounds are produced when palatine velum ups and blocks the airflow to nose, they are called plosive sound.

We defined the similarity index of consonants as the cosine value of the angle between the vectors representing whether the consonant has the corresponding feature.

By contrast, we defined the similarity of vowel part based on Letter-Sequence-Kernel.

3 Experiment

In order to estimate accuracy of the proposed similarity index, we conducted the experiment which compared it with the similarity of drug names based on human sensitivity. In this experiment, 20's university students answered the similarities of 20 pairs of drug names. For each pair of drug names, we played the subjects the sound of one of drug names, and after one-second pause, we played the sound of another. After subjects heard the sounds, they answered the similarity by means of Virtual Analogue Scale (VAS). VAS is a scale which measure subjective characteristics. The subject marks a point on the line between two end points, dissimilar and similar, to express the similarity that he felt. The response time was limited to 1.5 seconds for the reason to avoid the effects caused by memories. Figure 1 shows the correlation between the values of our similarity index and the averages of the subjects' answers.

The horizontal axis indicates the averages of the subjects' answers, and the vertical axis indicates our similarity index. This shows the tendency that the values of our index are higher than the similarities that the subjects felt.

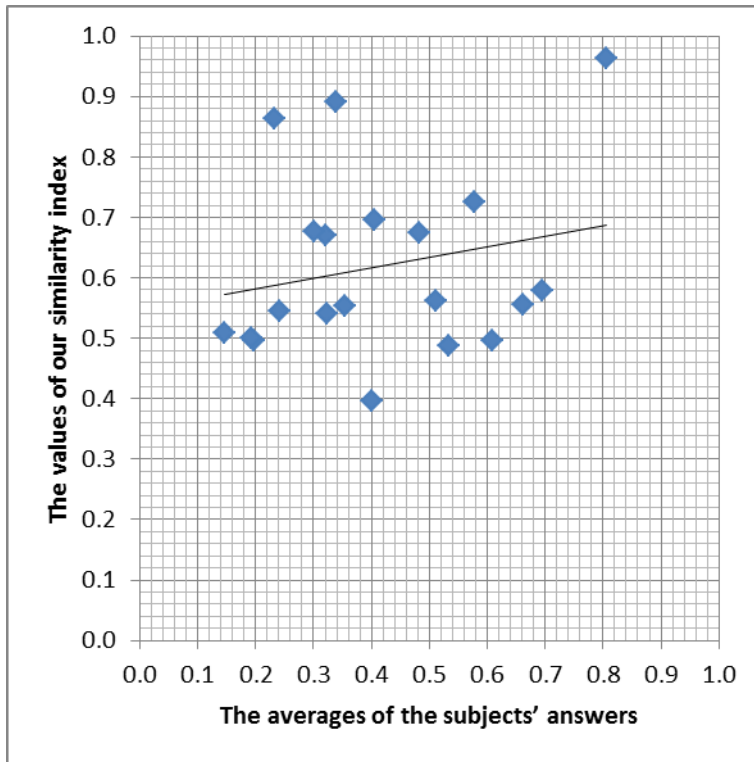


Fig. 1. The correlation between the values of our similarity index and the averages of the subject's answers

We investigated the cause of this problem and found that the contribution of consonant part generally gave high value to any pairs.

The consonant part of the proposed similarity is calculated based on the similarities between each pair of consonants included in the drug names. The cause can be our assumption that the phonetic features of consonants are equally taken in account to define the consonant similarity. We will investigate this in details in our future work.

In addition, the cause might originate in our definition of the similarity that was only based on consonant and vowel. The other factors which should be taken in consideration might be the coincidence of the characters at the head of drug names, the similarity of accent positions and so on.

4 Conclusion

In this paper, we proposed the similarity index of drug name based on sound-alikeness.

We used the property of Japanese letter sound, namely, each sound can be divided into consonant and vowel. We defined the drug name similarity index that consists of

a consonant part and a vowel part. As for the consonant part, we employed the idea used in articulatory phonetics, classification of consonant based on the features of IPA. We used the similarity of consonants to define the consonant part of similarity. The similarity index of consonants was defined as the cosine value of the angle between the vectors corresponding to them. The elements of the vectors represent whether the consonant has the corresponding feature.

In order to evaluate the proposed similarity, we conducted the experiment where the subjects heard sounds of two drug names and answered the extent of their similarity. We investigated a correlation between the measured similarities in this experiment and the values of our similarity index. As a result, we found that the consonant part of our similarity index was higher than the subjects felt. Its cause can be our assumption that the phonetic features of consonants are equally taken in account to quantify the consonant similarity.

In the future, we need to improve the similarity by taking account of acoustic phonetics and auditory phonetics. Additionally, we will research the factor that affects a similarity of drug names, whose candidates are the positions of similar characters, prolonged sound, accents and so on.

References

1. Nabeta, K., et al.: The Similarity Index of Medicine Names Based on Character Shape Similarity. *Transactions of Japan Society of Kansei Engineering* 10(2), 287–294 (2010)
2. Nabeta, K., Hatano, A., Ishida, H., Kimura, M., Ohkura, M., Tsuchiya, F.: The Similarity Index of Character Shape of Medicine Names Based on Character Shape Similarity (II). In: Jacko, J.A. (ed.) *Human-Computer Interaction, Part I, HCII 2011*. LNCS, vol. 6761, pp. 628–636. Springer, Heidelberg (2011)
3. Lambert, B.L., Chang, K.Y., Lin, S.J.: Effect of Orthographic and Phonological Similarity on False Recognition of Drug Names. *Social Science & Medicine* 52, 1843–1857 (2001)
4. Tsuchiya, F., Kawamura, N.: Development of the System to Objectively Evaluate Similarities of Drug Names. In: *Proceeding of APAMI – SJKMI and KOSMI Conference*, pp. 348–351 (2003)